

UNIVERSITY OF GENOVA

SCHOOL OF MATHEMATICAL, PHYSICAL AND NATURAL SCIENCE

PHYSICS DEPARTMENT



**Università
di Genova**

Ph.D. IN PHYSICS AND NANOSCIENCES - XXXVI CICLO

The Higgs, the Beauty and the Charm:

Improving Jet Flavour-Tagging and Higgs Boson Measurements with
Graph Neural Networks in the ATLAS Experiment at the LHC.

Candidate:

Martino Tanasini

Supervisors:

Dr. Andrea Coccaro

Prof. Carlo Schiavi

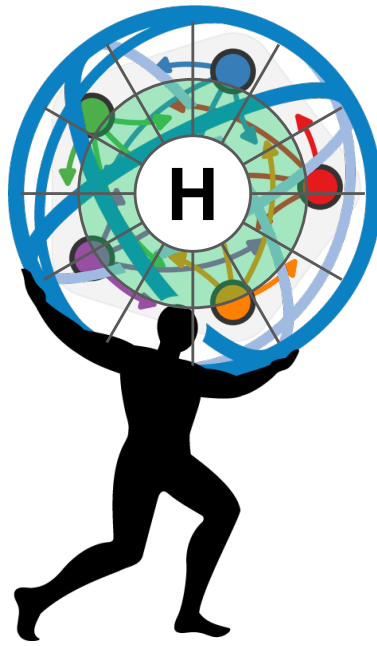
External referees:

Dr. Arnaud Duperrin

Prof. Marumi Kado

Academic year 2023-2024





*Io stimo più il trovar un vero, benché di cosa
leggiera, che 'l disputar lungamente delle massime
questioni senza conseguir verità nissuna.*

Galileo Galilei

Contents

Introduction	1
1 The Standard Model of particle physics	5
1.1 The SM Lagrangian	6
1.1.1 The Brout-Englert-Higgs mechanism	9
1.1.2 Fermion masses: the Yukawa term	12
1.2 SM parameter counting	13
2 The ATLAS experiment at the LHC	15
2.1 The Large Hadron Collider	15
2.2 ATLAS: A Toroidal LHC ApparatuS	18
2.2.1 Inner Detector	19
2.2.2 Calorimeters	22
2.2.3 Muon Spectrometer	23
2.2.4 Magnets	24
2.2.5 Triggering and data acquisition	24
2.3 Reconstruction of detector-level and physics objects	25
2.3.1 Detector-level objects	25
2.3.2 Physics objects	29
2.4 Data Quality monitoring	31
2.4.1 DQ infrastructure and operations	32
2.5 Monte Carlo generators for hadron colliders	34

3	Jet flavour-tagging in ATLAS	37
3.1	Identikit of a heavy-flavour jet	37
3.2	A recent history of jet flavour-tagging in ATLAS	39
3.2.1	Algorithms based on IPs of tracks	44
3.2.2	Algorithms based on displaced vertices	49
3.2.3	Soft Muon Tagger: SMT	55
3.2.4	The DL1 high-level algorithms	55
3.2.5	Performance	60
3.3	Developing SV1 with SVKine (and JFKine)	61
3.4	GN1 and GN2 : jet flavour-tagging using GNNs	71
3.4.1	The GN1 algorithm	71
3.4.2	The GN2 algorithm:	76
3.5	Calibration	79
3.6	Offline jet flavour-tagging DQ	85
3.6.1	A harsh beginning for the Run-3 DQ: exemplifying how monitoring DQ can help recover problematic data	87
4	Jet flavour-tagging @ high p_T	95
4.1	Jet flavour information in the distribution of clusters in the ID	101
4.2	Preliminary studies on cluster-based jet flavour-tagging within the ATLAS Collaboration	103
4.2.1	BDT using IP3D and clusters	104
4.3	Jet flavour-tagging with graphs and clusters	107
4.3.1	GNNs trained solely on clusters: the GN1ClusOnly and GN2ClusOnly algorithms	107
4.3.2	GN2+Clus_No_Aux : improving GN2 high p_T performance with clusters	118
4.3.3	GN2+Clus : outperforming the official GN2 with clusters	125
5	Truth jet flavour-tagging with GNNs	145
5.1	Jet flavour-tagging efficiency parametrization	146
5.1.1	Efficiency parametrization with (p_T, η) maps	147

5.1.2	Efficiency parametrization with GNNs	148
5.1.3	Event weighting	152
5.2	Performance	153
6	Measurements of $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays in the VH production channel	157
6.1	Motivation	158
6.1.1	Improving the $(W/Z)H, H \rightarrow b\bar{b}$ and $(W/Z)H, H \rightarrow c\bar{c}$ measurements in ATLAS	163
6.2	Identikit of the signal and the main backgrounds	164
6.2.1	Reconstructing the Higgs boson	164
6.2.2	Reconstructing the vector boson	169
6.2.3	The main backgrounds	173
6.3	Data and simulated samples	175
6.3.1	Data	175
6.3.2	Simulated samples	175
6.4	Physics objects selection	181
6.5	Event selection and categorisation	181
6.5.1	Triggers	183
6.5.2	Offline selections	183
6.5.3	Event categorisation	188
6.6	Corrections to the Higgs-boson candidate mass	194
6.7	Multivariate BDT to enhance signal vs background separation	197
6.7.1	Data and MC BDT output distributions in the SRs	205
6.8	Systematic uncertainties	205
6.8.1	Experimental uncertainties	205
6.8.2	Modelling uncertainties	209
6.9	Modelling V +jets events	212
6.9.1	Nominal and alternative MC samples	215
6.9.2	Deriving normalisation and relative acceptance uncertainties	229
6.9.3	V + jets shape uncertainties	233

6.10	Statistical treatment and inference	236
6.10.1	The profile binned likelihood function	236
6.10.2	$VH, H \rightarrow b\bar{b}/c\bar{c}$ fit model	240
6.11	GNN truth-tagging application in $VH, H \rightarrow b\bar{b}/c\bar{c}$	242
6.12	Blinded Results	249
6.12.1	$(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ combined fit	249
6.12.2	Expected results for the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ process	259
6.12.3	Postfit Distributions	259
Conclusions		270
“Eppur si muove...”		274
Appendices		275
A	Unblinded results of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ measurement	277
A.1	$(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ combined fit	277
A.2	Observed results for the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ process	279

Introduction

The Standard Model (SM) of particle physics is arguably the scientific theory that has resisted the greatest amount of falsification attempts in the history of humankind. Yet, many parameters that govern its predictions still exceed the reach of the experiments or are measured with coarse precision. Among these, the most important from the standpoint of this research are the couplings of the Higgs boson with the other particles of the SM. Such parameters are of no minor importance, as they govern, among countless other things, the stability of the proton, which is required for the existence of hydrogen; the radii of the atoms, which determine chemical reactions; and many radioactive decay rates and fusion reactions, that power the Sun and influence the lifetime of the stars [1]. Additionally, most of the theories attempting to go beyond the SM, which contains problematic aspects and cannot explain spectacular astrophysical and cosmological observations, predict some deviations of these couplings from their SM predictions. This means that a precise measurement of their values serves as a crucial probe to push our understanding of fundamental physics forward.

This thesis presents the single direct measurement expected to provide the most precise estimation to date of the $H \rightarrow b\bar{b}$ decay rate in the VH production channel, where V stands for the vector bosons W and Z . This estimation is expressed in terms of the signal strength $\mu_{VH, H \rightarrow b\bar{b}}$ of the $VH, H \rightarrow b\bar{b}$ process. Such signal strength, defined as the ratio of the observed cross-section of the $VH, H \rightarrow b\bar{b}$ process over the SM prediction, is expected to be measured as $\mu_{VH, H \rightarrow b\bar{b}}^{\text{exp}} = 1.00 \pm 0.13$, corresponding to a significance of the $VH, H \rightarrow b\bar{b}$ signal over the background-only hypothesis of 8σ . In the same VH channel, the measurement is expected to provide the most stringent 95% Confidence Level (CL) upper limit on the signal strength of the $VH, H \rightarrow c\bar{c}$ process set by the ATLAS Collaboration, namely 11_{-3}^{+4} . It is also expected to mark the first observation of the $WH, H \rightarrow b\bar{b}$ process, with a significance of 5.2σ over the background-only hypothesis; and the first observation of the $VZ, Z \rightarrow c\bar{c}$ process by the ATLAS Collaboration, with an expected significance of 5.5σ . The measurement, most noticeably sensitive to the coupling of the Higgs boson with the beauty and charm quarks, is performed using a dataset comprising 140 fb^{-1} of proton-proton collision data collected by the ATLAS experiment at the Large Hadron Collider (LHC) between 2015 and 2018.

Additionally, this thesis describes innovative algorithms, based on Graph Neural Networks (GNNs), that improve the efficiency with which hadronic jets containing b or c -hadrons

can be identified within the ATLAS detector, most noticeably when they are produced with high momenta. As quarks and gluons are never observed in isolation due to colour confinement, hadronic jets are the experimental observables that indicate the production of colour charges during collisions. Hence, these types of algorithms, termed jet flavour-tagging algorithms, play a crucial role in the analyses targeting final states featuring b or c -hadrons. **GN2+Clus**, a novel jet flavour-tagging algorithm introduced in this thesis, leverages not only the reconstructed tracks produced by charged particles in the ATLAS Inner Detector (ID) but also incorporates the clusters of silicon hits directly. **GN2+Clus** succeeds in improving the background jet rejection rates of the ATLAS state-of-the-art jet flavour-tagging algorithms of significant factors, with peaks of $\sim 80\%$ ($\sim 10\%$) of additional rejection when used to identify b -jets (c -jets) with projections p_T of their momenta over the plane orthogonal to the proton-proton collision axis of approximately $1 \text{ TeV} < p_T < 2 \text{ TeV}$. GNN-based jet flavour-tagging algorithms utilizing only the clusters of silicon hits in the ATLAS ID are also introduced, whose promising performance opens the door to future applications in the context of fast trigger pre-selection of jets, considering that, in such a context, tracking algorithms drain most of the available computational resources.

Modelling simulated backgrounds for processes selected with stringent jet flavour-tagging requirements can pose challenges, particularly when the selection discards a significant portion of the simulated statistics. In such cases, techniques that weight the events with their probabilities of being selected, rather than accepting or discarding them straightaway, enable a better usage of the simulated data. The precision of these techniques relies on an accurate parametrization of the dependency of the jet flavour-tagging efficiency function over the conditions under which the jets are produced. In this thesis, an innovative technique to parameterize this function using GNNs is described, along with its implementation in the context of the measurement of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ process, which marks its inaugural application.

Parallel to these research works, which unfolded over three years, I was active in other aspects of the ATLAS Collaboration activities, particularly in the context of the jet flavour-tagging group. In this thesis, two algorithms based on neural networks (NNs), that I developed to validate the effects of a change proposed for updating the design of one of the official ATLAS jet flavour-tagging algorithms, are described; along with the reasons that eventually led the group to reject the proposed modification. In addition, I have been a coordinator of the ATLAS offline Data Quality (DQ) jet flavour-tagging activities since 2022. These activities are illustrated as well.

This thesis is structured as follows. A complete mention of my contribution is provided for every not purely pedagogical chapter.

Chapter 1 describes the SM theory, focusing on the Brout-Englert-Higgs mechanism.

Chapter 2 describes the CERN accelerator complex and the ATLAS detector systems and operations, along with a brief introduction to Monte Carlo (MC) generators.

Chapter 3 provides a detailed overview of jet flavour-tagging in ATLAS. Among the contents

of this chapter, I personally developed and used the **SVKine** and **JFKine** algorithms to study the effects of a proposed change to the ATLAS secondary vertexing **SV1** algorithm as described in Sec. 3.3. Additionally, the offline jet flavour-tagging DQ operations described in Sec. 3.6 all refer to runs whose monitoring I personally coordinated, together with another coordinator and with colleagues volunteering for shifts. I maintained the code, regularly discussed the DQ at dedicated weekly meetings, and organized the shift schedule.

Chapter 4 describes the challenges encountered by the standard jet flavour-tagging algorithms when used to identify the flavour of jets produced with high momenta and describes novel algorithms that address these challenges by leveraging information regarding the distribution of clusters of silicon hits measured in the ATLAS ID. All the algorithms described in this chapter were developed by me. As a byproduct, to adapt them to my means, I contributed to developing the frameworks used to train and test the standard algorithms within the jet flavour-tagging group.

Chapter 5 describes the technique to parametrize the jet flavour tagging efficiency functions using GNNs. While adapting this technique for application in the measurement of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ process, I developed the code and contributed to releasing an ATLAS public document detailing the idea and its application and performance within official simulated event samples.

Chapter 6 describes the measurement of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ process. As already anticipated, I was among the main actors in implementing in this context the event weighting technique based on parametrizing the jet flavour-tagging efficiency functions with GNNs. Particularly, I trained the models used to weight the $V + \text{jets}$ events in the kinematic regime in which the Higgs boson decay products can be reconstructed and identified individually. These types of events, together with $t\bar{t}$ events, represent the dominant background of the measurement. I played a leading role in defining the final strategy for applying the method, described in Sec. 6.11. I had an important role in deriving the relative acceptance uncertainties affecting the contribution from $Z + \text{jets}$ events with zero or two leptons in the final state, kick-starting the code and providing preliminary estimations, as described in Sec. 6.9.2. I am among the main editors of the internal documentation that reflects the detailed studies being performed on the physics modelling of the events, and my contributions are reflected in Sec. 6.9. I performed Data/MC closure studies of events with two leptons in the final state. Finally, I studied the jet flavour-tagging efficiencies of the **DL1r** algorithm when evaluated on $V + \text{jets}$ events simulated with different alternative MC generators. This study supported the choice of adopting the nominal GNN model to apply the event weighting technique on both nominal and alternative events when computing the modelling uncertainties.

I defended this thesis on the 16th of May 2024 at the University of Genova. The $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ measurement described in Chapter 6 was later unblinded and published in [2]. Appendix A was added to this thesis after my defense for completeness, to present the final results obtained by the team.

Thousands of scientists worked for over a decade in the ATLAS Collaboration to develop its tools and use them to achieve outstanding scientific results. Their work is inextricably intertwined with the research presented in this thesis, which can be fully understood only when considering all the aspects of the ATLAS experiment. Serving justice to all the ingredients needed to achieve my results would be possible only by writing the equivalent of multiple theses: a deed that exceeds the scope of my own. Nevertheless, a significant part of the work involved in gaining a PhD participating in such a sophisticated experiment hides in understanding its details, to be able to collaborate effectively towards improved performance and new physics results. Describing large parts of the work of the other members of the Collaboration is not only crucial to understanding this thesis but also testifies to such hidden effort.

Chapter 1

The Standard Model of particle physics

The Standard Model of particle physics accounts for three of the four fundamental forces in Nature: the strong, electromagnetic, and weak forces. Gravity, whose effects operate on a distinct energy scale, is currently beyond its scope. Successfully unifying gravity and the other interactions within the same falsifiable theoretical framework, a quest leading into the landscapes beyond the SM, is among the greatest challenges of contemporary physics.

The elementary particles described by the SM are divided into fermions, with spin one-half and thus obeying the Fermi-Dirac statistics, which implies the Pauli exclusion principle; and bosons, with spin 0 or 1, which obey the Bose-Einstein statistics and can thus form condensates.

The SM comprises 12 fermions, divided into three generations, also called *flavours*, with increasing proper mass. Each generation is formed by one electrically charged lepton, namely the e^- , μ^- and τ^- leptons, and one neutrino, namely ν_e , ν_μ and ν_τ , along with an up-type (u, c, t) and a down-type (d, s, b) quark. Up and down-type quarks possess fractional electromagnetic charges of $+2/3$ and $-1/3$, respectively, and also bear an additional colour charge, which is associated to the *strong* force of quantum chromodynamics (QCD). For every fermion, there exists an antifermion counterpart with all charges inverted.

The interactions among fermions are mediated by spin-1 vector bosons. Three of these bosons, namely the $W^\pm$ and the Z , are associated to weak interactions; one, the photon (γ), mediates electromagnetism; and eight, the gluons (g), are associated to the strong force. Lastly, the Higgs boson (H), a spin-0 particle, is connected to the spontaneous symmetry-breaking mechanism that provides masses to the other elementary particles¹.

Fig. 1.1 schematizes the particle content of the SM.

¹Intriguingly, some authors advocate that the exchange of Higgs bosons should be accounted for as a fifth fundamental interaction [1]. Whether it should be is purely a matter of conventions.

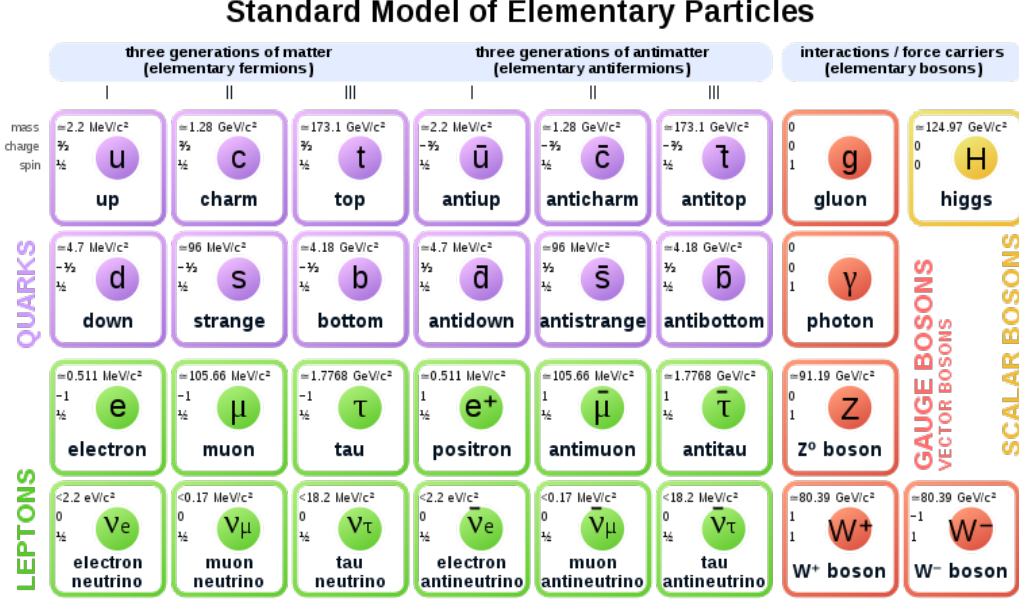


Figure 1.1: The particle content of the SM of particle physics. From [3].

1.1 The SM Lagrangian

The SM is a Quantum Field Theory (QFT) [4]. Its fundamental ingredients are 4-dimensional fields, whose local quantized excitations are the elementary particles. The dynamics of the fields comprising the SM can be derived from the Euler-Lagrange equations together with the Lagrangian density

$$\mathcal{L}_{\text{SM}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\bar{\psi}\not{D}\psi + (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) + \bar{\psi}_i y_{ij} \psi_j \phi + \text{h.c.} \quad (1.1)$$

$\mathcal{L}_{\text{SM}}$ is invariant under local gauge transformations of the $\text{SU}(3)_c \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y$ symmetry group and holds a global Poincaré symmetry, which guarantees that the SM describes physics coherently in every reference frame.

Each term of Eq. 1.1 is described in the following paragraphs.

Quantum electrodynamics Quantum electrodynamics (QED) describes the interactions between photons and the electrically charged fermions [5–7]. Its Lagrangian density can be derived by enforcing local gauge $\text{U}(1)$ symmetry on the Dirac Lagrangian density

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi; \quad (1.2)$$

where ψ represents the fermionic spinor, μ is the Lorentz index ranging from 0 to 3, ∂_μ denotes the Lorentz-covariant derivative, and γ^μ are the Dirac matrices².

$\mathcal{L}_{\text{Dirac}}$ is invariant under U(1) *global* gauge transformation, $\psi \rightarrow \psi' = e^{-ie\alpha}\psi$, where e is a transformation parameter. However, it is not invariant under U(1) *local* gauge transformations, *i.e.*, transformations depending on spacetime coordinates $x = (x^0, x^1, x^2, t)$: $\psi \rightarrow \psi' = e^{-ie\alpha(x)}\psi$. To ensure local gauge invariance, a gauge-covariant derivative is introduced to substitute ∂_μ in Eq. 1.2:

$$D_\mu = \partial_\mu + ieA_\mu, \quad (1.3)$$

where A_μ is an auxiliary spin-1 field which transforms as $A_\mu \rightarrow A_\mu - \partial_\mu\alpha(x)$. Introducing D_μ induces an interaction term between fermions and the auxiliary field A_μ . The latter can be identified with the electromagnetic field, with e representing the electric charge.

This procedure, which consists in enforcing local gauge invariance to a Lagrangian density by means of a gauge-covariant derivative, and observing the consequent interaction between the fermionic and gauge bosonic field, is at the core of the historical construction of the SM.

The kinetic term of the electromagnetic field is included by defining the electromagnetic tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and incorporating it in the QED Lagrangian density:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}i\partial\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - ie\bar{\psi}\not{A}\psi, \quad (1.4)$$

where $\not{\partial} := \gamma^\mu\partial_\mu$, and the same definition holds for $\not{A}$.

Quantum Chromodynamics Quantum Chromodynamics [8–10] describes quarks, gluons and their interactions. Quarks and gluons carry colour charge, which is of the red, green, or blue kind. The QCD Lagrangian density is constructed to be invariant under local SU(3) gauge transformations. This requires the inclusion of the gluon self-interacting gauge bosonic fields. Each of the three up-type and three down-type quark fields $q_k(x)$ of the SM transforms under the fundamental triplet representation of SU(3), where each component of the triplet corresponds to the colour quantum number for the red, green, and blue charges. The gluon field strength tensors $G_a^{\mu\nu}$ are defined for each of the eight Gell-Mann matrices λ_a that generate SU(3) as:

$$G_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu - g_s f_{abc} A_b^\mu A_c^\nu, \quad (1.5)$$

where A_a^μ represents the gluon fields, g_s is the strong coupling constant, and f_{abc} are the structure constants of SU(3). The gauge-covariant derivative is expressed as $D_\mu = \partial_\mu + ig_s \frac{\lambda_a}{2} A_a^\mu$.

²The $m\bar{\psi}\psi$ mass term is not invariant under the gauge transformations of the electroweak theory described in later paragraphs. Therefore, it is not part of $\mathcal{L}_{\text{SM}}$, where the invariant mass term for fermions is introduced through the Brout-Englert-Higgs mechanism. Hence, $m\bar{\psi}\psi$ will be omitted in the following of this paragraph.

The full QCD Lagrangian is then given by:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a + q_k i \not{D} q_k \quad (1.6)$$

The Lagrangian incorporates cubic and quartic terms of the gauge fields A_a^μ , which describe the self-interaction of the gluons.

Due to the presence of off-shell virtual particles in QFT, coupling constants in the SM are renormalized, and vary with the energy scale. The *running* of g_s is particularly relevant from the standpoint of the research presented in this thesis, as it explains the emergence of hadronic jets and allows the interpretation of the high-energy proton-proton scattering events at a hadronic collider like the LHC in terms of the protons' constituents. At lower energy scales (and hence at larger distance scales), the interaction is strong, leading to *colour confinement*, *i.e.*, the observation that any attempt to isolate colour charges involves so much energy that an additional quark-antiquark pair is produced. At higher energy scales (and hence smaller distance scales), *asymptotic freedom* prevails, with weaker strong interactions allowing perturbative calculations. Hadrons, the bound states of quarks, are colourless, meaning they are invariant under local SU(3) gauge transformations.

Electroweak interactions The weak and electromagnetic interactions are unified in the Glashow-Weinberg-Salam (GWS) model [11–13], whose Lagrangian density is invariant under local SU(2)_L ⊗ U(1)_Y gauge transformations. SU(2)_L is known as weak isospin, and U(1)_Y as weak hypercharge.

Each of the four generators associated with the SU(2)_L ⊗ U(1)_Y symmetry group, namely $T_a = \sigma_a/2$ for SU(2)_L (where σ_a symbolizes the Pauli spin matrices) and the scalar $Y = 1/2$ for U(1)_Y, corresponds to a vector gauge bosonic field: $W_{1,2,3}$ and B . The physically observable fields, *i.e.*, the $W^\pm$, the Z bosons and the photon, are linear superpositions of $W_{1,2,3}$ and B :

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2), \quad (1.7)$$

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \quad (1.8)$$

where θ_W is the Weinberg angle, also named the weak-mixing angle. θ_W relates the electromagnetic and weak coupling constants through

$$e = g \sin \theta_W = g' \cos \theta_W. \quad (1.9)$$

The specific linear combinations of the fields associated with the generators of the SU(2)_L ⊗ U(1)_Y symmetry group, that provide the observed physical fields, arise naturally in the context of electroweak spontaneous symmetry breaking (SSB). Electroweak SSB is the process by means of which the $W^\pm$ and the Z bosons, the weak force mediators, gain a proper mass. The mass of the weak interaction carriers explains its short range.

The electroweak charge operator Q can be expressed as a combination of T_3 and Y :

$$Q = T_3 + Y. \quad (1.10)$$

The weak force violates parity conservation, meaning weak interactions do not remain invariant under parity transformations. Only left-handed fermions participate in weak interactions. As neutrinos interact solely through the weak force, there are no right-handed neutrinos in the SM³.

1.1.1 The Brout-Englert-Higgs mechanism

The Brout-Englert-Higgs mechanism [14,15] is the SSB process that provides the mass terms of the electroweak gauge fields without breaking the SM local gauge invariance. It requires the introduction of an additional complex scalar field, denoted as ϕ , which transforms as a doublet under $SU(2)_L$ gauge transformations and has a hypercharge $Y = \frac{1}{2}$. Introducing ϕ additionally enables writing a symmetry conserving term in $\mathcal{L}_{\text{SM}}$, called $\mathcal{L}_{\text{Yukawa}}$, that incorporates the mass terms of all the SM fermion fields except the neutrinos.

The Lagrangian density of the Higgs sector is written as

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi) - \mathcal{L}_{\text{Yukawa}}, \quad (1.11)$$

where ϕ is expressed as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (1.12)$$

with $\phi_{1,2,3,4}$ being real scalar fields; and where the covariant derivative D_μ is the one associated with the $SU(2)_L \otimes U(1)_Y$ local gauge transformations:

$$D_\mu = \partial_\mu + igW_a^\mu T^a + ig'B_\mu. \quad (1.13)$$

$V(\phi)$, which is invariant under $SU(2)_L \otimes U(1)_Y$ local gauge transformations of the field, has a shape with crucial features for the SSB process. It is given by

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (1.14)$$

with $-\mu^2$ and λ being constants. A simplified version of $V(\phi)$, for the case in which ϕ is a complex scalar singlet rather than a doublet, is displayed in Fig. 1.2⁴.

To ensure an energetically bounded theory, *i.e.*, a theory whose minimum energy is finite, λ must be positive. If $-\mu^2$ is greater than 0, the quadratic term is interpreted as a mass

³A very natural minimal extension of the SM, as of today not corroborated by experiments, includes right-handed neutrinos and describes a process, the *seesaw* mechanism, by means of which left-handed neutrinos can get their exceedingly small masses (*e.g.*, $\nu_e < 2.2$ eV).

⁴In such a simplified scenario, $V(\phi)$ exhibits a $U(1)$ symmetry for phase rotations around the origin.

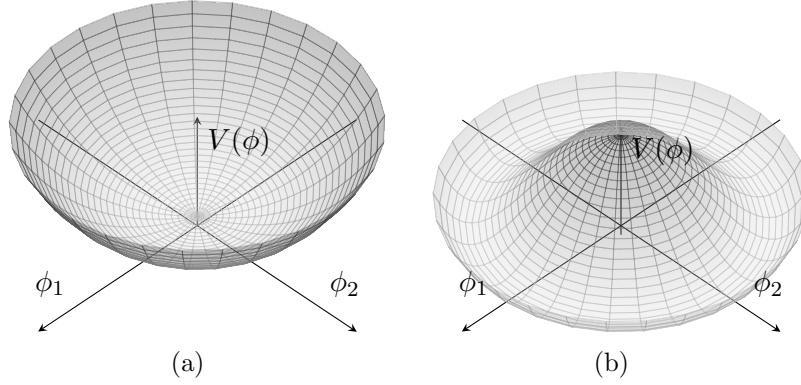


Figure 1.2: Higgs potential for a complex scalar singlet $\phi = \phi_1 + i\phi_2$, with $-\mu^2 > 0$ (a) and with $-\mu^2 < 0$ (b). In the first scenario, the minimum of the potential is in 0, and excitations around the minimum perceive the same symmetry held by the potential as a whole. In the second scenario, there is a continuum of equivalent minima, in the locus of field values where $|\phi| = \frac{\mu^2}{2\lambda} = v^2 > 0$. The surroundings of any minimum do not have the same symmetry of the potential as a whole, as moving in the radial direction yields different results than moving in that tangential to the valley of minima.

term for the scalar field, and $V(\phi)$ has its unique minimum in $\phi = 0$. In such a case, the local surroundings of the minimum maintain a $SU(2)_L \otimes U(1)_Y$ symmetry. In the simplified scenario of Fig. 1.2a, this means that exciting ϕ in any direction in the (ϕ_1, ϕ_2) plane yields identical results.

If $-\mu^2$ is smaller than 0, instead, the field is unphysical due to its negative mass. To obtain a physical interpretation of the Lagrangian density, the field is modelled in terms of perturbations around the minimum. In this case, $V(\phi)$ has a *sombrero* shape with a degenerate continuum of minima and hence of vacuum states. Albeit the locus of field values minimizing the potential is invariant under $SU(2)_L \otimes U(1)_Y$ gauge transformations, the symmetry is hidden in the proximity of any specific minimum value *spontaneously* chosen by the field. In the simplified scenario of Fig. 1.2b, this means that exciting the field in the radial direction from any equivalent minimum yields different results than moving in the direction tangential to the locus of minima. SSB always occurs in Nature when a set of equally favourable states respect the symmetries of the system, but perturbations around any realized minimum do not.

Any equivalent minimum of the sombrero potential is obtained for a non-zero field value ϕ_0 , which yields a non-null vacuum expectation value (vev) $v = \langle 0 | \phi_0 | 0 \rangle$ given by $\frac{\mu^2}{2\lambda} = v^2 > 0$. The vev value is determined by the relation

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}, \quad (1.15)$$

where G_F is the Fermi constant and m_W is the W boson mass. Such relation is derived from

the computation of the muon lifetime in the Fermi and SM theories, and yields $v = 246$ GeV.

Since all the different minima are equivalent, a gauge transformation allows choosing the minimum lying along the real numbers of the second component of the doublet. Applying Eq. 1.10 shows that this choice is invariant under the charge operator, indicating the persistence of an unbroken symmetry subgroup $U(1)_Q$ under which the ground state is invariant.

Choosing the unitary gauge⁵, excitations of the field around the minimum are expressed by:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.16)$$

where H is the famous Higgs field. Substituting this into Eq. 1.11, the linear combinations of the fields corresponding to the three massive electroweak gauge bosons, indicated in Eq. 1.7 and Eq. 1.8, are found to be associated to three mass terms; and the photon remains massless. These mass terms, together with those describing the interactions of the electroweak gauge bosons with the Higgs field, are:

$$\begin{aligned} \mathcal{L} = & \frac{g^2 v^2}{4} W_\mu^+ W^{-\mu} + \frac{g^2 v}{2} H W_\mu^+ W^{-\mu} + \frac{g^2}{4} H H W_\mu^+ W^{-\mu} + \\ & + \frac{(g^2 + g'^2) v^2}{8} Z_\mu Z^\mu + \frac{(g^2 + g'^2) v}{4} H Z_\mu Z^\mu + \frac{(g^2 + g'^2)}{8} H H Z_\mu Z^\mu, \end{aligned} \quad (1.17)$$

where

$$m_W^2 = \frac{g^2 v^2}{4} \quad m_Z^2 = \frac{(g^2 + g'^2) v^2}{4}. \quad (1.18)$$

The coupling between the electroweak and Higgs bosons is proportional to $y_V = \frac{2m_V^2}{v}$, and the Higgs boson mass is $m_H = v\sqrt{\lambda} + \mu$.

This is the Brout-Englert-Higgs mechanism. It maintains renormalizability and unitarity of the SM, while enabling the weak vector bosons to acquire their mass, a property confirmed by many direct and indirect observations. In summary, introducing an unphysical complex scalar field ϕ with a non-zero vacuum expectation value results in spontaneous symmetry breaking. The massless Nambu-Goldstone modes associated to the three generators of the broken $SU(2)_L$ symmetry, emerging during expansion around the true vacuum state, are cancelled out by means of a local gauge transformation to the unitary gauge. In such gauge, their degrees of freedom are eaten away by the vector bosons, allowing them to acquire the

⁵The unitary gauge explicitly removes the three unphysical Nambu-Goldstone modes, which are related to the broken symmetries. In this gauge, the removed degrees of freedom corresponding to the Nambu-Goldstone modes reappear as additional polarization modes of the gauge boson fields, which allow them to acquire mass. The Nambu-Goldstone modes are *eaten away* by the massive gauge boson fields.

additional polarization required to gain mass ⁶.

Within this sector of the SM, there are four fundamental parameters that need to be determined experimentally. These parameters can be expressed either as Lagrangian density parameters (g, g', v , and λ) or as physical observables ($m_Z, \sin \theta_W, m_H$, and e). In the surroundings of the real minimum, spontaneously chosen by the field, the system's symmetry is not manifested, leading to the excitations not adhering to the original symmetry. However, information about the symmetry is preserved through additional constraints on the theory's parameters. Before symmetry breaking, the potential in the system comprises two terms and two constants. After symmetry breaking, the potential comprises three terms, but only two constants persist, usually chosen to be m_H and v . This feature is the remnant of the symmetry of the potential, still perceived in the valley where it would otherwise be hidden.

1.1.2 Fermion masses: the Yukawa term

Fermion mass terms such as that appearing in the Dirac Lagrangian density (Eq. 1.2) violate the invariance of the theory under $SU(2)_L$ gauge transformations. Symmetry conserving mass terms for the fermions arise from Yukawa coupling terms⁷ between their fields and ϕ when the system undergoes SSB. Fermion masses are thus determined by their couplings with the Higgs field and by the Higgs field vev, which sets the fundamental mass scale of the SM.

The renormalizable and gauge-invariant coupling between a fermionic field ψ and ϕ is expressed as:

$$\mathcal{L}_{\text{Yukawa}} = -y_f(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi^\dagger\psi_L), \quad (1.19)$$

where y_f is the Higgs-fermion Yukawa coupling for the fermion field labelled with f .

In the case of the leptons, and considering only the first generation as an example, $\psi_L = (\nu_L, e_L)$ and $\psi_R = e_R$. After spontaneous symmetry breaking, the excitations of ϕ around its minimum are expressed in terms of a vev and of the Higgs field H (Eq. 1.16). Substituting the expression into $\mathcal{L}_{\text{Yukawa}}$ yields:

$$\mathcal{L}_{\text{Yukawa}} = -\frac{vy_e}{\sqrt{2}}\bar{e}e - \frac{y_e}{\sqrt{2}}\bar{e}He. \quad (1.20)$$

A mass term for the electron appears, with $m_e = \frac{vy_e}{\sqrt{2}}$. Neutrino masses cannot be addressed in the same way due to the absence of the right-handed neutrino ν_R ⁸. The second term of $\mathcal{L}_{\text{Yukawa}}$ describes a Yukawa interaction between the electron and the Higgs field.

⁶Other gauge choices, such as the R_ξ gauge, greatly simplify computations beyond the tree level, at the expense of providing an arguably less straight-forward interpretation of the vector boson mass terms. The physics predictions, by construction, do not depend on the gauge fixing.

⁷These are the only interaction terms in the Lagrangian density of the SM that are not linked to a gauge-covariant derivative.

⁸This follows from observing that $\bar{\nu}\nu = \bar{\nu}_L\nu_R + \bar{\nu}_R\nu_L$.

Down-type quarks acquire mass analogously to leptons. Conceptually, up-type quarks gain mass with the same process, but their Yukawa coupling is with $\tilde{\phi}$, the conjugate field to ϕ :

$$\tilde{\phi} = i\sigma_2\phi^\dagger = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1^\dagger \\ \phi_2^\dagger \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} v + H \\ 0 \end{bmatrix} \quad (1.21)$$

The Yukawa terms mix quarks from distinct generations. Physical particles propagate in their mass eigenstates q , which diagonalize the mass matrix. However, their electroweak interactions are diagonal in the weak eigenstates q' , which are linear superpositions of the mass eigenstates. The Cabibbo-Kobayashi-Maskawa (CKM) matrix quantifies this mixing, which determines the strength of flavor-changing weak currents. It is defined as:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \quad (1.22)$$

where $|V_{pq}|^2$ determine the probability of transitioning between the quark states p and q .

1.2 SM parameter counting

The SM is defined by 19 free parameters, which must be determined through experimental means and cannot be derived from fundamental principles. A possible breakdown of these 19 parameters includes:

- The 3 coupling constants governing the interactions of the strong, weak, and electromagnetic interactions: g_s , g , and g' .
- The 9 Yukawa couplings of the Higgs boson with the fermion mass eigenstates.
- 3 mixing angles and 1 CP-violating phase which fully determine the CKM matrix.
- 2 free parameters characterizing the Higgs potential: μ and λ .
- The QCD vacuum angle denoted as θ_{QCD} . This parameter has not been addressed in this chapter. It refers to a maximally CP-violating term that could be added to the QCD Lagrangian density without breaking its symmetry. Nevertheless, experimental data on the neutron electric dipole moment constrain this parameter to be exceedingly close to zero ($\theta < 10^{-13}$). Such remarkable fine-tuning is the core of the long-standing puzzle named *strong-CP problem*.

Different parameter sets can be chosen. For example, rather than the 9 Yukawa couplings, it is possible to choose the experimentally better-measured masses of fermions. Likewise,

the three coupling constants can be redefined in terms of the electromagnetic coupling $e = g \sin \theta_W$, the fine structure constant $\alpha = e^2/(4\pi)$, and the strong coupling constant $\alpha_s = g_s^2/(4\pi)$. Another option is to replace the two Higgs parameters originating from the Higgs potential with the mass of the Higgs boson and the vacuum expectation value of the ground state of the Higgs field.

Chapter 2

The ATLAS experiment at the LHC

2.1 The Large Hadron Collider

The Large Hadron Collider [16], an experimental facility of the Conseil Européen pour la Recherche Nucléaire (CERN), is the largest particle accelerator ever built by humans, boasting a circumference of 27 km. It accelerates protons or heavy ions to energy levels up to the TeV scale, and collides them in four different locations along its ring. Four sophisticated particle detectors, ATLAS [17], CMS [18], LHCb [19] and ALICE [20], are operational in the proximity of these interaction points to collect and measure the particles produced in the collisions and their subsequent decays.

The work presented in this thesis concentrates almost exclusively on proton-proton collision data collected with the ATLAS detector. Before being injected into the LHC, protons undergo a sequence of pre-acceleration stages, outlined in Fig. 2.1. Protons are sourced from a bottle of hydrogen gas, ionized with an electric field. A linear accelerator (initially Linac 2, replaced by Linac 4 in 2020) accelerates them to energies of 50 MeV using radio-frequency cavities. Subsequently, the Proton Synchrotron Booster (PSB), consisting of four synchrotron rings, accelerates the protons to 1.4 GeV, separates the beam into bunches and injects it into the Proton Synchrotron (PS). The PS further increases the beam energy to 25 GeV, followed by the Super Proton Synchrotron (SPS), where the protons reach an energy of 450 GeV. Finally, the beams are injected into the LHC, where two beam pipes are used for circulating the protons in opposite directions. Inside the LHC, protons are further accelerated while superconducting dipole magnets with an 8.33 T magnetic field maintain their circular trajectory. Magnets of higher-order in multipole expansion are used for additional shaping and focusing of the beams. The acceleration continues until two beam energies of 6.8 TeV are achieved. The nominal time-spacing between two subsequent bunches of protons in stable beam conditions is of 25 ns.

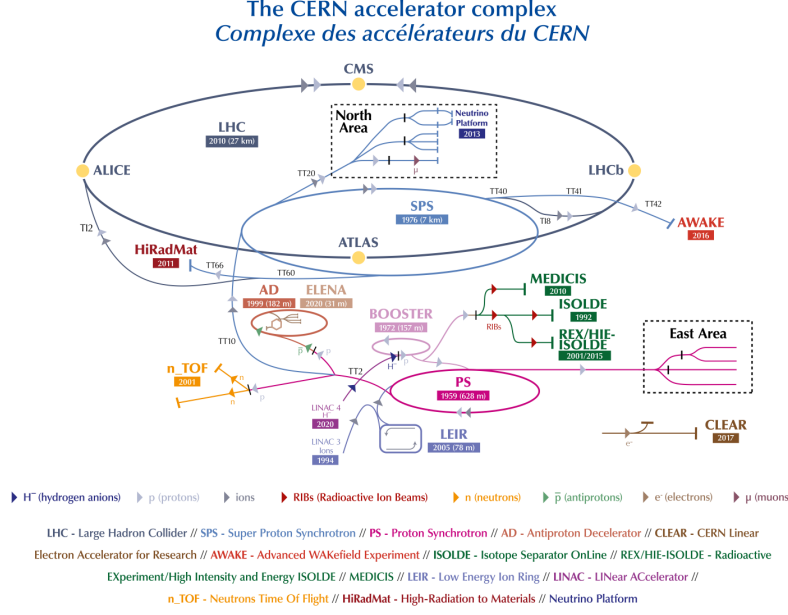


Figure 2.1: The LHC accelerator complex schematized [17].

The two most important quantities to define the operations of a particle collider are the instantaneous luminosity L_{inst} and the energy $\sqrt{s}$ of the colliding particles in their centre-of-mass reference frame. The instantaneous luminosity, which relates the rate dN/dt of a physics process to its cross-section σ as in

$$\frac{dN}{dt} = L_{\text{inst}} \times \sigma, \quad (2.1)$$

can be expressed in terms of the parameters of the collider as follows:

$$L_{\text{inst}} = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y}. \quad (2.2)$$

Here, N_1 and N_2 are the numbers of protons in the two colliding bunches, f is the revolution frequency of the collider, N_b is the number of bunches, and σ_x and σ_y are the widths of the beam density distributions along the x and y direction. The peak instantaneous luminosity delivered by the LHC at the ATLAS interaction point, $L_{\text{inst}} = 2.26 \times 10^{34} \text{cm}^2 \text{s}^{-1}$, was reached in May 2023 during Run-3, the last of the collider data-taking campaigns, which is still ongoing (Fig. 2.2a). The integrated luminosity $L_{\text{int}} = \int L_{\text{inst}} dt$ allows to calculate the total number N of events of a process with cross-section σ produced at a collider:

$$N = L_{\text{int}} \times \sigma. \quad (2.3)$$

Fig. 2.2b displays the integrated luminosity collected by the ATLAS experiment in each year of its operations. During Run-1, the 2011-2012 data taking campaign, the proton-proton

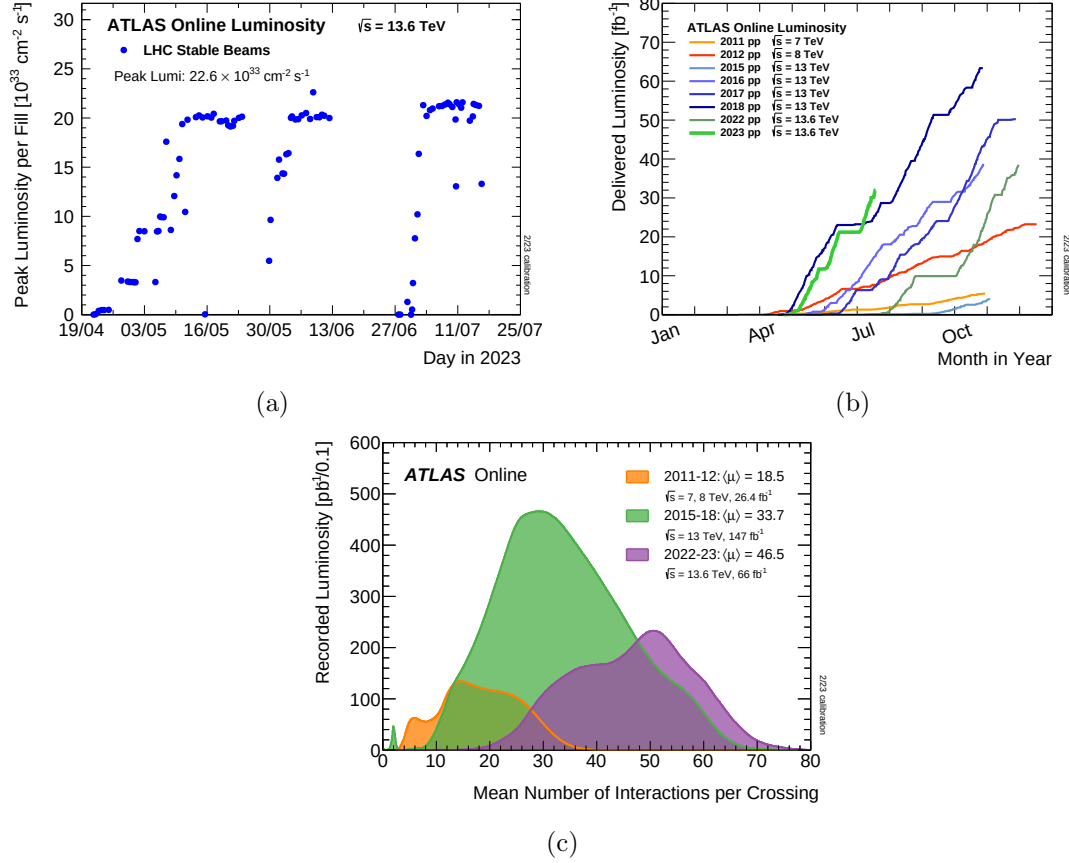


Figure 2.2: From [21]. (a) Per-fill peak instantaneous luminosity of proton-proton collisions observed in the ATLAS detector during the first half of 2023. A *fill* is defined as a single uninterrupted period in which beams circulate in the LHC. (b) Integrated luminosity collected by the ATLAS experiment in each year of its operations. The legend indicates the centre-of-mass energy $\sqrt{s}$ at which the proton-proton collisions occurred year by year. (c) Average number of pile-up interactions per event observed in the ATLAS experiment during the three Runs of the LHC.

collision data was collected at $\sqrt{s} = 7$ TeV and 8 TeV. During Run-2, which covered the 2015-2018 time span, $\sqrt{s} = 13$ TeV was achieved. The Run-2 dataset, with an integrated luminosity of $L_{\text{int}} = 140.1 \pm 1.2 \text{ fb}^{-1}$, is used to perform the calibrations of the jet flavour-tagging algorithms mentioned in Chap. 4 and the measurement of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ process detailed in Chap. 6. The highest energy ever achieved at a collider on Earth is $\sqrt{s} = 13.6$ TeV. This record is set at the Run-3 campaign, which started in 2022. The data quality activities described in Chap. 4 focus on Run-3 collisions.

In the ATLAS experiment, research mostly focuses on head-on collisions between two protons within a bunch-crossing, leading to hard scattering events between their partons. However, each hard scatter event is accompanied by additional soft interactions known as pile-up, occurring simultaneously. In-time pile-up involves soft interactions in the same bunch-crossing, while out-of-time pile-up comes from neighbouring bunch-crossings. Pile-up is quantified by the average number of interactions per event $\langle\mu\rangle$, defined as

$$\langle\mu\rangle = \frac{L_{\text{bunch}}\sigma_{\text{inel}}}{N_b f}, \quad (2.4)$$

where L_{bunch} is the average instantaneous luminosity per bunch crossing, and σ_{inel} is the inelastic scattering cross-section for proton-proton collisions. The distribution of μ observed in the ATLAS detector is showcased in Fig. 2.2c for the three Runs of the LHC. These distributions are later used to apply a per-year correction on the simulations to agree with data.

2.2 ATLAS: A Toroidal LHC ApparatuS

The ATLAS detector [17] is a multi-purpose particle detector, designed to explore a wide range of physics phenomena. With dimensions of 25 meters in diameter, a length of 44 meters, and a weight of 7000 tons, it stands as the largest detector at the LHC. ATLAS has a cylindrical structure comprising multiple layers, providing nearly complete 4π solid angle coverage. Fig. 2.3 schematizes the detector layout. All the detector systems are designed to achieve optimal performance for various physics analyses. The fundamental criteria that guided the design of the detector are fast electronics for efficient readout, high granularity, effective object reconstruction, and resolution.

ATLAS coordinate system ATLAS coordinate system is defined with the z -axis in the beam direction, the x -axis towards the centre of the LHC ring, and the y -axis pointing upwards from the interaction point. Using polar coordinates, the azimuthal angle ϕ spans the xy -plane, and the angle θ is measured upwards from the beam axis. It is convenient to substitute the angle θ with the alternative variable called rapidity, indicated with the y symbol:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (2.5)$$

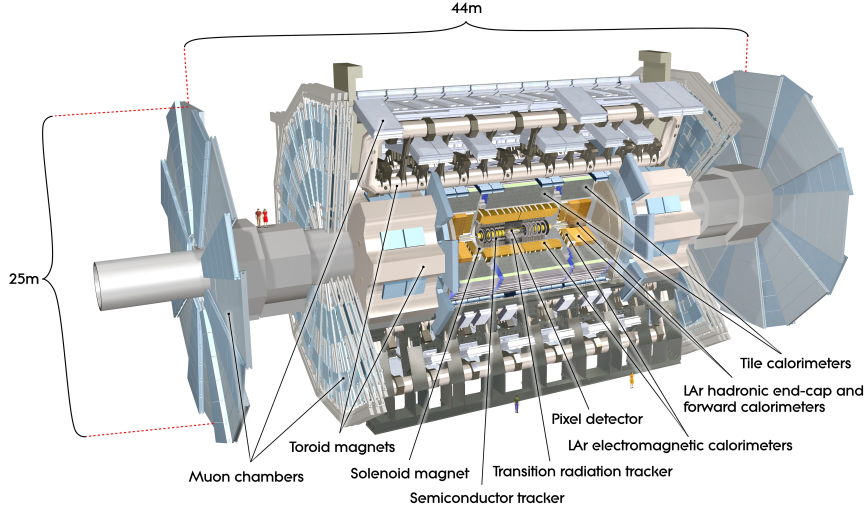


Figure 2.3: Layout of the ATLAS detector with its subsystems [22].

where p_z is the projection of a particle momentum on the z -axis. In proton-proton collisions, high-energy interactions occur between partons with unknown momenta in the beam direction z . Rapidity is advantageous since rapidity differences are invariant under Lorentz boosts along z . For highly energetic particles (*i.e.*, $p \gg m$), pseudo-rapidity $\eta = -\ln[\tan(\theta/2)]$ is a useful approximation of rapidity, that solely relies on geometrical information and is hence unrelated to any specific particle. Another important kinematic variable is ΔR , indicating the angular separation between two objects in the detector based on η and ϕ :

$$\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2} \quad (2.6)$$

Due to the unknown momenta of the partons along the z -axis, energies and momenta of the particles are commonly measured in the transverse xy -plane, defining $p_T = p \sin(\theta)$ and $E_t = E \sin(\theta)$.

2.2.1 Inner Detector

ATLAS innermost subsystem is the Inner Detector, schematized in Fig. 2.4. The ID is used to track charged particles from the interaction point to the calorimeters and measure their momentum by analyzing the track curvature in the solenoid magnetic field. It covers the region with $|\eta| < 2.5$ and consists of three subsystems: the Pixel Detector, the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

The ID momentum resolution σ_{p_T} is given by the formula:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T [\text{GeV}] \oplus 1\% \quad (2.7)$$

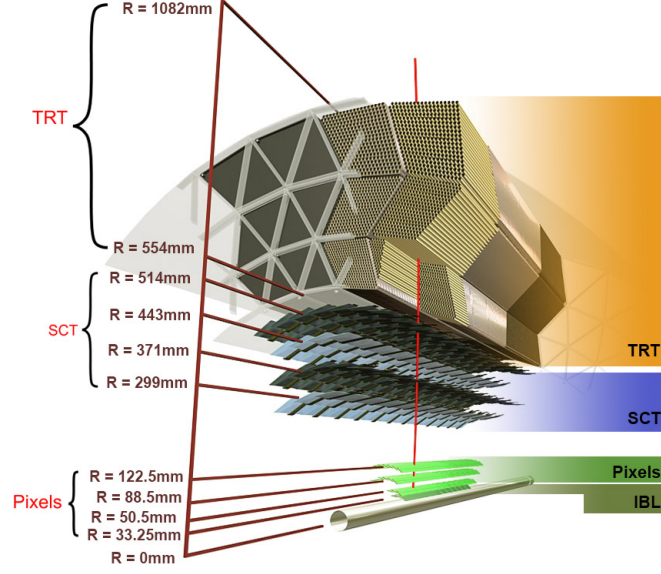


Figure 2.4: Cutaway view of the three subsystems comprising the ATLAS ID [23].

where $\oplus$ denotes the pythagorean sum.

Pixel detector The inner part of the ID is the silicon pixel detector, consisting of 4 cylindrical layers and 2 end-caps with 3 disc layers each. The barrel layers are positioned between 33.25 mm and 122.5 mm around the beam pipe, covering $|\eta| < 2.5$. The pixel detector is crucial for track reconstruction and for primary and secondary vertex identification.

The Insertable B-layer (IBL) is the innermost layer, introduced between Run-1 and Run-2, featuring the highest granularity with a pixel size of $50 \mu\text{m}$ in the $r - \phi$ direction and $250 \mu\text{m}$ in the z -direction. The three remaining layers have a pixel size of $50 \mu\text{m}$ in the $r - \phi$ direction and $400 \mu\text{m}$ in the z -direction. This results in an expected hit resolution of $8 \mu\text{m}$ and $10 \mu\text{m}$ in the $r - \phi$ direction and $40 \mu\text{m}$ and $115 \mu\text{m}$ in the z -direction for the IBL and the three remaining layers, respectively. Charged particle leaves an average of 4 hits in the pixel detector.

The IBL plays a vital role in jet flavour-tagging, which is the focus of Chap. 3 and 4. The impact parameter (IP) resolution of tracks with low- p_T ¹, essential for jet flavour-tagging, improves significantly with the inclusion of the IBL, due to its smaller pixel size and because it provides an additional hit point close to the beam 2.5. All jet flavour-tagging studies in this thesis cover Run-2 and Run-3 and include the IBL.

The hits measured in the pixel detector, together with those measured in the SCT, provide the additional information that, used alongside with tracks, can be leveraged to improve

¹The IP of a track is defined as its closest distance to the primary vertex (Fig. 2.10).

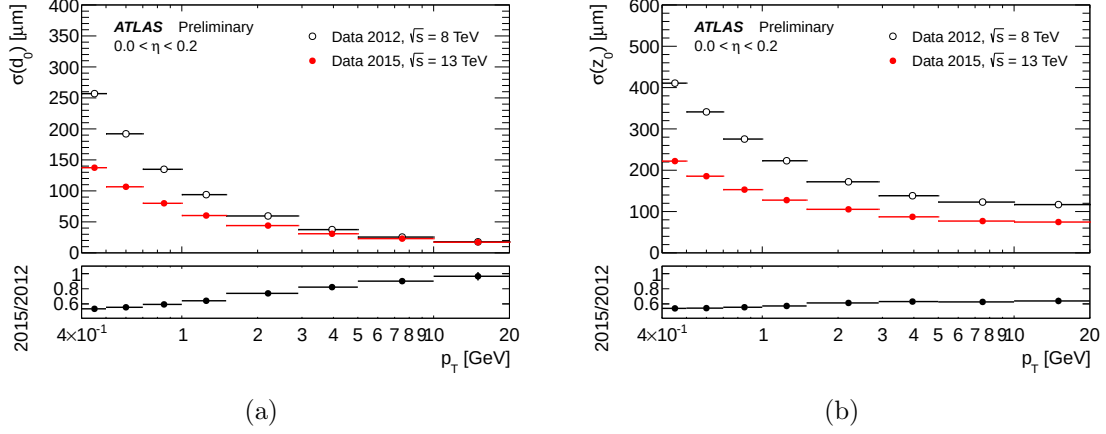


Figure 2.5: Transverse (a) and longitudinal (b) track IP resolution measured from data as a function of p_T in the $0.0 < \eta < 0.2$ region [24]. Run-2 data, collected at a $\sqrt{s} = 13$ TeV centre-of-mass energy with the ID including the IBL, are compared to Run-1 data, with $\sqrt{s} = 8$ TeV and prior to the introduction of the IBL. The IP resolution has been unfolded to remove the contribution from the vertex resolution.

jet flavour-tagging performance at high- p_T by means of the GN2+Clus algorithm detailed in Chap. 4.

Semiconductor Tracker The SCT uses silicon microstrip modules with an average strip pitch of $80 \mu\text{m}$. Each module has two sensor layers tilted at a small stereo angle of 40 mrad , enabling 2D hit reconstruction. The barrel region has four cylindrical layers of strip modules, and each end-cap consists of nine discs of modules. The SCT’s intrinsic accuracy is $17 \times 580 \mu\text{m}^2$ in the barrel region and $580 \times 17 \mu\text{m}^2$ in the end-caps, measured in $r - \phi \times z$ coordinates. On average, charged particles create 8 hits in the SCT.

Transition Radiation Tracker The TRT is a straw-tube tracker comprising cylindrical drift tubes (straws) with a 4 mm diameter. Each straw contains a gold-plated anode wire and is filled with a Xenon-based gas mixture. As a charged particle passes through the tube, it ionizes the gas. The resulting electrons move towards the anode wire, creating a detectable signal that can be compared against two adjustable thresholds. The TRT provides $r - \phi$ hit information in the barrel region and $z - \phi$ information in the end-caps with an intrinsic accuracy of $130 \mu\text{m}$ per straw via drift time measurements. On average, a charged particle leaves 36 hits in the TRT.

The TRT serves not only for measuring the momentum of charged particles but also for particle identification. The spaces between the straws is filled with polymer fibres (foils) in the barrel (endcaps). When a relativistic particle crosses the boundary between the two

materials, it can emit radiation known as transition radiation, dependent on the Lorentz boost $\gamma \sim E/m$. This allows determining the particle mass for particles with the same momentum. In the ATLAS detector, it aids in distinguishing between electrons and pions with momenta ranging from 1 to 100 GeV². The drift tubes measure the transition radiation photons through the photoelectrons they generate, identified as distinct large energy deposits.

2.2.2 Calorimeters

The calorimeters are used to measure the energy of neutral and charged particles, as well as missing transverse momentum. Hermetic in ϕ and covering the $|\eta| < 4.9$ range, the ATLAS calorimeters are of the sampling type, consisting of multiple layers of active and passive materials. The passive material, typically a dense substance with a high atomic number, induces particle showers. Active layers then measure the energy deposited by these showers, detecting ionization or scintillation light. Liquid Argon (LAr) serves as the primary active material in ATLAS [26].

Fig. 2.6 illustrates the layout of the ATLAS calorimeters. The Electromagnetic Calorimeter (ECAL) measures energy deposited by electrons, photons, and some jet energy. It utilizes lead as the passive material and liquid argon as the active material [27]. Electromagnetic showers, governed by the electromagnetic interaction, are characterized by the radiation length X_0 , representing the average distance over which the energy of a high-energy electron reduces to $1/e$ of its original value. The ATLAS electromagnetic calorimeters have a depth of at least $22 X_0$.

Hadronic calorimeters are employed to measure the energy of hadronic showers. The hadronic endcap and forward calorimeters use LAr as the active material, with copper plates in the endcap serving as passive material [28]. The forward calorimeter employs both copper and tungsten as passive materials. In the barrel, the tile calorimeter uses steel as the absorbing material and scintillating plastic tiles as the active material [29]. Hadronic showers in the detector are characterized by the nuclear interaction length λ , representing the mean distance traveled by a hadron before interacting with a nucleus. The depth of the ATLAS hadronic calorimeters is approximately 10λ . Consequently, both the electromagnetic and hadronic calorimeters effectively contain particle showers.

The energy resolution of a calorimeter is expressed as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c. \quad (2.8)$$

The first term is the stochastic term which captures fluctuations associated with Poisson statistics, such as shower or sampling fluctuations. The second term is the noise term,

²During Run-2, sections of the TRT had to be operated with an Argon-based gas mixture, less expensive than Xenon, due to irreparable gas leaks. This decision came at the expense of losing these sections' particle identification capability since Argon does not adequately absorb transition radiation photons [25].

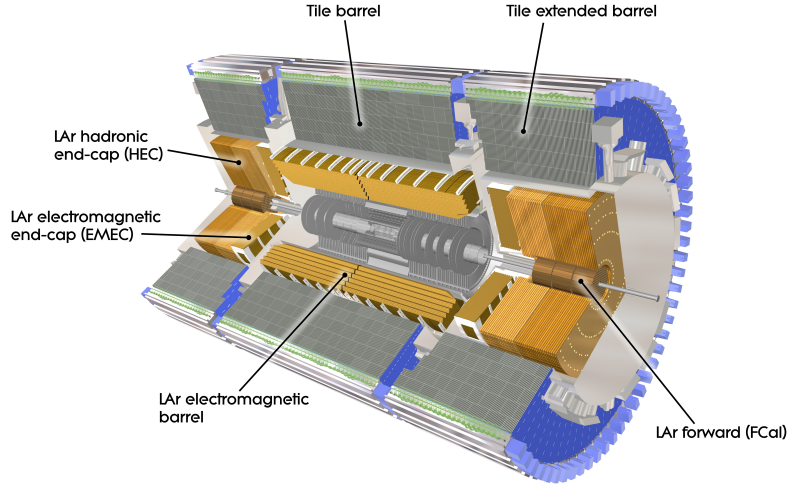


Figure 2.6: Cutaway view of the ATLAS Calorimeter system [30].

accounting for electronic noise in the detector readout. The constant term describes instrumental effects, including inefficiencies in signal collection. For the electromagnetic calorimeter, the nominal energy resolution is approximately $\frac{\sigma_E}{E} \approx 10\%\sqrt{E} + 0.7\%$ [31]. In the case of the hadronic calorimeter, the energy resolution is poorer due to the complexity of hadronic showers: $\frac{\sigma_E}{E} \approx 50\%\sqrt{E} + 3\%$ [31].

2.2.3 Muon Spectrometer

The Muon Spectrometer (MS) is the outermost subsystem in the ATLAS experiment. Its primary function is to measure the momentum of muons, which constitute the majority of charged particles traversing the detector after passing through the hadronic calorimeter. Muon momenta are measured through the curvature of their trajectories in the toroidal magnet system. Precision in measured muon momentum ranges from 3% for low p_T muons ($2 < p_T < 250 \text{ GeV}$) to 10% for high p_T muons. Different detectors within the muon system perform precision tracking and provide fast triggering information.

Precision Tracking Chambers The Monitored Drift Tube (MDT) detector system is used for precision tracking [32]. It spans both the barrel and endcap regions up to $\eta < 2.7$ in multiple layers, providing a spatial resolution of $35 \mu\text{m}$. Additionally, the high occupancy region from $2.0 < \eta < 2.7$ is covered by Cathode Strip Chambers (CSC) [33], which are multi-wire proportional chambers with segmented cathodes. The CSCs can achieve a spatial resolution of $60 \mu\text{m}$ in the r direction and 5 mm in the ϕ direction.

Fast Trigger Chambers The muon trigger system incorporates two types of detectors. In the barrel region, three layers of Resistive Plate Chambers (RPCs) [34] are used, while the endcap features 3-4 layers of Thin Gap Chambers [35]. The combination of these two systems enables a muon-based hardware trigger through the coincidence of hits in different layers, along with an additional measurement of the muon trajectory.

2.2.4 Magnets

In addition to the detector systems, the magnet system plays a crucial role in enabling the measurement of momenta and charges of particles. The magnets bend the trajectory of charged particles through the Lorentz force, depending on their momentum and charge. The ATLAS detector relies on two magnet systems. The central solenoid magnet, positioned between the ID and the calorimeters, produces a constant magnetic field of 2 T along the z direction. This superconducting magnet, composed of NbTi, is cooled using liquid helium to a temperature of 1.8 K. Secondly, there is the extensive toroidal magnet system integrated into the MS, comprising one barrel toroid and two end-cap toroids, each with eight coils. The toroidal magnets generate an inhomogeneous magnetic field, approximately 0.5 T in the central region and 1 T in the end-cap regions, respectively.

2.2.5 Triggering and data acquisition

The frequency with which bunches of protons collide at the LHC is ~ 40 MHz. The data generated in the ATLAS subsystems is too large to be integrally stored on tape. Furthermore, most analyses focus on a subset of events which are of particular interest. Therefore, a multi-layered trigger and data acquisition (TDAQ) system is in place to select and store the events meeting specific criteria. Fig. 2.7 schematizes the ATLAS TDAQ chain.

The trigger system is broadly divided into the level one trigger (L1) and the high level trigger (HLT) [36]. The L1 trigger is a specialized hardware trigger that uses mainly information from the MS and reduced calorimeter information to find signatures of high- p_T objects such as muons, electrons, jets, or missing transverse energy. It reduces the peak event rate to approximately 100 kHz and defines preliminary regions of interest in the detector. These regions are then forwarded to the HLT. During the decision-making process of the L1 trigger, event information is temporarily stored in custom front-end (FE) buffers on the detector before being transmitted to read-out-drivers (ROD) located outside the detector. The HLT, a software-based trigger, operates on a dedicated computing farm and has access to the entire detector information, in addition to the specific regions of interest identified by the L1 trigger. The HLT further decreases the event rate to about 1.5 kHz at full capacity, and the selected events are permanently stored and transmitted to the CERN Tier-0 computing site for offline reconstruction. While the HLT makes decisions, the event information is held in the read-out-system buffer.

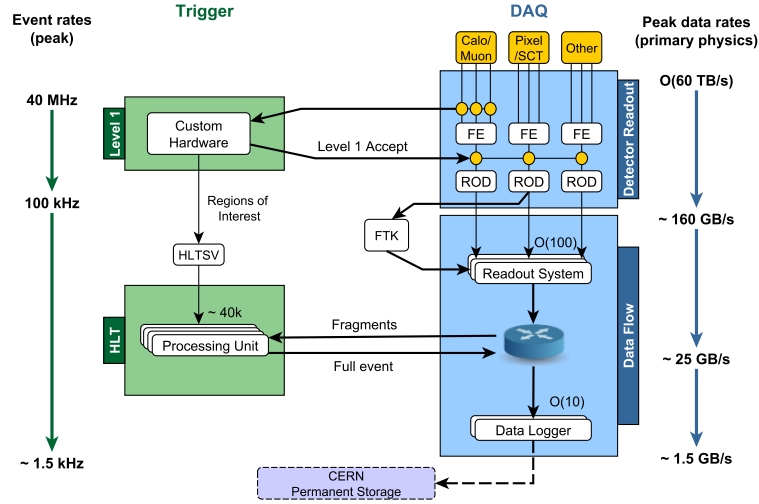


Figure 2.7: Functional diagram of the ATLAS Trigger and Data Acquisition system in Run 2 showing expected peak rates and bandwidths through each component. From [36].

2.3 Reconstruction of detector-level and physics objects

Proton-proton collisions occur at the centre of ATLAS, and the resulting particles traverse through the detector, leaving distinct signatures in various sub-detectors, as depicted in Fig. 2.8. These signatures include hits in different layers of the ID and muon spectrometer, as well as energy deposits in the calorimeter. Signals from the detectors undergo processing and reconstruction, forming what are termed “physics objects.” These objects are then matched to final-state particles or observables within ATLAS, such as electrons, muons, or jets. The reconstruction process consists of multiple steps. Initially, detector-level objects like space-points, tracks, vertices, and calorimeter clusters are reconstructed. These objects are then combined to reconstruct high-level objects corresponding to different particles, possibly applying additional identification and isolation criteria. Finally, the physics objects are calibrated to address any detector effects that may differ between data and simulation.

2.3.1 Detector-level objects

Hits, clusters and space-points Charged particles traversing the ID deposit energy within the pixel layers. This energy excites electrons in the valence band of the silicon semiconductors, causing them to move to the conduction band. By applying electric potentials, the excited electrons can be collected, generating an electric current that serves as a signal indicating energy deposition in the pixel. The signal is read out when the electric

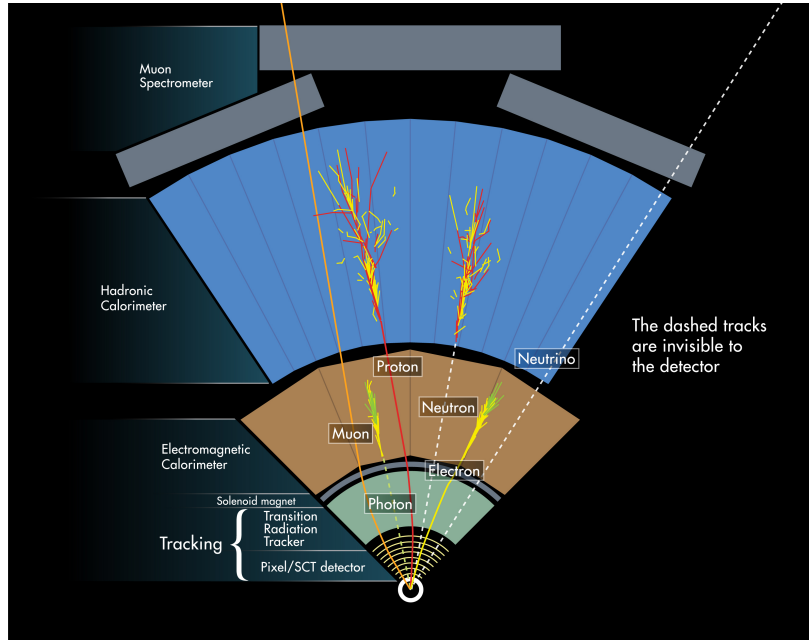


Figure 2.8: Signatures that different particles write in the ATLAS detector [37]

current comprises more electrons than a calibrated threshold which is varied with the aging of the detector [38]. In such cases, the signal is termed *hit*. The hit records 2D spatial information about the (η, ϕ) coordinates of the energy deposition. Its distance from the beamline r is inferred by the ID layer in which it occurs.

As a charged particle moves through a pixel sensor, the charge is typically collected in more than one pixel. Neighbouring activated pixels are grouped into *clusters* through a clustering algorithm. Clusters are categorized as *single-particle* when they result from the charge deposited by one particle, or *merged* when they are formed by multiple particles. When used to reconstruct tracks, additional categorizations are introduced. Clusters are *split* if a splitting algorithm divides them into sub-clusters, each used as an individual measurement for a different track. Alternatively, they are *shared* if they are not split but used by more than one track candidate. Fig. 2.9 schematizes cases of single-particle and split clusters.

From each cluster, the *space point* is computed, representing the point of intersection of the particle with the detector. This estimation involves a charge interpolation technique based on the charge distribution in different hits. Space points are the detector-level objects used to reconstruct tracks.

Tracks and vertices in the ID Charged particles moving through the approximately uniform magnetic field produced by the solenoid in the ID follow a helicoidal trajectory

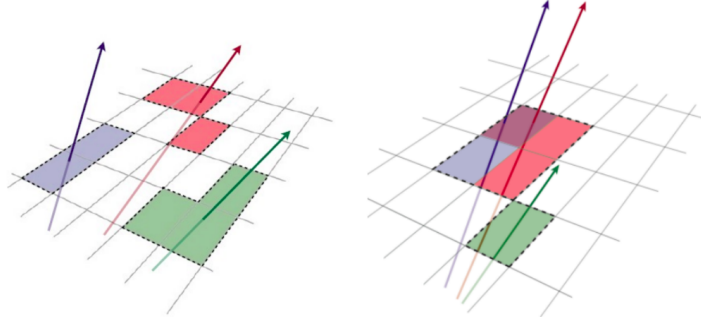


Figure 2.9: Sketch of single-particle (left) and split clusters (right) of silicon hits in the ATLAS ID. From [38].

described by five parameters (Fig. 2.10). One possible parametrization commonly adopted in high-energy physics uses:

- the transverse and longitudinal IPs d_0 and z_0 , *i.e.*, the distance of the point of closest approach of the curve to the z -axis, and its distance to the $x - y$ plane containing the collision point;
- ϕ and θ , *i.e.*, the azimuthal and polar angles at the point of closest approach;
- q/p_T , *i.e.*, the ratio of the reconstructed charge over the transverse momentum of the particle producing the track. This value is connected to the track's curvature through $p_T = qBR$, where B represents the magnetic field and R represents the radius of the osculating circle of the curve.

Tracks are the detector-level objects targeting these helices. Tracks are reconstructed using the three-dimensional space points generated from the clusters in the Pixel and SCT detectors and of the drift circles produced in the TRT. Two algorithms are used. An inside-out algorithm aims at extending the tracks from the pixel layers outwards. Its goal is the efficient reconstruction of charged particles emerging from hard-scatter events or the decay of very short-lived particles. The algorithm begins with a triplet of hits in the Pixel and SCT detectors, and, factoring in information on the material and magnetic field configuration, extrapolates the track from the seeds to the end of the SCT using a combinatorial Kalman filter. In cases of ambiguity, multiple track candidates are considered. An ambiguity solver applies quality criteria to the track candidates, ranking them based on factors such as fit quality or applying penalties for missing hits. In the final step, the surviving tracks are extended to the TRT and combined with its hit measurements. A complementary outside-in algorithm [40] tracks charged particles produced in decay chains of long-lived particles like K_s , photon conversions, or hadronic interactions lacking sufficient hits in the innermost layers. This algorithm initiates from the TRT measurements and progressively adds silicon hits.

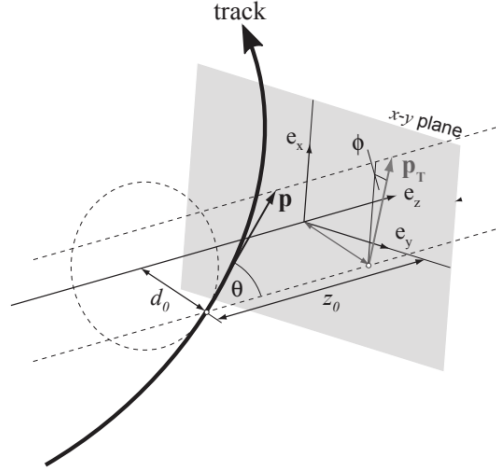


Figure 2.10: Parametrization of a helicoidal trajectory in a uniform magnetic field. From [39].

Tracks used for physics analyses must fall within the acceptance of the ID ($|\eta| < 2.5$), have p_T greater than 0.5 GeV, and satisfy quality-driven selections on the number of silicon detector hits and association with the primary vertex.

Vertices are the detector-level objects describing the locations where interactions occur. They are reconstructed targeting points within the detector from which multiple tracks originate. The Primary Vertex (PV) represents the position of the primary proton-proton interaction per bunch-crossing. Its reconstruction involves an iterative finding and fitting procedure [41, 42]. The key parameters guiding this procedure are the d_0 and z_0 of each track. The process begins with a vertex seed, and tracks are incrementally added if they are compatible with the vertex position. This position is updated iteratively by weighting candidate tracks based on their compatibility with the vertex position. All tracks associated with a given vertex are refit under the assumption that they originate from that vertex. Tracks incompatible with the identified vertex then serve as seeds for another vertex. This procedure repeats until no further vertices can be constructed. Valid PV candidates must be associated with at least two tracks with p_T greater than 0.5 GeV. Among these candidates, the PV is selected as having the maximal $\sum_{\text{tracks}} p_{T_i}^2$.

Topological calorimeter clusters Energy deposits in calorimeter cells estimated to originate from the shower of a single particle caused by its interaction with the calorimeter material are grouped in topological calorimeter clusters [43]. Energy deposits from adjacent cells are reconstructed to create 3D clusters representing the particle showers within the active calorimeter volume, utilizing only significant signals standing out from electronic noise and other fluctuations such as pile-up. The resulting topological clusters can represent a

full or fractional response to a single particle, a merged response to several particles, or a combination of both.

2.3.2 Physics objects

Jets Due to colour confinement, colour charges can never be observed as free states. Instead, they undergo a cascade process called *hadronisation*, which results in a spray of collimated colour-neutral particles. These cascades, which hint at the production of a quark or gluon within the detector, are aggregated in reconstructed objects called *jets*. Jets are by far the most important physics objects from the standpoint of this research.

Jets are reconstructed by clustering constituent objects such as tracks or calorimeter clusters using a jet clustering algorithm. In ATLAS, the anti- k_t algorithm [44] implemented in FASTJET [45] is predominantly used. Objects can be associated to jets using their angular separation ΔR or via ghost association techniques [46] where the object is assigned null momentum and re-clustered into the jet after it is formed.

Different types of jets have been used throughout the history of the ATLAS collaboration. Those used in this thesis are Particle-Flow (PFlow) jets and variable radius track jets. PFlow jets are constructed from particle-flow objects [47] using the anti- k_t algorithm. Particle-flow objects combine information from both the ID and the calorimeters. After applying an energy-subtraction algorithm to remove overlap between momentum and energy measurements made in the ID and the calorimeters, tracks are matched to topological calorimeter clusters. The expected energy in the calorimeter, deposited by the particle that created the track, is computed based on the cluster position and track momentum. A probability is computed that the particle deposited energy in multiple calorimeter clusters, and possibly additional clusters are added to the track+cluster system to recover the full shower energy. Incorporating tracking information in the jets improves the energy resolution at low transverse momenta and the position resolution at high transverse momenta. Moreover, tracking information can be used to mitigate pile-up contamination. PFlow jets clustered using the anti- k_t algorithm with a radius parameter $R = 0.4$ are named *small- R* jets. Those using $R = 1.0$ are named *large- R* jets. Small- R jets aim at clustering the particles stemming from the hadronization of a single colour charge. In this thesis, focusing only on the sections describing my personal contribution, they are used to train the **SVKine** and **JFKine** jet flavour-tagging algorithms detailed in Sec. 3.3; to monitor the quality of the jet flavour-tagging performance on the data collected during Run-3 within the ATLAS Data Quality group, as described in Sec. 3.6; to train the GNN-based jet flavour-tagging algorithms which are the object of Chap. 4, and to reconstruct the $b\bar{b}$ and $c\bar{c}$ pairs produced in the decay of the Higgs-boson candidate in the kinematic region of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ analysis where the two quarks can be resolved, as described in Chap. 6. large- R jets contain the $b\bar{b}$ and $c\bar{c}$ decay products of Higgs-bosons produced with high p_T in the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ analysis, since the pair is too collimated to resolve the individual quarks with small- R jets. They are also used to reconstruct the events comprising the training sample of the GNN-based truth

flavour-tagging algorithm described in Chap. 5.

Variable-radius track-jets are clustered using the anti- k_t algorithm applied on tracks. Track-jets are matched to large- R jets as sub-jets and serve the purpose of identifying large- R jets containing b -hadrons. The radius parameter varies with transverse momentum, allowing for a wider cone (up to $R = 0.4$) for low- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T track-jets [48]. The narrower cone allows to resolve very collimated sub-jets within the large- R jet. Variable-radius track jets are used to reconstruct the large- R jets' substructures in Chap. 5 and 6.

Electrons Electrons are reconstructed through a combination of ID tracks and topological clusters in the ECAL [49]. Electrons are termed *prompt* when they originate from the primary vertex and *non-prompt* when they are produced in decays or subsequent interactions within the detector. Electron signatures can be mimicked, for instance, by misreconstructed photons and pions. Therefore, it is crucial to establish identification and isolation criteria to select real electrons with high efficiency and purity.

Electrons are reconstructed by selecting tracks compatible with the expected energy loss of electrons within the detector. Tracks are matched to energy deposits in the ECAL if their extrapolated trajectories meet the barycenter of a calorimeter cluster. When this happens, tracks are refitted factoring in the energy that is lost to the electrons via bremsstrahlung. After reconstructing the candidates, they are further scrutinized by means of a likelihood-based *identification* algorithm that pins-down prompt electrons and mitigates contamination from non-prompt electrons and electrons stemming from photon conversion. Increasingly stringent working points of the algorithm are defined with different thresholds on the likelihood discriminant. Each working point provides different reconstruction efficiency and background rejection. Typically, the activity surrounding prompt electrons is very low. To further reduce the background, *isolation* criteria can be defined by limiting the sum of the transverse momenta of the ID tracks or the energy of calorimeter clusters in a cone around the electron candidate.

Photons Photons' electromagnetic showers result in energy deposits comparable to the electrons'. However, being electromagnetically neutral, photons do not produce a track originating from the primary vertex to the energy deposits. Due to their similarities, photon reconstruction proceeds in parallel with the electron reconstruction [50]. Photons fall within two categories: converted photons generate an e^+e^- pair within the tracker volume, and the pair then showers in the ECAL. These photons can be identified by detecting a conversion vertex originating from a pair of tracks with opposite charges. In contrast, unconverted photons do not leave a track in the ID. An ambiguity solver is employed to differentiate between electrons and converted and unconverted photons.

Muons Muon reconstruction uses muon tracks in the MS, ID tracks, and characteristic signatures in the calorimeters [51]. MS tracks are formed by connecting straight-line segments, identified via a Hough transform, into an approximately parabolic trajectory. The track is then obtained via a χ^2 fit which factors in the interactions of the muon with the material. Reconstructed muon candidates are *combined* when they are matched to ID tracks. Combined muons are further fitted with combined ID and MS hits, accounting for calorimeter-induced energy loss.

Muon candidates are further scrutinized to identify prompts and reject non-prompt backgrounds such as muons produced in pion, kaon and semi-leptonic heavy-flavour hadron decays. Combined muon identification leverages discrepancies in p_T and charge measurements between the MS and the ID, and the χ^2 of the combined track fit. Selections on ID and MS hit numbers are also applied. As for the electrons, isolation requirements can be added to suppress backgrounds further.

Taus τ leptons are not collider-stable particles³. They decay hadronically via $\tau \rightarrow \nu_\tau +$ hadrons 66% of the times and leptonically through the $\tau \rightarrow l\nu_l\nu_\tau$ ($l = e, \mu$) process in the remaining 34% cases. Leptonic τ decays are identified by reconstructing muons and electrons. Hadronic decays, denoted as τ_{had} , are categorized based on the number of charged particles they produce. Approximately half of the total branching fraction of the τ lepton results in a single charged track (*one-prong* topology), while around 15% of the decays exhibit a *three-prong* final state. Hadronic τ are reconstructed as jets, and identified using a Recurrent Neural Network (RNN) trained to discriminate the one or three-prong final states from the hadronic background. WPs with increasingly stringent identification criteria are defined.

Neutrinos as Missing Transverse Momentum An imbalance in the transverse momentum of the final state may arise due to the incomplete measurement of particles. Most noticeably, neutrinos interact too weakly to be detected at LHC detectors. Hence, they are measured using the events' missing transverse momentum E_T^{miss} , defined as the negative vector sum of the transverse momenta of all photons, leptons, jets, and tracks associated with the PV [52].

2.4 Data Quality monitoring

All the data collected by the ATLAS detector are subject to a thorough Data Quality (DQ) scrutiny, to ensure they meet the tight standards required for usage in scientific analyses.

³*i.e.*, particles whose average decay length is much longer than the detector dimensions when produced at the LHC.

The monitoring is performed by the DQ group, which is responsible for providing the Collaboration with the list of events that can be used for physics analyses.

Before the beginning of Run-3, I was appointed offline jet flavour-tagging DQ coordinator, a role that I share together with a colleague. This section introduces the operations of the DQ groups. Sec. 3.6 is dedicated to the detailed description of the activities of the offline jet flavour-tagging DQ group.

2.4.1 DQ infrastructure and operations

From the moment in which the protons are injected in the LHC beam pipe, data collected by the ATLAS detector are subject to continuous inquiry to extract the maximum amount of information beneficial for the physics programme of the Collaboration.

Typically, ATLAS continuously collect data during a single LHC fill. Each of the datasets taken by ATLAS while it is continuously recording is referred to as *run*. Runs are further divided in *luminosity blocks* (LBs), which are periods of data taking in which the detector *conditions* are considered constants. Generally, LBs last approximately 60 s. The detector conditions refer to the status of the trigger in use, the configuration of the subsystem, and other information that can vary during a run, such as the calibration constants of the subsystems. Conditions are stored in the ATLAS conditions database according to *intervals of validity*, which note the beginning and end of the time interval to which they apply.

Fig. 2.11 schematizes the DQ workflow [53]. The DQ operations begin *online*, meaning that monitoring starts already during real-time data taking. Data passing trigger requirements are divided into *streams*, according to the sets of trigger selections and prescales satisfied by their events. Different streams serve different purposes, some related to specific calibrations and others for collecting data potentially interesting for physics analysis. Additionally, the *express* stream contains a subset of the data collected in the others. It is characterized by a small rate (~ 20 Hz) of events: since it contains those that are used for the first stages of the DQ assessment, the full reconstruction of the stream must be as fast as possible. A subset of the express stream of the data is reconstructed online and used to produce distributions of the features of the reconstructed physics objects. These are compared to reference distributions obtained during previous runs free of DQ issues and ideally sharing similar conditions. They are checked 24 hours a day by the crew of online DQ *shifters*, which are provided with a dedicated monitor in the ATLAS Control Room. In addition to the physics objects distributions, online DQ shifters monitor histograms related to the trigger and data acquisition status. All the monitoring histograms are updated every few minutes with the additional data. The work of the online shifters can help minimise potential data loss. For instance, if the hit occupancy per unit area in the ID system is uniform, but a “hole” is present in the histograms monitoring track occupancy, a local calibration issue could be present, which would be difficult to spot at the desks of the various subsystems involved.

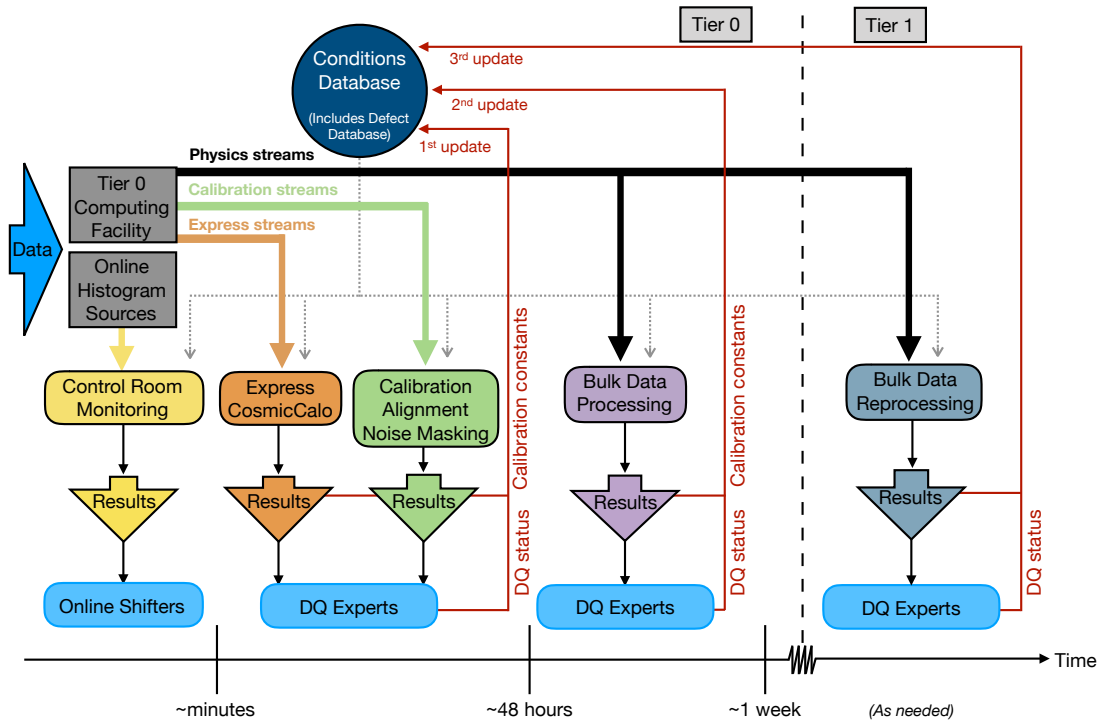


Figure 2.11: Schematic diagram illustrating the nominal operations workflow for the DQ assessment of ATLAS data [53]. Online histogram sources include the HLT farm, the data acquisition system, and full reconstruction of a fraction of events accepted by the trigger.

The offline DQ stages occur once the data has been collected and reconstructed. The offline DQ group is divided into subgroups, responsible for the various subsystems of the detector and for monitoring detector-level and physics objects reconstructed by combining signals in various systems. Experts decide the set of histograms relevant for assessing the DQ from their groups' perspective. Furthermore, they identify which subset should be watched by the online shifters in the Control Room.

The first step of the offline DQ workflow is the *calibration loop*. The monitoring histograms are produced with the data in the express and calibration streams. The scrutiny of the express stream data allows an early check of the data collected during the run. This is important in cases where actions can correct DQ issues before the reconstruction of the full physics streams, for instance, by updating calibration constants in the condition database.

Calibration streams are used by the detector experts to verify the conditions of the subsystems. Additionally, they are used to identify the constant related to the alignment of the ID subsystems and to identify the location of the *beam spot*, the luminosity-weighted centroid of which defines the average collision point.

When the calibration loop is completed, and the conditions database is updated, the processing of the physics streams is launched. The DQ subgroups perform the final DQ assessment on the full data statistics, checking all their monitoring histograms, and ensuring the updated conditions are correctly applied. For instance, Fig. 2.12 displays the d_0 of tracks as a function of ϕ in the express stream and the physics stream. The positions of the PVs are computed with respect to the beam spot. The beam spot is known less precisely in the express stream: its location is coarsely determined online with the few events reconstructed immediately. In the physics stream, the beam spot location is found with a higher resolution via the calibration loop. The result is that, in the express stream, some dependence of d_0 on ϕ can be seen. The dependence disappears thanks to the calibration.

After this final assessment, the data that pass the DQ checks are provided to the Collaboration via dedicated lists, called Good Run Lists (GRLs). All the DQ procedures can be repeated when a reprocessing of the data occurs with an updated conditions database or with a new version of the software. This can happen in cases in which a new understanding of the conditions allows to improve in the DQ efficiency of a given dataset, as is the case for the example provided in Sec. 3.6.1.

2.5 Monte Carlo generators for hadron colliders

Testing a theory at a collider requires comparing the collision data with theoretical predictions. Obtaining such predictions involves extremely complex computations, most noticeably solving multi-dimensional integrals. The complexity is exacerbated at a proton-proton collider such as the LHC, where QCD effects in the non-perturbative regime play a significant role in shaping the final state. To handle such complexity, numerical methods, which

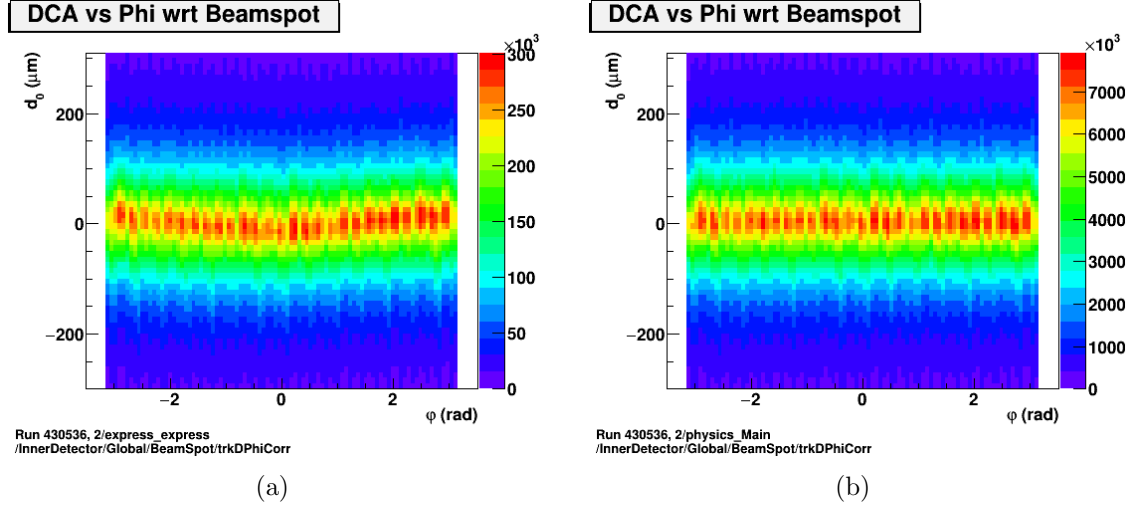


Figure 2.12: Transverse IP d_0 - termed Distance of Closest Approach (DCA) in the title of the plots - versus ϕ for charged-particle tracks. (a) For the first processing of the express stream, the beam-spot location is not yet known accurately and d_0 is calculated relative to the beam-spot position as determined online. (b) A correctly determined beam spot results in a flat distribution after the application of updated conditions during the calibration loop, since here d_0 is calculated relative to the reconstructed beam spot.

primarily rely on MC techniques, are employed. They result in a full simulation of all the physics processes at play.

First, the hard scattering interaction is determined by calculating the hadronic cross-section $\sigma_{pp \rightarrow X}$, defined by

$$\sigma_{pp \rightarrow X} = \sum_{a,b} \int_0^1 dx_1 \int_0^1 dx_2 f_a(x_1, \mu_F) f_b(x_2, \mu_F) \int d\sigma_{ab \rightarrow X}(x_1 P_1, x_2 P_2, \mu_R, \mu_F). \quad (2.9)$$

Here, the *parton distribution functions* (PDFs) $f_i(x_i, \mu_F)$ describe the probability of a parton i carrying a fraction x_i of the proton momentum P_i to engage in a hard scattering collision with momentum $p_i = x_i P_i$. PDFs depend on the energy scale of the collision, but are independent from the hard-scattering interaction which takes place, a phenomenology termed *factorization*. Factorization ultimately arises from the running of α_s [54], which allows to picture the proton-proton collision as an interaction between asymptotically free high energy partons picked from a lower energy non-perturbative proton system. The factorization scale, μ_F , can be intuitively described as the energy below which long-distance non-perturbative effects take over. $\sigma_{ab \rightarrow X}$ indicates the cross section of the $ab \rightarrow X$ process, where a and b are partons of the colliding protons, which depends on the renormalization scale μ_R , involved in the determination of the renormalized value of α_s . Many measurements in ATLAS focus in measuring parameters which enter $\sigma_{ab \rightarrow X}$ as parts of the Lagrangian

density used to determine it. This is the case for the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ analysis detailed in Chap. 6.

In the hard scatter, highly energetic particles are generated. When these are gluons and quarks, they are initially too energetic to form bound states, leading to the production of additional quarks and gluons through radiation or gluon splitting. This process, known as *parton shower generation*, is simulated separately using perturbative QCD calculations. Once the energy of the gluons decreases to the order of $O(1)$ GeV, quarks and gluons form bound states through hadronization. Perturbative QCD is no longer applicable at these energy scales, and phenomenological QCD models are employed to simulate the hadronization. Additionally, the possibility of initial state radiation (ISR) and final state radiation (FSR) in the form of gluons, or, less frequently, photons, emitted from the ingoing and outgoing partons, must be considered. Alongside the hard scattering, interactions among the remaining partons present in the proton are simulated as the *underlying event*. Finally, pile-up collisions are accounted for by overlaying soft QCD events to the simulated event before further processing.

Various MC generators on the market implement these computations, either simulating all described steps or interfacing with each other to simulate different stages. Once events are generated, they are processed through a simulation of the ATLAS detector and undergo the identical reconstruction performed on real collision data; while storing the truth information regarding every simulated particle.

The ATLAS simulation chain consists of four steps [55] implemented within the ATHENA [56] software framework. It consists in event generation using MC generators, calculation of the matrix element for the hard scatter process, parton shower, and hadronization. The detector simulation, the most computationally intensive step, uses the GEANT4 toolkit [57] to simulate the full detector geometry, material, and response. Particle interactions with the detector material are recorded and digitized to simulate the detector output, and stored as *Raw Data Objects*. Particles are finally reconstructed with the same methodology applied to data.

Chapter 3

Jet flavour-tagging in ATLAS

Jet flavour-tagging is the method used in particle physics to identify the flavour of hadrons¹ within a jet. In simulation, small- R jets containing b -hadrons with $p_T > 5 \text{ GeV}$ and $\Delta R(\text{hadron, jet axis}) < 0.3$ are labelled as b -jets, those containing c -hadrons but no b -hadrons are labelled as c -jets, those containing τ leptons but no b or c -hadrons are labelled as τ -jets, and the remaining jets are labelled as light-flavour jets.

Colour charges are never observed in isolation, thus, the detection of heavy-flavour hadrons within jets clustered in observed data is essential to test whether a heavy-flavour quark was produced during a proton-proton collision. This is enabled by the application of jet flavour-tagging algorithms, which harness the distinct features of heavy-flavour hadrons and their decays.

Heavy-flavor quarks hold significant importance in studying both the SM and potential phenomena beyond it. Consequently, jet flavour-tagging plays a central role in various areas of particle physics, including investigating the properties of the Higgs boson, searching for new physics beyond the SM, and exploring the characteristics of quarks and gluons themselves [58–60].

3.1 Identikit of a heavy-flavour jet

Heavy-flavour hadrons can be identified by means of their distinctive characteristics. One crucial feature is their long lifetime, allowing them to travel considerable distances before decaying. Additionally, they have relatively large masses and high decay multiplicities to stable, charged particles, and they carry a large fraction of the p_T of the jet to which they

¹Hadrons inherit their flavour from their valence quarks. b -hadrons and c -hadrons, *i.e.*, hadrons containing either a b or a c valence quark, are termed *heavy-flavour* hadrons, as opposed to light-flavour hadrons. Top quarks do not form bound states.

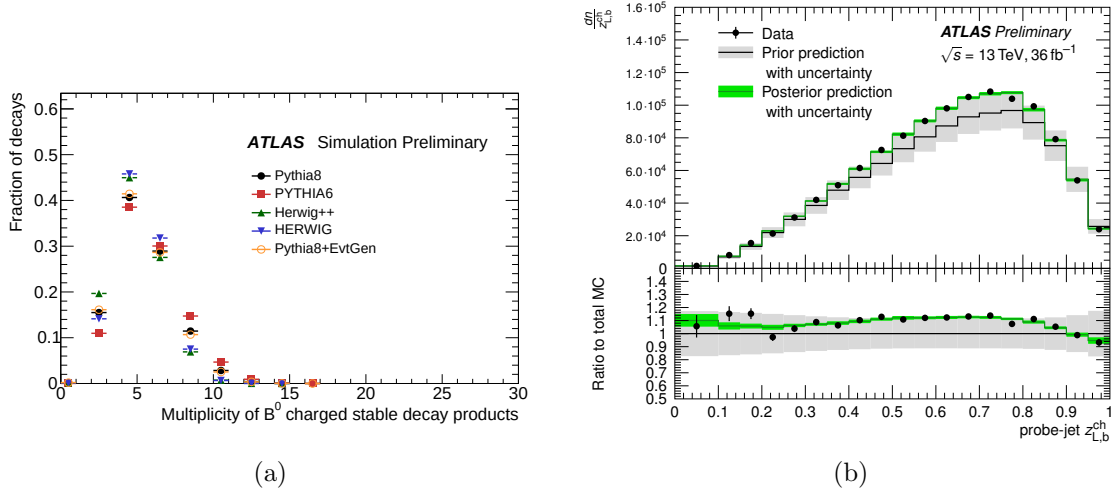


Figure 3.1: (a) Decay multiplicity of the B^0 hadron into charged stable products compared for different MC generators [62]. Noticeably, albeit the mode value for all the generators falls within the bin that comprises the measured average [63], the predictions of the different generators vary by as much as 20%. Differences such as this pose challenges when characterizing the performance of the jet flavour-tagging algorithms described in later sections, as they affect the efficiency with which heavy-flavour jets can be identified. These challenges are addressed using dedicated MC-MC calibrations. (b) Fraction of the jet charged momentum carried by the b -hadron decay products along the direction of the jet charged momentum, measured using proton-proton collision data unfolded at particle level for comparison with the MC predictions [64].

are associated. Furthermore, they have a relatively high semi-leptonic decay² fraction. b -hadrons, for instance, have a lifetime of around 1.5 ps [61]: it means that those produced with a momentum of 30 GeV, on average, decay only after 2.5 mm - well above the spatial resolution of the tracking system of the ATLAS detector. Their mass is of about 5 GeV [61], and they have a decay multiplicity of around 5 charged particles, as shown for different MC generators in Fig. 3.1a). Fig. 3.1b displays the fraction of the jet charged momentum³ carried by the b -hadron decay products along the direction of the jet charged momentum, which peaks at around 70%. Finally, b -hadrons undergo semi-leptonic decays around 10% of the times.

Fig. 3.2 illustrates the typical structure of a heavy-flavour jet. The extended lifetime of the heavy-flavour hadron results in a displaced secondary vertex, and in cases where the

²Semi-leptonic decays are decays of heavy hadrons into lighter hadrons by emitting a lepton and the associated neutrino.

³The jet charged momentum is computed in the measurement summing the momenta of the tracks associated to the PV and to the candidate secondary vertices reconstructed within the jet using the SV1 algorithm described in Sec. 3.2.2.

b -hadron decays to a c -hadron, a tertiary vertex might also be present. The tracks stemming from these displaced decays of heavy-flavour hadrons exhibit large IPs due to their long lifetimes and large masses. Additionally, a soft lepton⁴ from a semi-leptonic decay is depicted.

c -jets have features similar to b -jets, but on a smaller scale. They are lighter, have a shorter lifetime, and lower decay multiplicities, resulting in a comparable topology that makes distinguishing between the two challenging.

In contrast, light-flavour jets primarily consist of tracks produced from quark hadronisation, originating mostly from the primary vertex. These distinctions enable clear differentiation between heavy-flavour jets and light-flavour jets.

However, distinguishing heavy-flavour from light-flavour jets can be challenging since certain processes can mimic heavy-flavour content in the latter. As an example, displaced vertices may also occur in light-flavour jets due to various factors. These factors include random combinations of tracks, interactions with detector material, photon conversions, the presence of long-lived particles, and erroneous association of tracks coming from another PU collision to the jet under study (Fig. 3.3).

3.2 A recent history of jet flavour-tagging in ATLAS

The performance of jet flavour-tagging algorithms developed by the ATLAS Collaboration has greatly improved in recent years, harnessing the spectacular advancements that have been achieved in the application of artificial intelligence techniques to all kind of fields. Broadly speaking, this performance can be quantified by measuring the proportion of heavy-flavour jets correctly identified against the rate of rejection of background jets⁵. The performance is firstly studied on MC simulated jets, where the truth flavour label can be accessed, and then is corrected to account for the difference between simulation and data by means of dedicated calibrations. With this definition of performance in mind, the scope of the huge leap achieved in a short time can be fully appreciated by observing the chart in Fig 3.4. The chart is about b -tagging performance, which means that b -jets are seen as signal, and c and light-flavour jets are regarded as background. In less than ten years, the background rejection rates improved fourfold. In this chapter, all of the algorithms mentioned in the chart are detailed.

When I begun my PhD, the algorithm called DL1d had just been developed. This section is dedicated to the description of what had happened before. Such description is due, since

⁴Soft leptons, such as those produced in semi-leptonic decays of heavy-flavour hadrons, are defined by having a p_T detectable and of the order of ~ 10 GeV, depending on the transverse momentum of the decaying heavy-flavour hadron. They are opposed to the harder leptons produced in the leptonic decays of gauge vector bosons with transverse momenta comparable to the heavy-flavour hadrons’.

⁵Even though much refined characterizations are presented for the algorithms that I personally developed, which are the object of Sec. 3.3 and of Chap. 4.

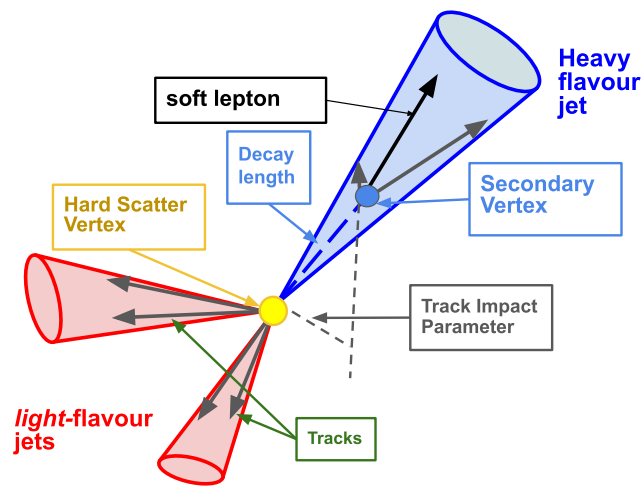


Figure 3.2: Identikit of an heavy-flavour jet. The long lifetime of the heavy-flavour hadron results in a decay displaced from the hard scatter vertex. Tracks emerging from the decay are produced with a large IP. The decay is semi-leptonic in about 10% of the cases for b -hadrons, for which a soft lepton is present. Tracks associated to the light-flavour jets, instead, stem from the hard scatter vertex.

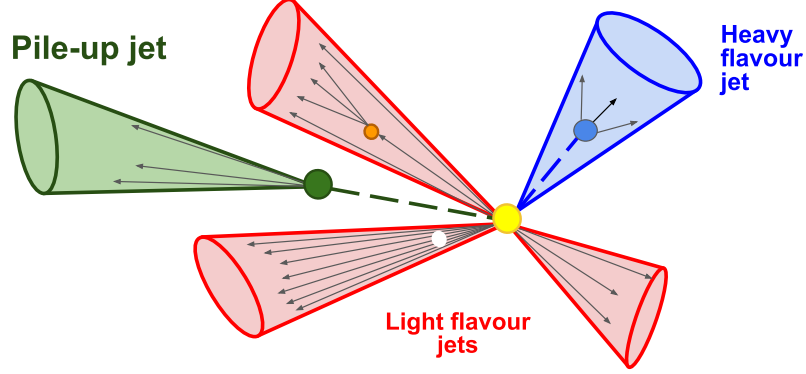


Figure 3.3: Different ways in which heavy-flavour content can be mimicked within light-flavour jets: fake secondary vertex emerging from a random combination of tracks in dense environments; real secondary vertex from processes such as interaction with the detector, photon conversion, decay of long-lived particles; pile-up jet associated to the primary vertex.

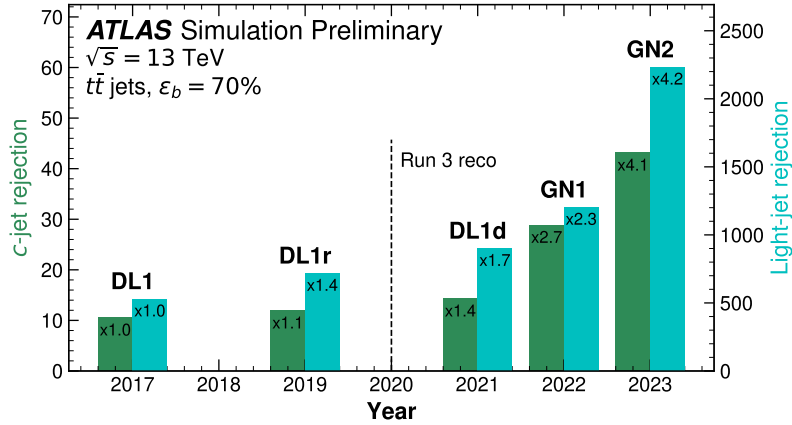


Figure 3.4: b -tagging performance of the subsequent algorithms developed within the ATLAS Collaboration in recent years. The background rejection rates underwent a fourfold improvement in a very limited time span. From [65].

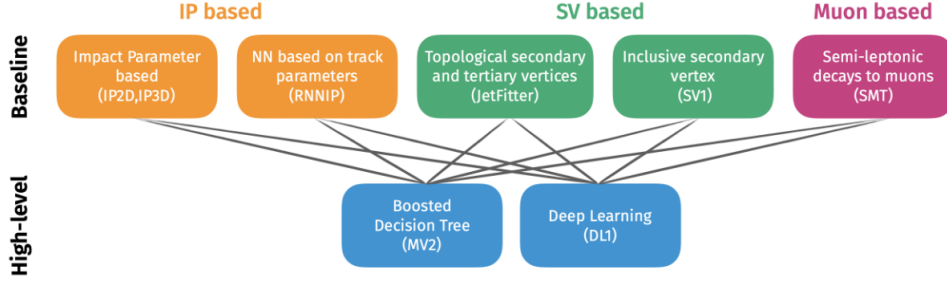


Figure 3.5: Jet flavour-tagging low-level and high-level algorithms available before Run-3. Low-level algorithms are manually optimized to target specific features of the heavy-flavour jets. Their outputs are combined by multivariate high-level algorithms into a final discriminant variable used for jet flavour identification. Image from the ATLAS internal flavour-tagging documentation web site.

the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ measurement, detailed in Chap. 6, uses the DL1r algorithm to identify the flavours of the jets; and because I trained the SVKine and JFKine algorithms, presented in later sections, motivated by use-cases that mostly refer to the pre-GN1 paradigm. Furthermore, flavour-tagging data quality activities, which I have coordinated together with a colleague since the beginning of Run-3, still monitor the DL1d performance.

Before GN1, jet flavour-tagging algorithms followed a two-stage approach, involving a vast set of intermediate algorithms targeting specific features of the heavy-flavour hadrons. Their names and relations are illustrated in Fig. 3.5. First, *baseline* algorithms (also known as *low-level* algorithms) interpret specific features of tracks, secondary vertices and leptons associated to the jets to provide variables useful for selecting jet flavours. Then, their outputs are collectively used as inputs of multivariate *high-level* algorithms that produce the discriminant variable which is used for the final jet flavour identification.

MC samples for jet flavour-tagging developments Before describing the jet flavour-tagging algorithms, a brief presentation of the two samples of simulated jets used to develop them is provided. They are composed of jets produced in $t\bar{t}$ and $Z' \rightarrow q\bar{q}$ events respectively⁶, to populate the low and high regions of the jet p_T spectrum, covering the 20 GeV to 6 TeV interval.

The $t\bar{t}$ sample [66] was generated using the POWHEG BOX v2 tool. It provides matrix elements for producing top-quark pairs at the next-to-leading order (NLO) in terms of the strong coupling constant α_s . It uses a first-gluon-emission cut-off scale parameter, h_{damp} , set at 1.5 times the top quark’s mass, which was taken as $m_t = 172.5$ GeV. POWHEG BOX is interfaced to PYTHIA 8.230 using the A14 set of tuned parameters and parton distribu-

⁶The Z' is a hypothetical, non-SM resonance with features engineered ad hoc for the benefit of the jet flavour-tagging studies.

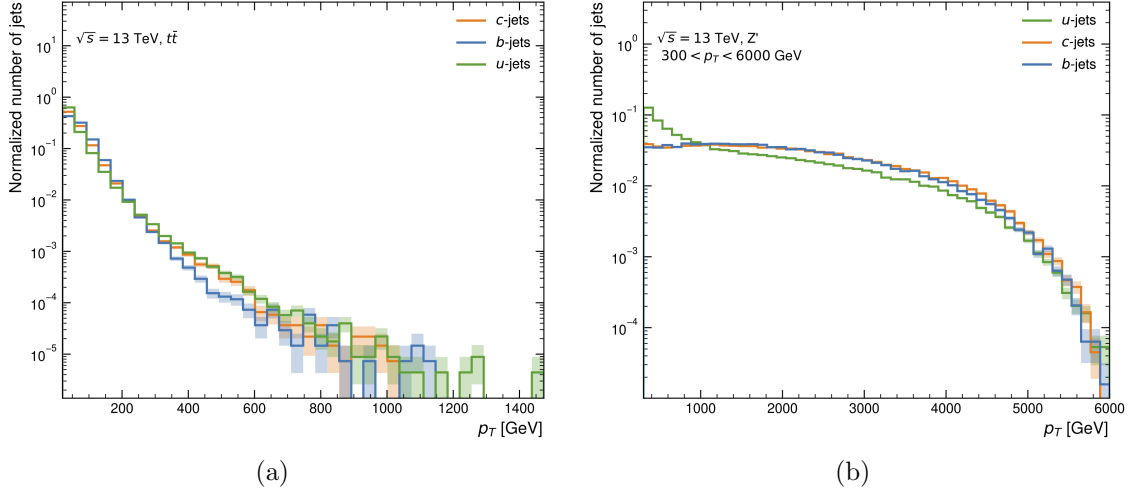


Figure 3.6: p_T distribution of simulated jets in the $t\bar{t}$ (a) and Z' (a) samples used to develop ATLAS jet flavour-tagging algorithms.

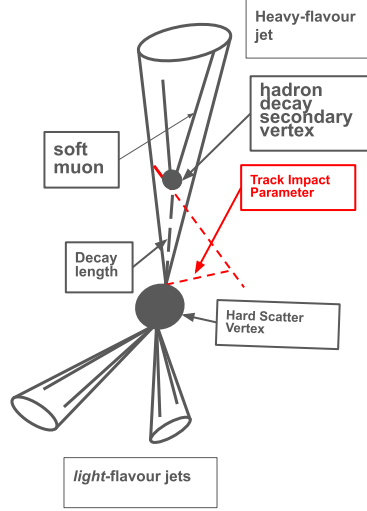
tion functions from NNPDF3.0NNLO (for matrix elements) and NNPDF2.3LO (for parton shower). Only events where at least one W boson decays leptonically are included. This criterion ensures a high fraction of b -jets, c -jets, and light-flavour jets in the jet population.

The $t\bar{t}$ sample contains very few jets at the TeV scale (Fig. 3.6a). The Z' sample [66] was designed specifically to populate also such a high- p_T region of the phase space. It was generated using PYTHIA 8.212 with the A14 set of tuned parameters for the underlying event and the leading-order NNPDF2.3LO parton distribution functions. Unlike the $t\bar{t}$ sample, which simulates a SM process, the Z' sample uses a hypothetical resonance, the Z' , with a modified production cross-section in the hard-scattering process that broadens its natural width. The result is a sample with a fairly flat p_T spectrum in the 250 GeV to 1.5 TeV interval (Fig. 3.6b). The Z' is forced to decay to b -jet, c -jet and light-flavour jet pairs with equal probability to obtain a sample uniformly populated by jets of each flavour.

The simulated events in both samples are processed through the full ATLAS detector simulation based on GEANT4.

Jet flavour-tagging specific object selections The second excursus preceding the presentation of the jet flavour-tagging algorithms regards the specific selections on the detector-level objects and jets that are used to develop them. Jet flavour-tagging primarily relies on efficiently reconstructing tracks, PVs and jets. Hence, specific criteria are imposed to select these objects [66].

Tracks are required to have $p_T > 500$ MeV and $|\eta| < 2.5$. Furthermore, they must satisfy additional “Loose” criteria [67], based on their hit content, which improve rejection of fake and poorly reconstructed tracks.



143

Figure 3.7: Tracks emerging from a displaced decay of a heavy-flavour hadron can have a large IP.

The reconstruction of the primary vertex [42,68] holds paramount importance for jet flavour-tagging, as it sets the reference point from which the tracks' IPs and the displacement of secondary vertices are calculated. Hence, at least one PV in each event is required.

Jets are PFlow jets with a radius parameter of $R = 0.4$ and with $p_T > 20$ GeV. When dealing with jets that fall within the $20 \text{ GeV} < p_T < 60 \text{ GeV}$ range and have $|\eta| < 2.4$, an extra Jet Vertex Tagger (JVT) [69] cut of $\text{JVT} > 0.5$ is applied. JVT is a multivariate discriminant designed to assign a low score to jets with a large fraction of high-momentum tracks from pile-up vertices. It aims at suppressing the number of pile-up jets in the dataset. Jets overlapping with an electron or muon originating from a W -boson decay are discarded.

Low-level algorithms are divided into three broad categories. The first consists of track-based taggers that exploit the large IP of tracks originated by heavy-flavour hadron decays. They are called **IPxD**, **RNNIP** and **DIPS**. The second category consists of algorithms that explicitly reconstruct displaced vertices: **SV1** and **JetFitter**. Finally, **SMT** targets the features of the leptons emerging from the semi-leptonic decays of the heavy-flavour hadrons.

3.2.1 Algorithms based on IPs of tracks

Algorithms based on IP of tracks leverage the long decay path of the heavy-flavour hadrons, which results in a displaced vertex. The IP can be larger for tracks emerging from such vertices compared to tracks originating from the PV (Fig. 3.7).

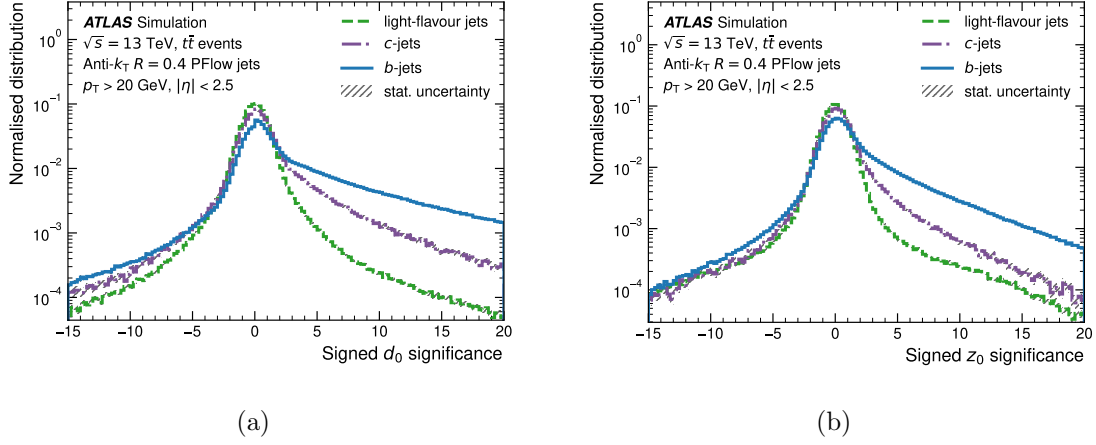


Figure 3.8: The (a) transverse and (b) longitudinal signed IP significance of tracks for b -jets, c -jets and light-flavour jets in $t\bar{t}$ events. [66].

So-called lifetime-signed significances are computed from the transverse and longitudinal IP: $s_{d_0} = d_0/\sigma_{d_0}$ and $s_{z_0} = z_0 \sin \theta / \sigma_{z_0} \sin \theta$ (Fig. 3.8). They represent the IP divided by its associated uncertainty. The sign is established by extrapolating the track to the primary vertex. A negative sign is assigned if the jet axis needs to be extended backwards from the primary vertex to intersect the track or its projection. This scenario arises mostly due to resolution effects. A positive sign is given otherwise.

The IPxD algorithms: The IP2D and IP3D algorithms [70], collectively termed IPxD, compute per-track probabilities p_b , p_c , and p_u of being part of a jet of a specific flavour based on probability density functions derived as described later in this paragraph. These probabilities are subsequently merged in log-likelihood ratio (LLR) discriminants.

Tracks are categorized based on their hit content quality, and from these classifications, probability density functions are derived to compute p_b , p_c , and p_u for each track. These functions are drawn from the MC distribution of s_{d_0} for IP2D and the joint (s_{d_0}, s_{z_0}) MC distribution for IP3D. Therefore, IP3D can encapsulate correlations between s_{d_0} and s_{z_0} .

Finally, the LLR discriminants for various jet-flavour hypotheses are calculated as the sum of the logarithms of the ratios of the per-track probabilities:

$$\text{IPxD}_{l,c,cl} = \sum_{i \in \text{tracks}} \log \left(\frac{p_{b,b,c}^i}{p_{l,c,l}^i} \right) \quad (3.1)$$

As indicated in Eq. 3.1, three distinct discriminants are computed to differentiate between b -jets and light-flavour jets, b -jets and c -jets, as well as c -jets and light-flavour jets. IP2D

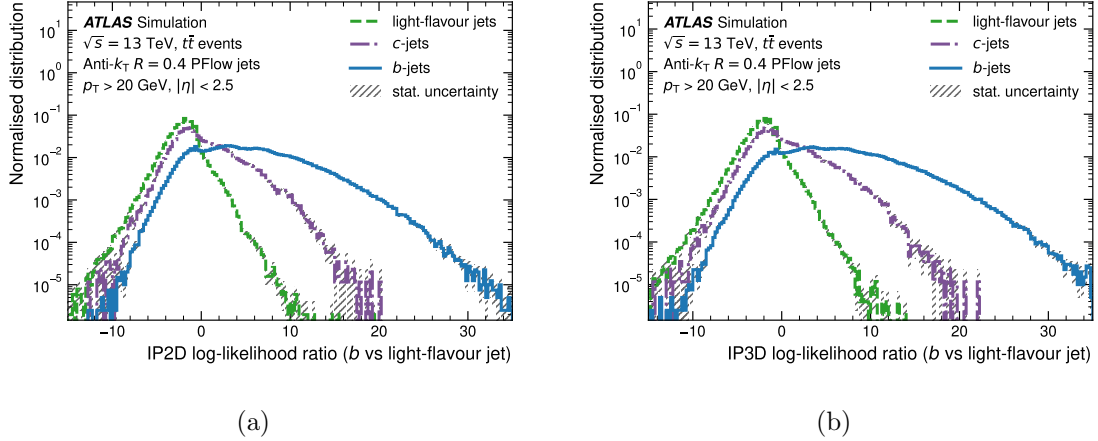


Figure 3.9: The log-likelihood ratio for the (a) IP2D and (b) IP3D b-tagging algorithms for b -jets, c -jets and light-flavour jets in $t\bar{t}$ events. The log-likelihood ratio shown here is computed as the ratio of the b -jet to light-flavour jet hypothesis probabilities. Jets with no tracks are not included in the plot, but assigned a large negative value. [66]

and IP3D b vs light-flavour jet discriminants are shown in Fig. 3.9.

The RNNIP algorithm: Applying the LLR technique used by the IPxD algorithms to higher-dimensional track representations while accounting for the correlations between track properties is nearly impossible. As the number n of dimensions increases, the statistics required to estimate the probability density functions used to compute the probabilities p_b , p_c , and p_u of each track from the MC simulations scales exponentially with n ⁷. This limitation also impedes an application of the technique accounting for correlations between different tracks in the jet. In cases where a displaced decay is present, the IPs of the emerging tracks are correlated, as they are more likely larger than those of prompt tracks.

RNNIP [71] is a Recurrent Neural Network (RNN) operating on multivariate representations of tracks. Thanks to its *long term short term* memory (LSTM) network, it preserves correlations between different tracks. In addition to s_{d_0} and s_{z_0} , selected tracks are represented by the fraction p_T^{frac} of the jet- p_T that they carry, by their angular distance ΔR to the jet axis, and by their hit content.

RNNs operate on ordered sets: tracks are ordered by decreasing $|s_{d_0}|$. Hence, each jet is represented by an ordered sequence of tracks, and each track by a vector of variables. Sequences are fed to the RNN schematized in Fig. 3.10: a LSTM cell converts the variable-length track sequence into a 50-dimensional vector. This vector feeds into a fully connected layer, which provides three outputs. The outputs are finally normalized through a softmax layer. They represent the probabilities that RNNIP assigns to each jet of being a b , c or

⁷This phenomenon is sometimes referred as *the curse of dimensionality*.

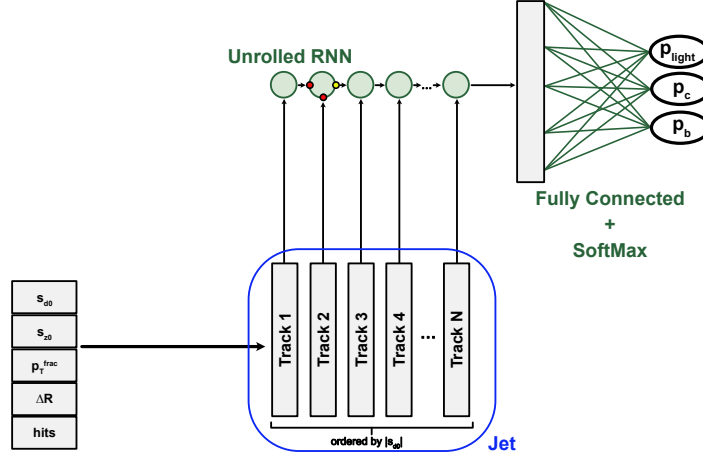


Figure 3.10: Schematic drawing of the RNNIP neural network architecture [71]. The s_{d0} and s_{z0} input variables correspond to the lifetime-correlated signed transverse and longitudinal IP significances, while p_T^{frac} and ΔR represent the fraction of transverse momentum carried by the track relative to the jet and the angular distance between the track and the jet axis, respectively.

light-flavour jet (Fig. 3.11).

The DIPS algorithm: The sequence in which tracks in jets are provided as input to RNNIP during training can influence its performance. Such ordering is chosen empirically, as there is no inherently ideal method for arranging tracks in jets. DIPS [72] is a DeepSets-based NN that retains the advantages of the RNNIP algorithm over IPxD, while encoding permutation invariance among the tracks in the jet. This is achieved through the architecture outlined in Fig. 3.12. Each track undergoes processing by a NN Φ , which extracts its relevant features to construct a new, higher-dimensional representation. Then, these tracks are summed, and the result is passed through a NN F , which delivers the p_b , p_c , and p_u probabilities allocated to the jet.

The DIPS algorithm achieves permutation invariance of tracks through a summing layer. This offers a more natural set representation and enables parallel processing of the tracks via Φ . As a result, training and evaluation times are significantly reduced, thereby facilitating more refined optimisations. The initial version of DIPS was trained on tracks selected and represented in the same manner as RNNIP, enabling a direct comparison of their performances. During the optimisation process, a looser track selection was found to enhance performance. Moreover, adding d_0 and $z_0 \sin \theta$ to the set of variables used for track representation proved beneficial.

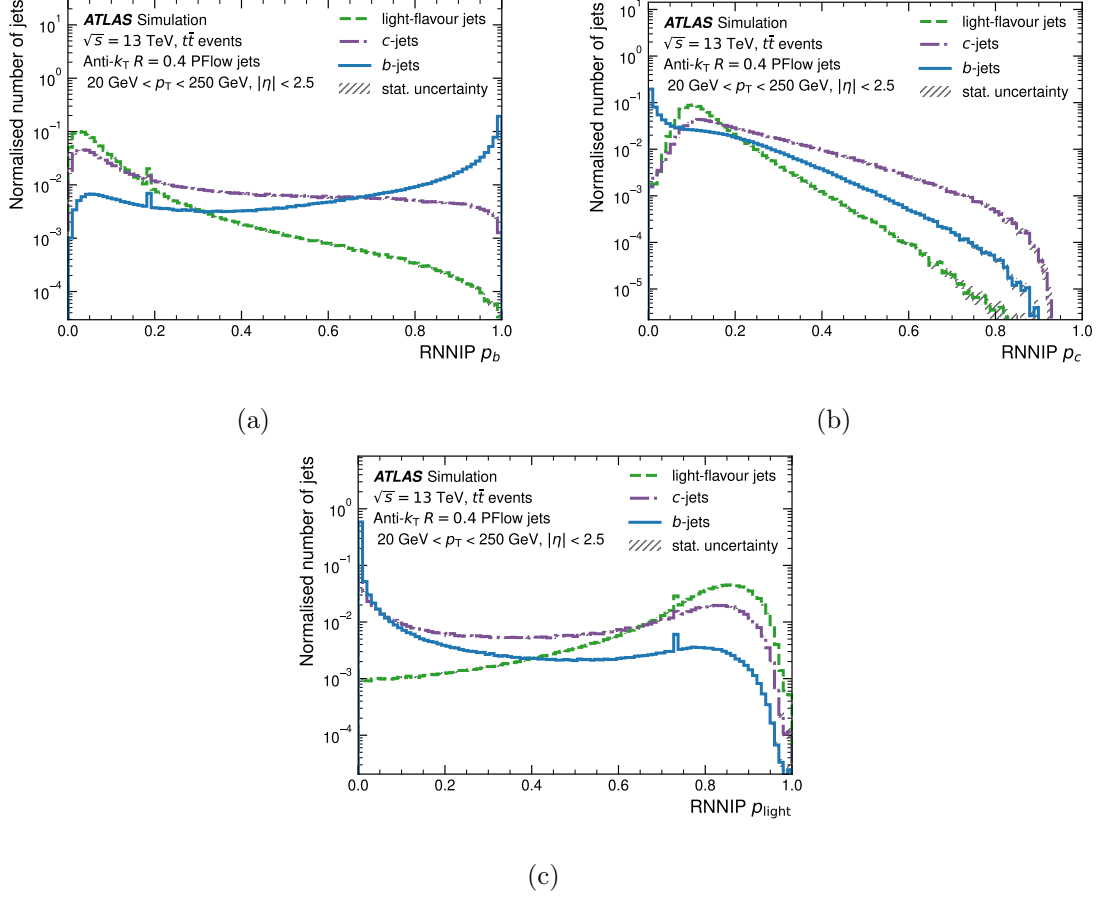


Figure 3.11: Distributions of the outputs of the RNNIP b -tagging algorithm for b -jets, c -jets and light-flavour jets evaluated on a $t\bar{t}$ sample: (a) p_b , (b) p_c , and (c) p_u . The spikes at $p_u \sim 0.73$ and $p_b \sim 0.18$ originate from jets with no tracks. [66]

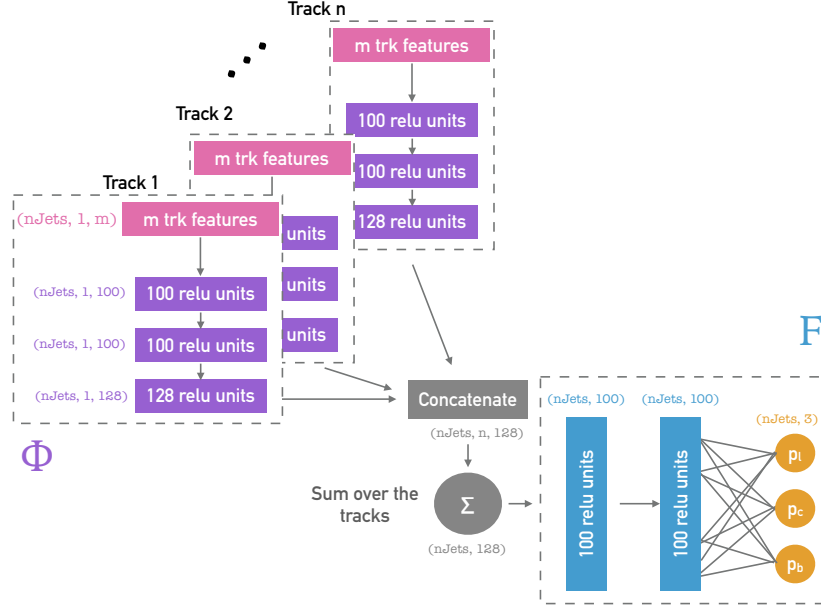


Figure 3.12: Architecture of the DIPS algorithm [72].

3.2.2 Algorithms based on displaced vertices

The IP of tracks can hint at the presence of displaced vertices. However, the illustration in Fig. 3.13 shows that not all tracks emerging from displaced decays have a large IP. Additionally, Fig. 3.3 indicates that the existence of secondary vertices may be linked to processes other than heavy-flavour hadron decays. Consequently, it is crucial to identify the displaced vertices, evaluate their origin, and employ features like track multiplicity and mass as supplementary means to identify the heavy-flavour jets.

SV1, the Secondary-Vertex-Tagging algorithm: The SV1 algorithm [73] reconstructs a single secondary vertex within jets. Most b -hadrons decay into c -hadrons [61], which then undergo subsequent decays. When these two decays occur in close proximity, SV1 merges the vertices. If the decays are distant, the algorithm reconstructs the vertex with the highest track multiplicity. Although this method approximates the b to c -hadron decay chain, it retains the majority of information needed for jet flavour-tagging.

The algorithm unfolds in stages. Initially, it classifies tracks using the Track Classification Tool (TCT). The TCT employs a Boosted Decision Tree (BDT) to represent tracks, using as input features their s_{d_0} and s_{z_0} , η , momentum perpendicular to the jet axis, share of the jet's transverse momentum, hit content, and the p_T of the jet to which they are associated. The BDT outputs three weights for each track, denoting the probability it assigns to the track of originating from a heavy-flavour hadron decay (W_{HF}), fragmentation (W_{FG}), or

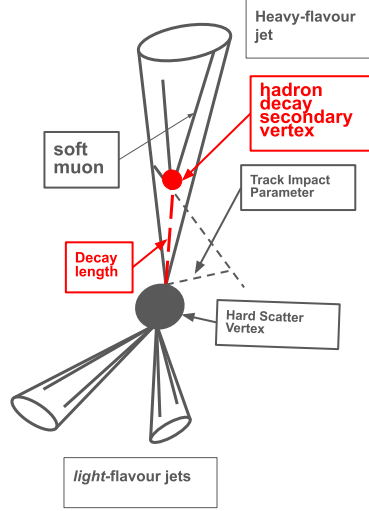


Figure 3.13: The relatively long lifetime of heavy-flavour hadrons often leads to the presence of displaced vertices in heavy-flavour jets.

other processes (W_{GB} ⁸). Tracks with $W_{\text{HF}} < 0.1$ or $W_{\text{GB}} > 0.7$ are discarded.

Then, all the possible couples of tracks originating from nearby points are paired in two-track vertices. A two-track vertex cleaning procedure follows, where two-track vertices are selected based on their vertex score, defined as

$$\text{BVrtScore} = \left[\left(W_{\text{HF}}^{\text{trk1}} W_{\text{HF}}^{\text{trk2}} \right) \right] \left[\left(1 - W_{\text{FG}}^{\text{trk1}} (1 - W_{\text{FG}}^{\text{trk2}}) \right) \right] \left[\left(1 - W_{\text{GB}}^{\text{trk1}} (1 - W_{\text{GB}}^{\text{trk2}}) \right) \right]. \quad (3.2)$$

A lower cut on **BVrtScore** is tuned to eliminate two-track vertices compatible with the decays of K_s^0 or Λ , photon conversions, or hadronic interactions with the detector material.

In the final phase, an iterative method is employed to reconstruct a secondary vertex from all the tracks forming the selected two-track vertices. During each iteration, the track with the highest χ^2 in the fit is removed. This process continues until a satisfactory vertex χ^2 and an invariant mass lighter than 6 GeV are achieved.

Eight discriminant variables that characterize the reconstructed secondary vertex are calculated to serve as inputs for the high-level taggers. These variables include the invariant mass, track multiplicity, and the energy fraction (*i.e.*, the ratio of the energy of the tracks linked to the vertex compared to the total energy of the tracks in the jet), along with the transverse decay length, the decay length and associated 3D significance, the count of two-track vertices reconstructed within the jets, and the ΔR between the jet axis and the

⁸GB stands for “garbage”.

direction of the secondary vertex relative to the primary vertex. Six of these distributions, evaluated on the $t\bar{t}$ simulated sample, are illustrated in Fig. 3.14.

JetFitter, the topological multi-vertex finding algorithm: As opposed to *SV1*, which assumes the presence of just one SV in heavy-flavour jets, *JetFitter* [74] aims at simultaneously reconstructing the full b - c hadron decay chain, assuming that the decay vertices lie on the line defined by the trajectory of the b -hadron, initialised as the jet axis. The assumption implies that all the particles emerging from the decay of the heavy-flavour hadrons must originate from a point on such line. This constraint allows the reconstruction of incomplete topologies in which only a single track from the decay of the heavy-flavour hadron is selected.

A track selection is designed to minimize the tracks not originating from the heavy-flavour hadrons decay. Tracks compatible with the PV are removed. Then, track pairs are fitted to identify two-track vertices. Tracks forming two-track vertices located in regions occupied by the pixel layers are discarded as they likely emerge from interactions with the material. Tracks forming vertices compatible with photon conversions or K_S^0 and Λ^0 decays are also discarded.

An iterative process of vertex finding and fitting is used to identify the most probable number of vertices in the decay chain and fit their positions and the b -hadron trajectory. Initially, all tracks are considered as distinct one-track vertices. A modified Kalman filter [75] fit is applied, extracting information about the full decay chain, encoded in the position of the PV, the flight direction of the b -hadron, and the distances of the secondary vertices to the PV. Vertices with a χ^2 contribution to the overall fit below a 0.1% probability are removed.

Next, a clustering step evaluates all pairs of vertices to determine the probability that each pair belongs to a common vertex and that the vertex is part of the decay chain. This is computed using the same fitting technique. The two vertices with the highest probability are merged, and the process repeats with the new decay chain structure. The iterations continue until all pairs of vertices above a certain vertex mass-dependent threshold have been combined.

In the final stage, tracks that did not form any two-track vertices in the initial iteration are evaluated if they meet strict p_T and IP criteria. The fit procedure is reapplied to see if these tracks can form additional single-track vertices along the b -hadron trajectory.

The algorithm's output is a topology where tracks are uniquely associated with vertices along the path of the b -hadron. Two sets of discriminant variables are extracted to be inputs for high-level taggers. The first set, including the track multiplicity at the *JetFitter* displaced vertices, the invariant mass of tracks associated with these vertices, their energy fraction and their average three-dimensional decay length significance, is evaluated on all the jets. Some of these variables are displayed in Fig. 3.15. The second set is related only to jets in which a single secondary vertex is reconstructed with an intermediate charged

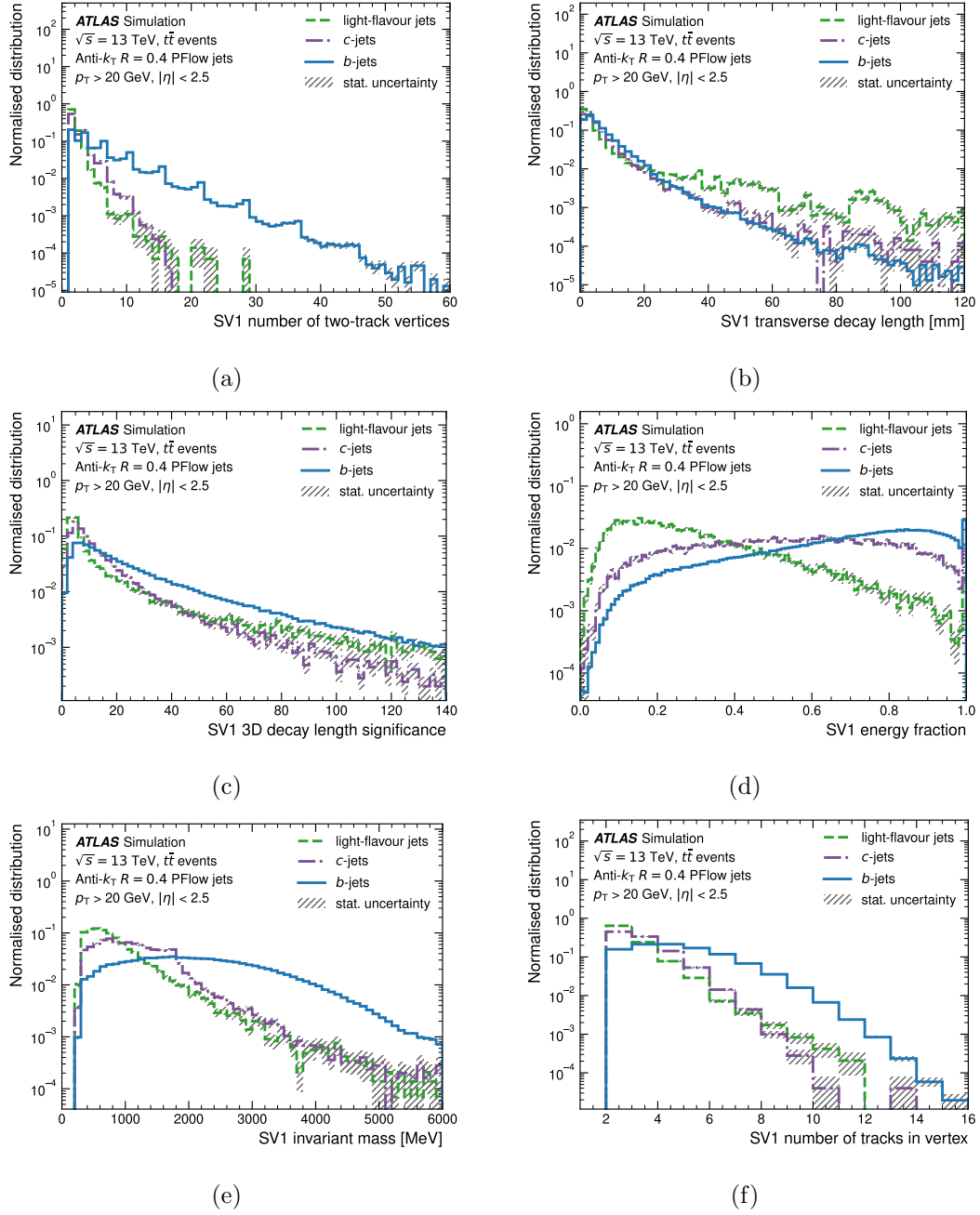


Figure 3.14: Properties of secondary vertices reconstructed by the SV1 algorithm for b -jets, c -jets and light-flavour jets in $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the 3D decay length significance defined as the significance of the distance between the primary vertex and displaced vertex, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated to the vertex. From [66].

decay multiplicity and a decay distance similar to b -hadrons'. Such a set was introduced to improve further the c -tagging performance of the algorithm [70]. It includes the invariant mass and the energy of the tracks associated to the secondary vertex, as well as their rapidity with respect to the jet axis and with respect to the heavy-flavour hadron trajectory.

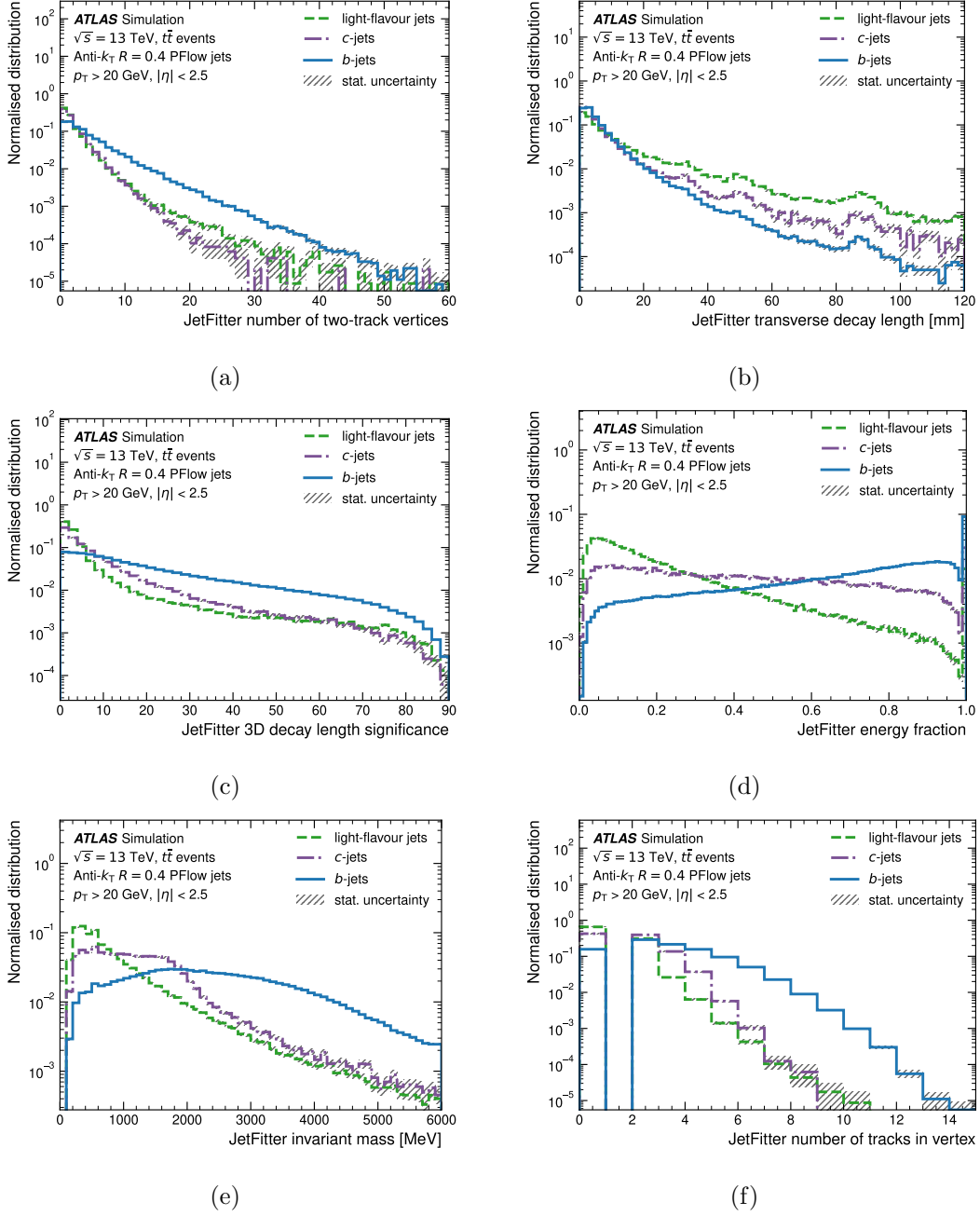


Figure 3.15: Properties of secondary vertices reconstructed by the **JetFitter** algorithm for b -jets, c -jets and light-flavour jets in $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the average 3D decay length significance, defined as the significance of the average distance between the primary vertex and displaced vertices, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated with the vertex. From [66].

3.2.3 Soft Muon Tagger: SMT

As anticipated before, soft muons have an enhanced presence in b -jets, due to the semi-leptonic decays of b -hadrons ($BR(b \rightarrow \mu\nu X) \sim 11\%$) and c -hadrons emerging from decays of b -hadrons ($BR(b \rightarrow c \rightarrow \mu\nu X) \sim 10\%$). SMT [70] uses the features of these soft muons to improve b -tagging.

Only combined muons within a cone of $\Delta R < 0.4$ from the jet axis are considered, further subject to selections regarding their p_T , IPs and $|\eta|$. In light-flavour jets, muons meeting these conditions are predominantly those from the decays of prompt W -boson, light hadrons, light and strange mesons, or τ leptons. Misidentified energetic hadrons that reach the MS also constitute a small but sizeable background.

A set of variables has been defined to identify muons related to a heavy-flavour hadron semi-leptonic decay. Nowadays, these variables have been tested to train some yet uncalibrated versions of the high-level taggers. They include the ΔR between the muon and the jet axis, d_0 , and p_T^{rel} , the projection of the muon's p_T onto the jet axis (Fig. 3.17). The p_T^{rel} is typically larger for particles from the decays of heavy particles, so it's useful for spotting muons from semi-leptonic b -hadrons' decays, but less effective for those from c -hadrons.

3.2.4 The DL1 high-level algorithms

The distinct information provided by low-level taggers, which target specific features of heavy-flavour jets, is used to train the multivariate high-level algorithms. In Fig. 3.5, two series of high-level taggers are included: MV2 [76] and DL1 [66]. The former, whose members consist of BDTs, was used in many ATLAS analyses during Run-2. As MV2 is outperformed by DL1 and was not used for this research, it is not further detailed.

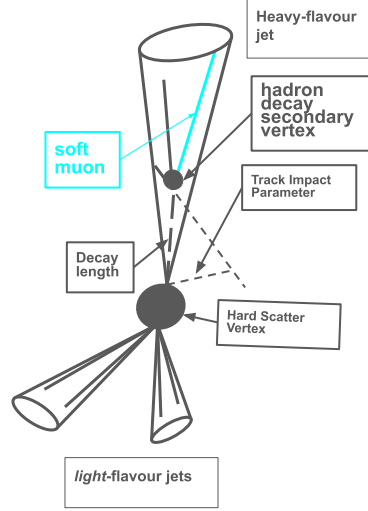
The DL1 algorithms are fully connected, multi-layer feed-forward NNs. They assign three scalar outputs to each jet, p_b , p_c , and p_u , which signify the probabilities that they evaluate for the jet being a b , c , or light-flavour jet. These probabilities are then combined into two distinct discriminant variables, D_b and D_c , as in

$$D_{b/c} = \ln \left(\frac{p_{b/c}}{f_{c/b} p_{c/b} + (1 - f_{c/b}) p_u} \right), \quad (3.3)$$

where f_c and f_b are adjustable parameters governing the trade-off between the rejection rates of the two classes of background jets⁹. Selections based on D_b and D_c define regions, allowing for the classification of jets as b or c -tagged with progressively increasing confidence.

The algorithms are trained on a simulated *hybrid* sample, composed 70% of $t\bar{t}$ events and 30% of Z' events. Both p_T and $|\eta|$ of jets are included among the training variables, allowing the NN to learn their correlations with the low-level taggers' outputs. However,

⁹*i.e.*, c -jets and light-flavour jets when it comes to b -tagging through D_b , and b -jets and light-flavour jets when it comes to c -tagging through D_c .



146

Figure 3.16: b -hadrons and c -hadrons originating from b -hadrons' decays often undergo semi-leptonic decays. The resulting leptons help identify the b -jets that contain them.

the training dataset is resampled¹⁰ to ensure that the p_T and η distributions of all jet flavours are uniform, preventing the models from identifying their flavours predominantly based on the kinematic distributions of the signal and background jets.

The two algorithms most relevant from the standpoint of this research, DL1r and DL1d, are described in the following paragraphs.

The DL1r algorithm: DL1r, where the **r** stands for RNNIP, uses as inputs the IP2D, IP3D, RNNIP, SV1 and JetFitter outputs, along with the jet p_T and $|\eta|$. The network architecture is schematized in Fig. 3.18. The full list of training variables is found in Tab. 3.1, while Tab. 3.2 enumerates the hyperparameters that shape the NNs architecture and define its training campaign. Fig. 3.19 displays the D_b and D_c distributions evaluated with the DL1r algorithm on the $t\bar{t}$ simulated sample.

The DL1d algorithm: DL1d substitutes the RNNIP outputs within DL1r with a joint architecture that integrates DIPS and DL1, as shown in Fig. 3.20. Tracks are processed with a DIPS block that estimates p_b , p_c , and p_u . The last hidden layer of the block is also concatenated¹¹ to a vector created by processing the same jet features used by DL1 through one layer. The resulting vector is then processed through the NN $\mathcal{U}$, which yields additional estimates for p_b , p_c , and p_u .

¹⁰More information about an analogous resampling procedure is provided in Sec. 4.3.1.

¹¹In this context, concatenating two vectors means joining them together into a higher-dimensional single vector.

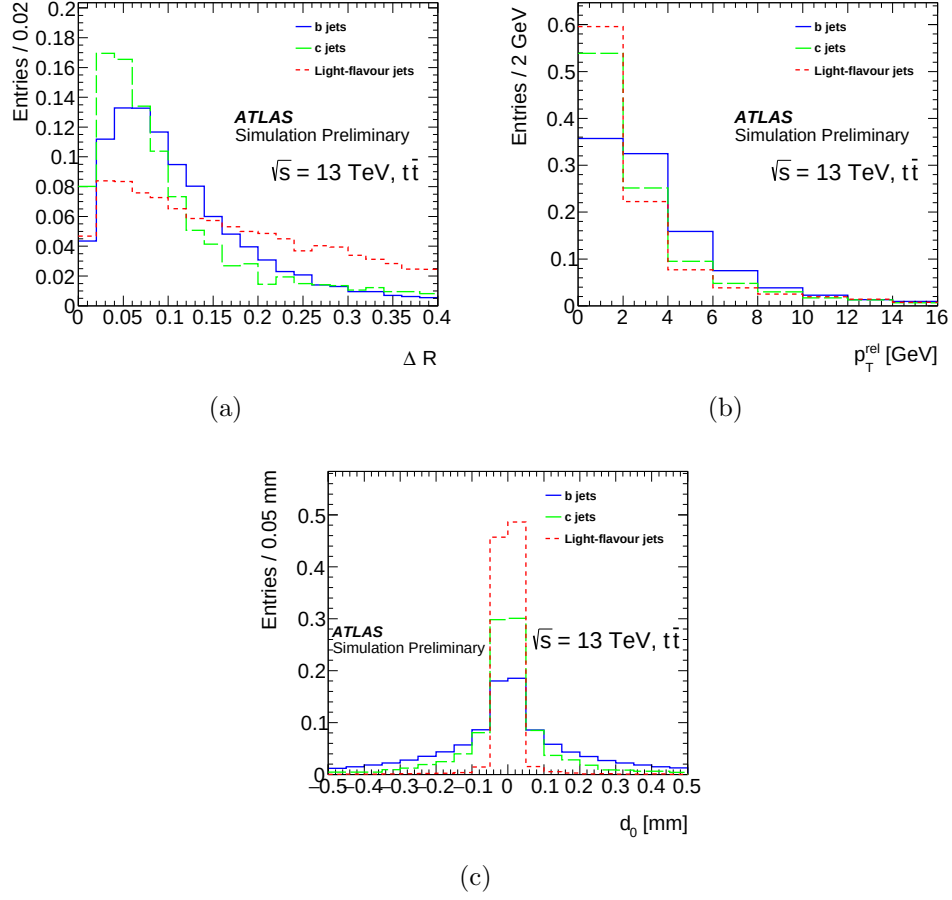


Figure 3.17: ΔR (a), p_T^{rel} (b) and d_0 (c) normalised distributions for reconstructed muons associated to b -jets (blue), c -jets (green) and light-flavour jets (red). From [70].

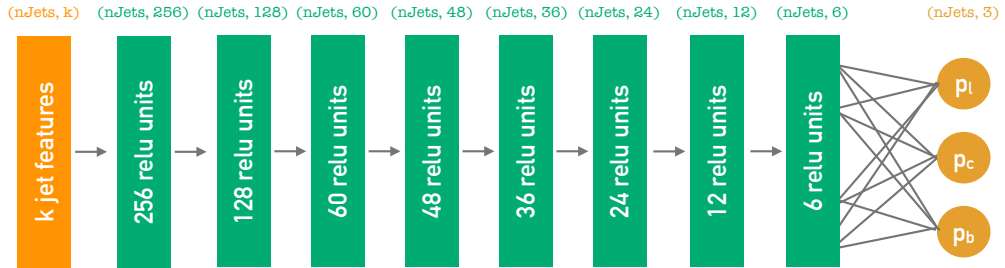


Figure 3.18: Architecture of the DL1r algorithm. Image from the ATLAS internal flavour-tagging documentation web site.

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $ \eta $
IP2D, IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio of the b -jet to light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio of the b -jet to c -jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio of the c -jet to light-flavour jet hypotheses
RNNIP	P_b	b -jet probability
	P_c	c -jet probability
	P_{light}	light-flavour jet probability
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\text{SV})$	Jet energy fraction of the tracks associated with the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{2\text{TrkVtx}}(\text{SV})$	Number of two-track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JetFitter	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Jet energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and the vector sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{2\text{TrkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{1\text{-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices
	$L_{xyz}(2^{\text{nd}})(\text{JF})$	Distance of 2^{nd} vertex from PV
	$L_{xy}(2^{\text{nd}})(\text{JF})$	Transverse displacement of the 2^{nd} vertex
	$m_{\text{Trk}}(2^{\text{nd}})(\text{JF})$	Invariant mass of tracks associated with the 2^{nd} vertex
	$E(2^{\text{nd}})(\text{JF})$	Energy of the tracks associated with the 2^{nd} vertex
	$f_E(2^{\text{nd}})(\text{JF})$	Jet energy fraction of the tracks associated with the 2^{nd} vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}})(\text{JF})$	Number of tracks associated with the 2^{nd} vertex
	$\eta_{\text{trk}}^{\text{min,max,avg}}(2^{\text{nd}})(\text{JF})$	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex

Table 3.1: Input variables used by the DL1r algorithm.

Hyperparameter	Value
Number of input variables	31
Number of hidden layers	8
Number of nodes [per layer]	[256, 128, 60, 48, 36, 24, 12, 6]
Learning rate	0.01
Training batch size	15 000
Activation function	ReLU
Number of training epochs	200
Free (trainable) parameters	59 275
Fixed parameters	1 140
Training sample size	22 M jets

Table 3.2: List of optimised hyperparameters shaping the DL1r architecture and defining its training campaign.

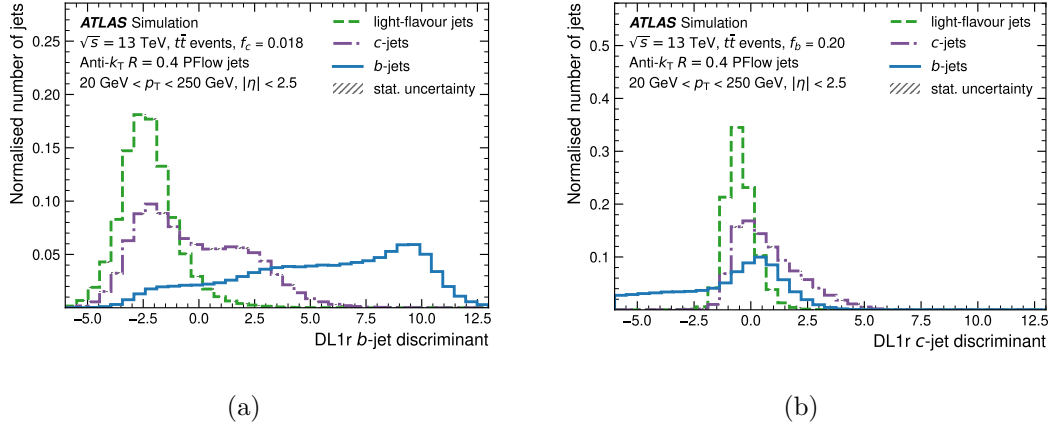


Figure 3.19: D_b and D_c distributions evaluated with the DL1r algorithm on the $t\bar{t}$ simulated sample. From [66].

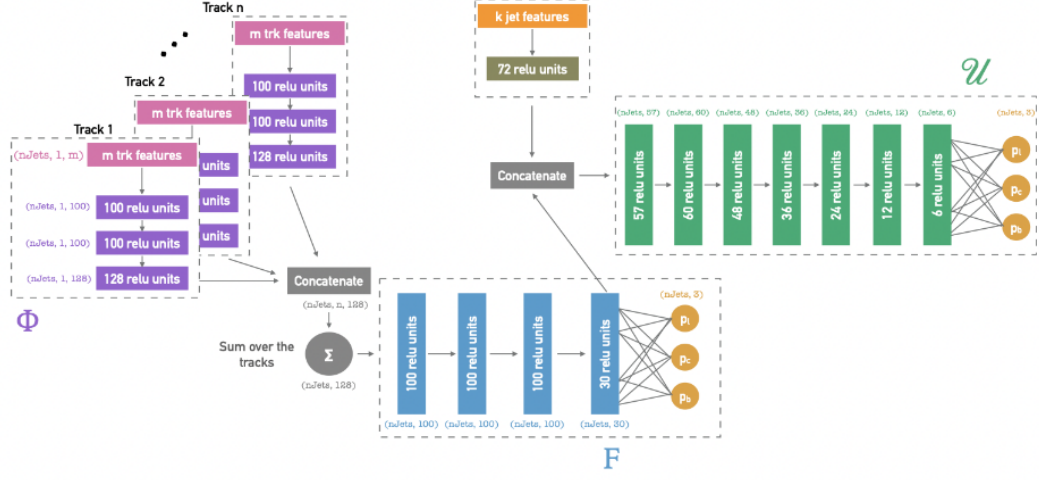


Figure 3.20: Architecture of the DL1d algorithm. Image from the ATLAS internal flavour-tagging documentation web site.

Both the p_b , p_c , and p_u estimates have a dedicated loss. That assigned to the DIPS block is only aware of the features of the tracks, while the loss of the $\mathcal{U}$ network is affected by both the track and jet features. These two terms are then merged into a single loss function, which is employed to perform full back-propagation, simultaneously optimizing the $\mathcal{U}$ and DIPS parameters. The DIPS loss allows to evaluate the standalone DIPS performance and helps optimise the related part of the network. The outputs of the $\mathcal{U}$ network are then combined in the D_b and D_c discriminant variables. Their distributions, evaluated on the $t\bar{t}$ simulated sample, are shown in Fig. 3.21.

3.2.5 Performance

The performance of jet flavour-tagging algorithms can be graphically represented through Receiver Operating Characteristics (ROC) curves. These curves plot the efficiency of correctly tagging signal jets against the rejection rates of background jets. These values are derived by selecting thresholds in the D_b and D_c distributions and then calculating the proportion of signal and background jets above it. These proportions are the flavour-tagging efficiencies ϵ_f of the algorithm at the threshold:

$$\epsilon_f = \frac{\text{number of jets of flavour } f \text{ above the threshold}}{\text{total number of jets of flavour } f}. \quad (3.4)$$

Each point on the curve corresponds to a specific threshold and, thus, to particular jet flavour-tagging efficiencies. The signal jet efficiency of a given threshold is also called the

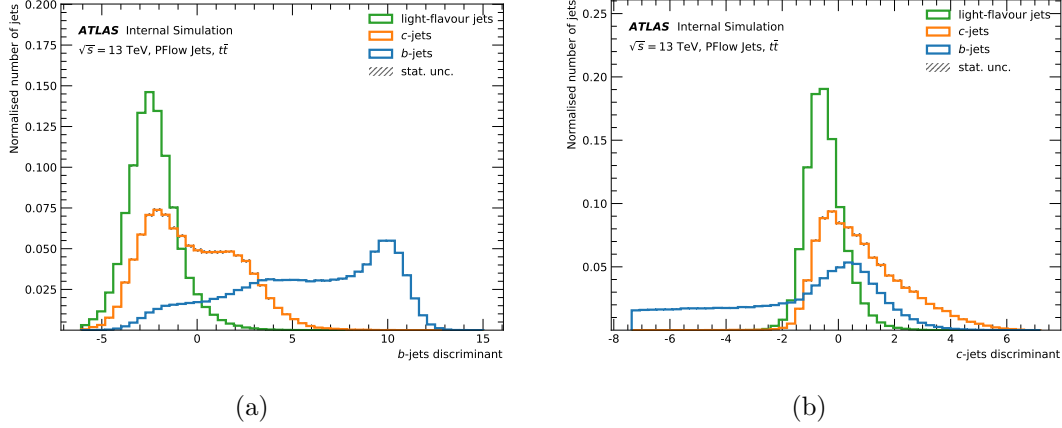


Figure 3.21: D_b and D_c distributions evaluated with the DL1d algorithm on $t\bar{t}$ simulated events.

Working Point (WP). The background rejection rates, displayed on the y axis, are computed as the inverse of the background efficiencies. A background rejection rate of r indicates that, in a random set of r background jets, on average, one is *mistagged*, meaning it is erroneously identified as signal.

Fig. 3.22 shows b -tagging ROC curves for the RNNIP, DL1r and DL1d algorithms evaluated on jets from the $t\bar{t}$ simulated sample. Fig. 3.22a compares the performance of DL1r and RNNIP evaluated on simulated Run-2 events. Two algorithms named SVKine and JFKine also appear in the plot. These are sub-networks of DL1 that I developed to assess the standalone jet flavour-tagging performance of the SV1 and JetFitter algorithms. They are described in full detail in the next section. The ROC curves clearly illustrate the benefit of combining the low-level information through high-level algorithms: DL1r achieves background rejections up to 4 times larger than the standalone low-level taggers. Fig. 3.22b compares the performance of DL1r and DL1d evaluated on simulated Run-3 events. DL1d surpasses the background rejection rates achieved with DL1r of factors up to 30%.

3.3 Developing SV1 with SVKine (and JFKine)

A consequence of the two-stage approach characterizing the DL1 algorithms is that all the low-level tagger algorithms must be maintained when a change occurs in one of the subsequent steps involved in the reconstruction of the detector-level and physics objects that they use.

When I joined the ATLAS Collaboration, a major upgrade of ATHENA was ongoing, involving substantial changes in the object reconstruction procedure. The effect of these changes on the jet flavour-tagging algorithms had to be assessed. When deemed necessary,

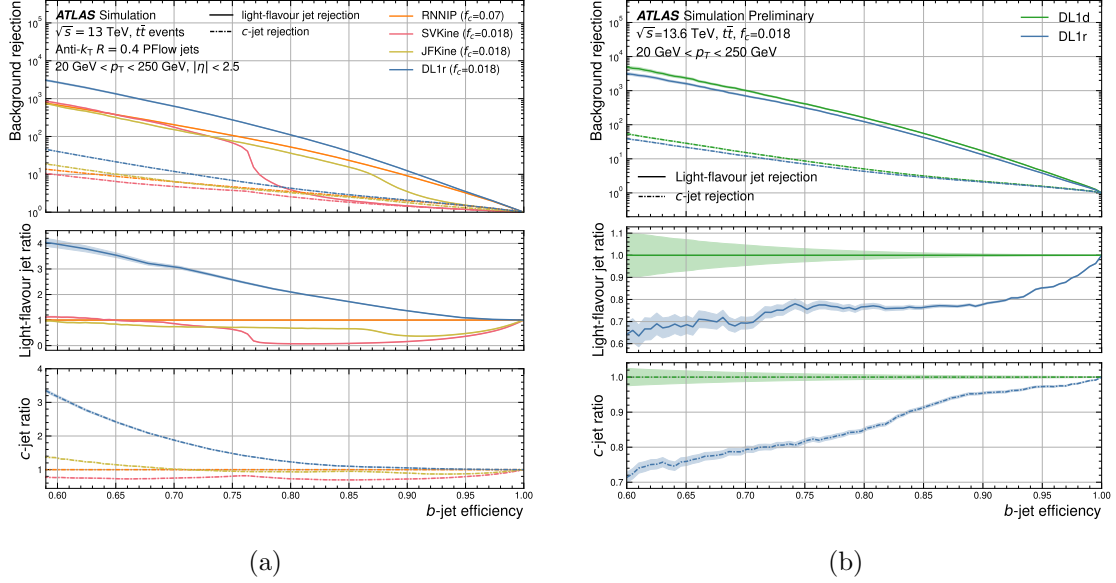


Figure 3.22: (a) b -tagging ROC curves of the DL1r, RNNIP, SVKine and JFKine algorithms evaluated on simulated jets from the $t\bar{t}$ sample. The statistical uncertainties of the rejection rates are calculated using binomial uncertainties and are indicated as coloured bands. The lower two panels show the ratio of the light-flavour jet rejection and the c -jet rejection of the algorithms to RNNIP. SV1 and JetFitter have secondary-vertex-finding efficiencies of approximately 80% and 90%, respectively; this causes the rapid growth of their light-flavour jet rejection rates around these values. Figure produced by me and published in [66]. (b) The light-flavour jet (solid line) and c -jet (dashed line) rejection rates for the DL1r and DL1d algorithms, evaluated on $t\bar{t}$ events simulated with the LHC Run-3 conditions. From [65].

a re-optimization of the parameters defining the flavour-tagging algorithms was performed.

SVKine and **JFKine** were developed in this context. They are NNs trained only on the variables listed under the sections **Kinematics**, **SV1** and **JetFitter** of Tab. 3.1. Their scope was to guarantee a fast turnaround in the optimization process of **SV1** and **JetFitter** for the new **ATHENA** release. The hyperparameters shaping the architecture of the NNs and defining their training campaign are enumerated in Tab. 3.3.

Hyperparameter	Value
Number of input variables	10/17 SVKine / JFKine
Number of hidden layers	4
Number of nodes [per layer]	[36, 24, 12, 6]
Learning rate	0.01
Training batch size	15 000
Activation function	ReLU
Number of training epochs	200
Training sample size	9.5 M jets

Table 3.3: List of hyperparameters that shape the architecture of the **SVKine** and **JFKine** algorithms and define their training campaign.

SVKine and **DL1r** were used in concert to test the benefits of relaxing the cut on **BVrtScore**, applied in the workflow of the **SV1** algorithm to discard two-track vertices not compatible with the decay of an heavy-flavour hadron (Sec. 3.2.3). The value of such cut, historically, was tuned before developing the **DL1r** algorithm. Relaxing it produces two concurring effects: an increased proportion of reconstructed “good” secondary vertices, composed of tracks emerging from the decay of heavy-flavour hadrons, at the cost of increasing also the proportion of “bad” secondary vertices, composed of tracks produced in other processes. The optimization aimed at identifying the value of the cut that would maximise the jet flavour-tagging performance of **DL1r**.

Two WPs of the Single Secondary Vertex Finding (SSVF) step of the **SV1** algorithm, referred to as *Standard* and a *Loose*, were defined. They correspond to cut values on **BVrtScore** (Eq. 3.2) of 0.015 and 0.005, respectively. Fig. 3.23 illustrates the features of the secondary vertices reconstructed with the *Loose* and *Standard* **SV1** WP, as well as those reconstructed by **JetFitter**. Three figures of merit were proposed to characterize the quality of the vertexing performance, all relying on the true information about the process producing the simulated reconstructed tracks:

- the *recall*: the ratio of tracks associated to the secondary vertex and produced by decay products of heavy-flavour hadrons, over the total number of tracks produced by decay products of heavy-flavour hadrons;

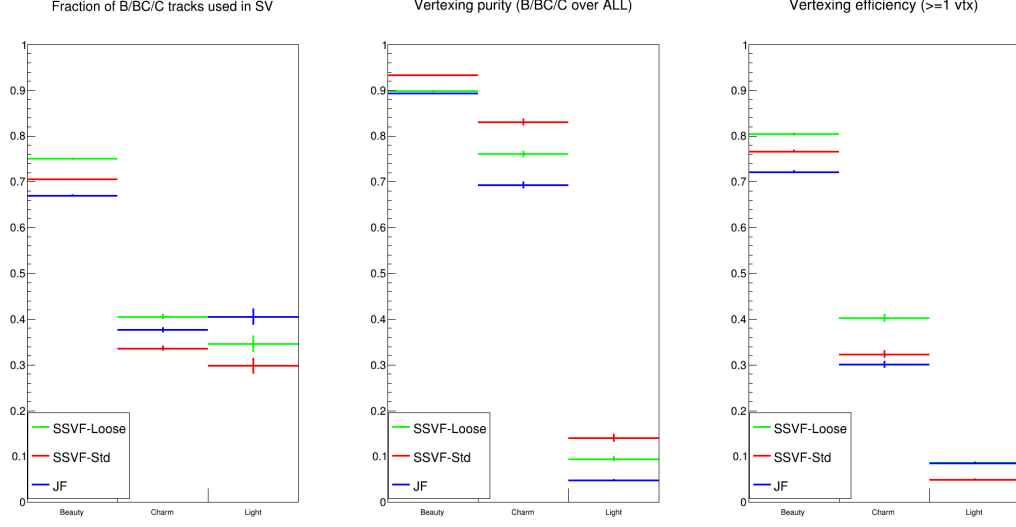


Figure 3.23: Recall, vertex purity and vertexing efficiency of secondary vertices reconstructed with the *Loose* and *Standard* working points of the SV1 algorithm, and with *JetFitter*. In this figure, B, BC and C tracks indicate tracks known at the truth level to originate from the decay of b -hadrons, c -hadrons produced in b -hadrons decays, and other types of c -hadrons respectively. JF stands for *JetFitter*. Loosening the cut on *BVrtScore* increases recall and vertexing efficiency for all the jet flavours and decreases vertex purity.

- the *vertex purity*: the ratio of tracks associated to the secondary vertex and produced by decay products of heavy-flavour hadrons, over the total number of tracks associated to the secondary vertex;
- the *vertexing efficiency*: the fraction of jets with at least a reconstructed secondary vertex.

Choosing the *Loose* WP increases recall and vertexing efficiency for all the jet flavours, but reduces vertex purity.

The increase in vertexing efficiency achieved when loosening the cut is balanced by the lower quality of the reconstructed vertices. “Good” secondary vertices were defined as those formed by at least 50% of tracks from the decays of heavy-flavour hadrons. Figure 3.24 shows the “good” vertexing efficiency as a function of the *BVrtScore* cut value for the different jet flavours. The efficiency measured in light-flavour jets decreases more for tighter cuts when compared to heavy-flavour jets.

SVKine was trained multiple times utilizing the SV1 variables obtained when applying different cut values on *BVrtScore*. This was done to assess how the interplay between varying

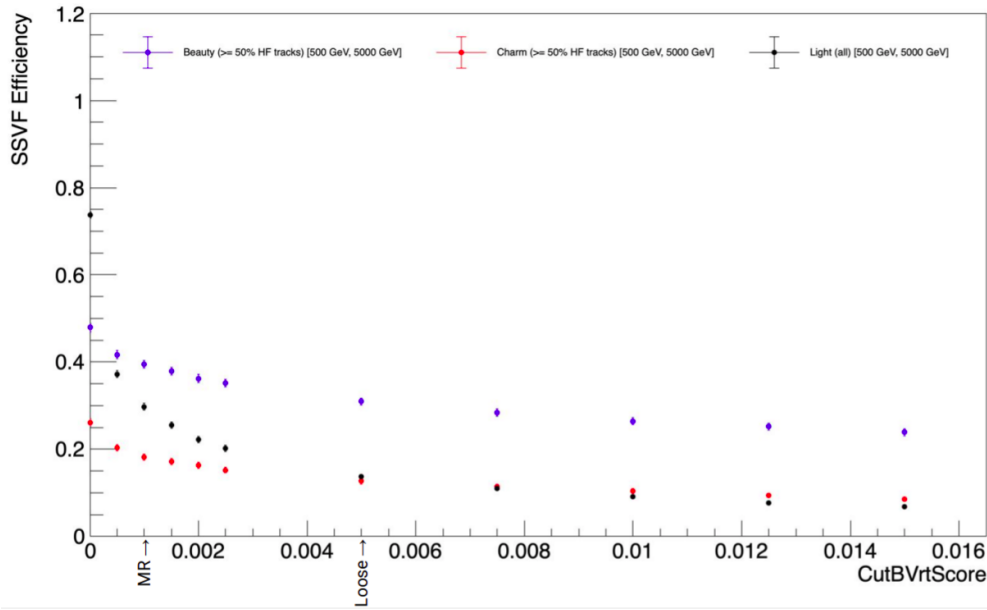


Figure 3.24: “Good” vertexing efficiency, *i.e.*, vertexing efficiency for vertices composed of at least 50% of tracks stemming from the decay of a heavy-flavour hadron, as a function of the BVrtScore cut value in b (blue), c (red) and light-flavour jets (black). The Standard and Loose WPs correspond to cut values of 0.015 and 0.005, respectively.

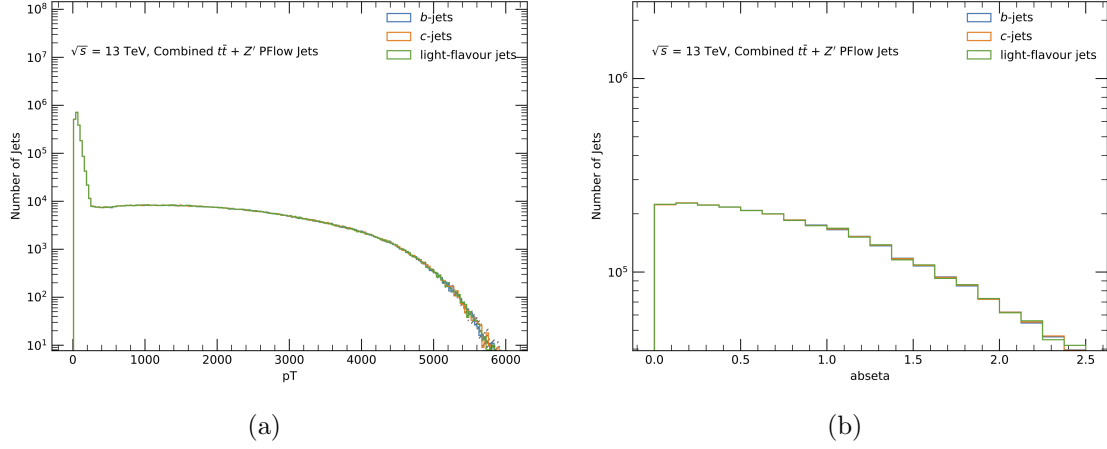


Figure 3.25: p_T and $|\eta|$ distributions of jets in the hybrid $t\bar{t}$ and Z' sample used to train **SVKine** and **JFKine**. The c -jets and light-flavour jets spectra are resampled to match the b -jets'.

vertexing efficiency, purity and recall affects the jet flavour-tagging information contained in the reconstructed vertices, when harnessed by a NN capable of exploiting the correlations among the various vertex features. The training is performed on a hybrid $t\bar{t}$ and Z' sample, resampled to match the p_T and $|\eta|$ distributions of c -jets and light-flavour jets to the b -jets' (Fig. 3.25).

Fig. 3.26 shows the b -tagging ROC curves computed with the different versions of **SVKine**, when it is evaluated on the simulated $t\bar{t}$ and Z' samples. The mark of the concurring effects obtained when varying the **BVrtScore** cut is clearly visible in the behaviour of the light-flavour jet rejection rate as a function of the b -tagging efficiency. ROC curves evaluated using tighter cut values achieve better light-flavour jet rejection rates at lower efficiency values. However, these rates drop sooner when approaching higher efficiencies compared to algorithms using looser cut values. These drops in rejection rates at increasing efficiencies for loosening **BVrtScore** cut values, particularly evident when looking at the curve evaluated on the $t\bar{t}$ sample, indicate that when **SV1** is not capable of providing secondary vertices, no discriminant information is available. In other words, when **SVKine** is evaluated at a b -tagging efficiency greater than the proportion of b -jets in which a secondary vertex is reconstructed, it identifies as b -jets also those with no reconstructed secondary vertex, which cannot be distinguished from the light-flavour jets by the algorithm. This is clear when comparing **SVKine**'s D_b distributions evaluated on the $t\bar{t}$ sample using the standard cut of **BVrtScore** = 0.015 (Fig. 3.27a) and using no cut at all (Fig. 3.27b). The former figure displays distributions in which two distinct zones can be identified. Jets with reconstructed secondary vertices are found at D_b values larger than 0. This region is populated mostly by b -jets, followed by c -jets and a very small amount of light-flavour jets. Jets with no reconstructed secondary vertex populate the region with $D_b < 0$ instead, which is dom-

inated by light-flavour jets. b -tagging WPs of **SVKine** with ϵ_b values greater than those corresponding to the drops in the light-flavour jet background rejection rates use thresholds set in the $D_b < 0$ zone. Any of these thresholds gives a definition of b -tagged jet which applies to significant fractions of the background jets as well, greatly deteriorating the rejection rate. In Fig. 3.27b, no **BVrtScore** cut is imposed on the reconstructed secondary vertex. The result is that most of the background jets contain at least a secondary vertex as well (Fig. 3.24). In this case, the D_b distributions are more similar for the different jet flavours and the separation between the zones of the distributions populated by jets with and without reconstructed vertices is less evident.

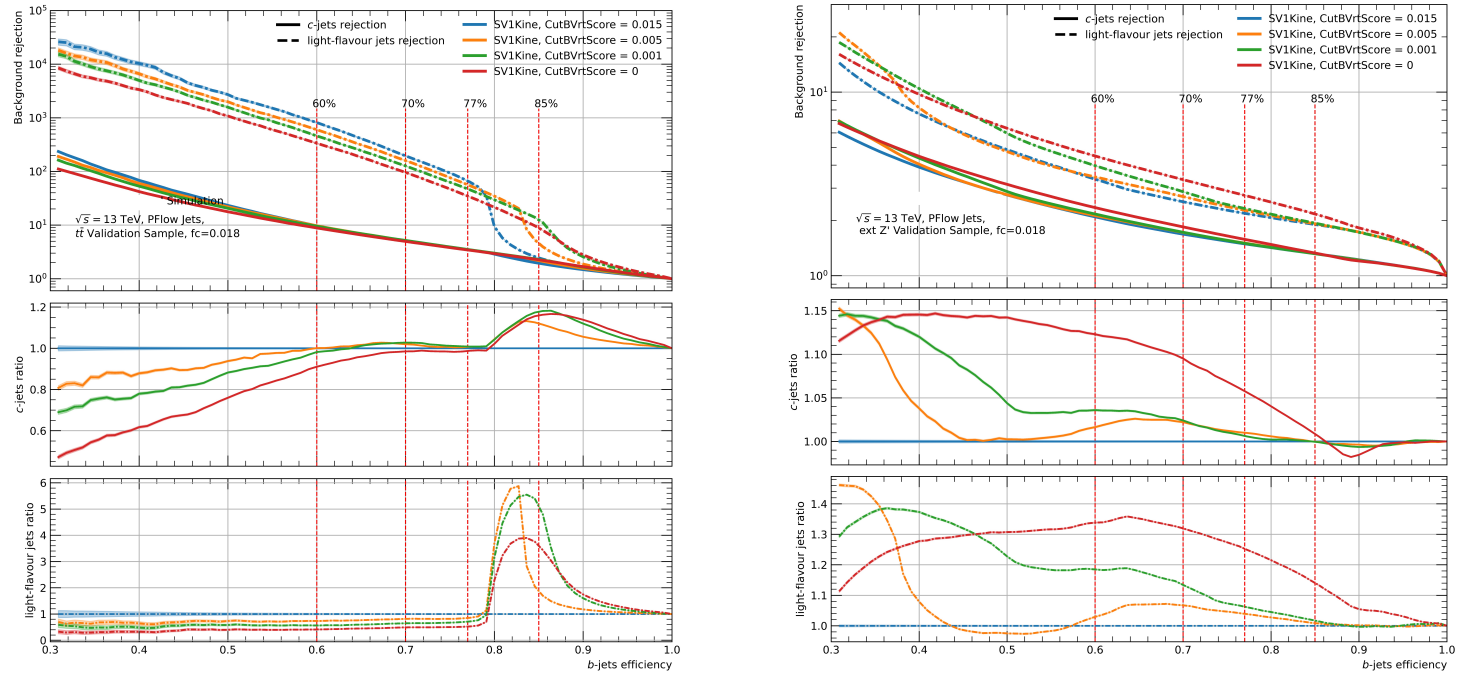


Figure 3.26: ROC curves computed with the different versions of **SVKine** on the simulated $t\bar{t}$ (left) and Z' (right) sample. The drops in background rejections are located at the b -tagging efficiency values ϵ_b corresponding to the proportion of b -jets in which **SV1** reconstructs a secondary vertex.

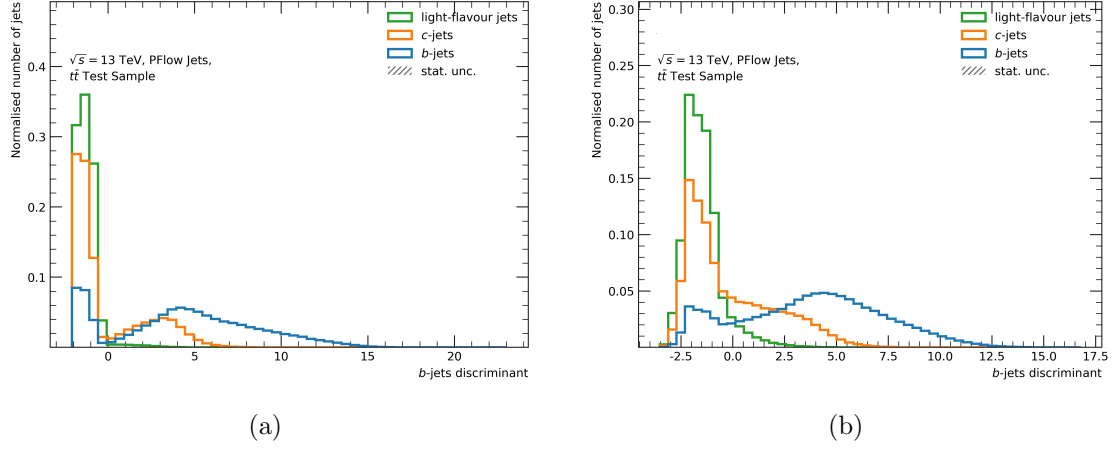


Figure 3.27: $SVKine$'s D_b distributions evaluated on a $t\bar{t}$ sample using the standard cut of $BVrtScore = 0.015$ (a) and using no cut at all (b).

Following the characterisation of the vertices reconstructed with the different $BVrtScore$ cut values and of their impact on the $SVKine$ performance, $DL1r$ was retrained with the various cut values as well. This was done to understand whether the final high-level tagger benefited more from a higher vertexing efficiency or from better vertex purity. The training was done on the same hybrid sample used to train $SVKine$ and $JFKine$, using the same parameters as of Tab. 3.2. Fig. 3.28 shows the ROC curves of the resulting NNs when evaluated on the $t\bar{t}$ and Z' sample. Loosening the $BVrtScore$ cut values increases the background rejection rates only for ϵ_b values smaller than 60% and larger than 85%, which are respectively the tightest and loosest $DL1r$ b -tagging WPs used in ATLAS physics analyses. This aligns with what is shown in Fig. 3.22a: in the 60% to 85% ϵ_b region, the discrimination is mostly driven by RNNIP.

Considering all these studies, it was decided to maintain the *Standard* cut on $BVrtScore$.

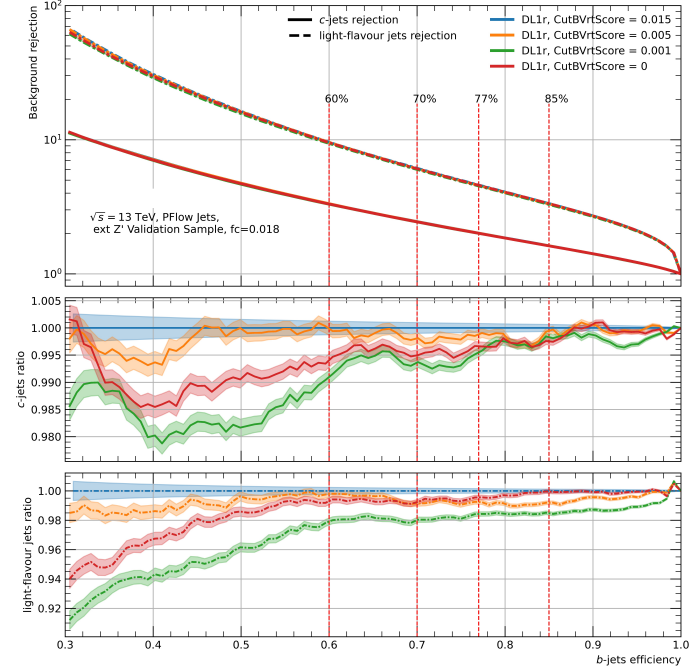
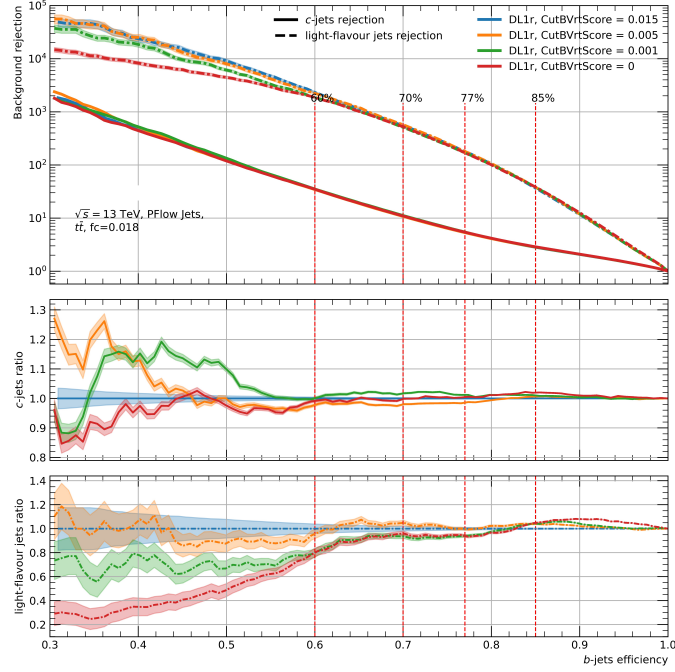


Figure 3.28: ROC curves of the DL1r algorithm trained using different BVrtScore cut values in the SV1 workflow, as evaluated on the $t\bar{t}$ (a) and Z' (b) simulated samples.

3.4 GN1 and GN2: jet flavour-tagging using GNNs

The introduction of GNNs allows to abandon the two-stage low-level to high-level tagger paradigm. GN1 [77] and its successor GN2, the ATLAS state-of-the-art jet flavour-tagging algorithm, are monolithic algorithms that operate directly on tracks. With a single training, they are capable of modelling the internal structure of tracks and vertices inside the jets and predicting their flavours. This is enabled by including auxiliary training objectives, namely the classification of the origin of the tracks and the pairing of those originating from the same vertex. Auxiliary tasks guide the network towards an understanding of the physics features that differentiate the jet flavours. In doing so, they effectively replace the low-level algorithms. Fig. 3.29 compares the design concepts of two-step high-level taggers such as DL1r and monolithic taggers such as GN1.

3.4.1 The GN1 algorithm

GN1 is the first GNN-based jet flavour-tagging algorithm introduced in ATLAS. It operates on fully connected graph representations of jets, with the tracks associated to the jets serving as the nodes.

In the graph, tracks are represented as vectors of variables. Tab. 3.4 enumerates the variables used to train GN1 and GN1Lep. The latter is a version of GN1 also using the leptonID variable indicating whether a track was used to reconstruct an electron or muon, which helps leverage the semileptonic decays of heavy-flavour hadrons. The p_T and η of the jets, which are global features of the graphs, are concatenated to each track along with the other variables.

The architecture of the GN1 algorithm is schematized in Fig. 3.30. Tracks are fed into an initializing NN, transforming their initial vector representation into a higher-dimensional one. This first step is very similar to the initialization of tracks provided by the DIPS block to the $\mathcal{U}$ network in the workflow of DL1d (Sec. 3.2.4). The obtained vectors serve as nodes to build a fully connected graph, in which each node neighbors every other node. A GNN then operates on the nodes as follows. Each node representation h_i is updated by aggregating its features and those of his $\mathcal{N}_i$ neighbouring nodes¹² using a $n = 2$ multi-head attention mechanism [78, 79]. First, the *receiver* node h_i and each of the neighbouring $\mathcal{N}_i$ *sender* nodes h_j are fed to the fully-connected linear layers $\mathbf{W}_r$ and $\mathbf{W}_s$, that generate the updated representations $\mathbf{W}_r(h_i)$ and $\mathbf{W}_s(h_j)$. These vectors are split into two parts of equal dimensions, $\mathbf{w}_r(h_i)^l$ and $\mathbf{w}_s(h_j)^l$, where l indexes the two heads of the multi-head attention mechanism. They are used to compute edge scores $e^l(h_i, h_j)$ for each node pair according to the formula

$$e^l(h_i, h_j) = \mathbf{a}^l \cdot (\theta\{\mathbf{w}_r(h_i)^l + \mathbf{w}_s(h_j)^l\}). \quad (3.5)$$

¹²Since the graph is fully-connected, including self-connections of individual nodes, $\mathcal{N}_i$ is always equal to the number of tracks in the jet.

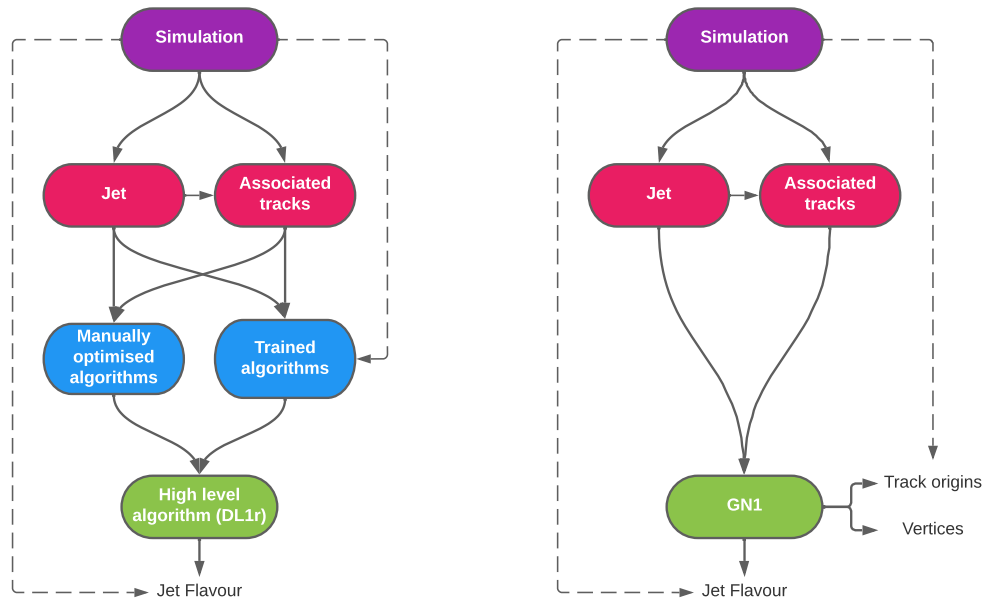


Figure 3.29: Comparison of the DL1r (left) and GN1 (right) design [77]. The first approach uses low-level algorithms (shown in blue) to produce outputs fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimized low-level algorithms, additional truth information from the simulation is used as auxiliary training targets for GN1 instead. The solid lines represent reconstructed information, whereas the dashed lines represent truth information.

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
$d\eta$	Pseudorapidity of the track, relative to the jet η
$d\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(d_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used to reconstruct an electron or muon (only for GN1Lep)

Table 3.4: Input features to the **GN1** model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. The track leptonID is an additional input to the **GN1Lep** model.

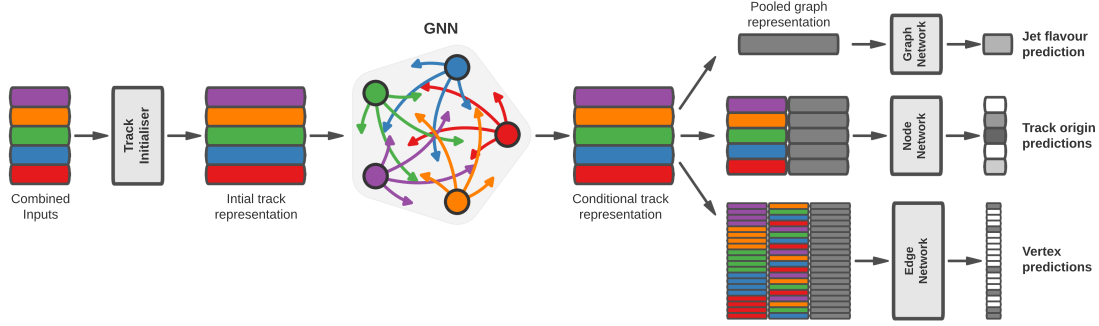


Figure 3.30: The network architecture of GN1 [77]. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the probability of each track pair belonging to a common vertex.

Here, θ is a non-linear activation function, and $\mathbf{a}^l$ is a learned vector. The edge scores $e^l(h_i, h_j)$ of each node i are combined into a vector of dimension $\mathcal{N}_i$. They are used to compute attention weights a_{ij}^l through the softmax function, which forces the sum of their components to add up to 1:

$$a_{ij}^l = \text{softmax}_j \{e^l(h_i, h_j)\}. \quad (3.6)$$

The updated node representation $h_i'^l$ is the sum over the $\mathcal{N}_i$ vector representations $\mathbf{w}_s(h_j)^l$ of the neighbouring nodes, weighted by the attention weights and passed through a sigmoid function

$$h_i'^l = \sigma \left\{ \sum_{j \in \mathcal{N}_i} a_{ij}^l \mathbf{w}_r(h_j)^l \right\}. \quad (3.7)$$

This set of operations defines a single layer of the GNN. GN1 uses three of such layers. For the first two layers, the two representations of each node, $h_i'^1$ and $h_i'^2$, associated with the two different heads of the attention mechanism, are concatenated into a single vector

$$h_i' = h_i'^1 \oplus h_i'^2, \quad (3.8)$$

that serves as input to the following layer. For the last layer, they are averaged instead:

$$h_i' = \frac{1}{2} (h_i'^1 + h_i'^2). \quad (3.9)$$

The representation of each node is then concatenated to its representation as it was before processing through the GNN layers. The resulting vectors are then combined in three

Network	Hidden layers	Output size
Node classification network	128, 64, 32	7
Edge classification network	128, 64, 32	1
Graph classification network	128, 64, 32, 16	3

Table 3.5: A summary of **GN1**’s different classification networks used for the different training objectives. The hidden layers column contains a list specifying the number of neurons in each layer.

Truth Origin	Description
Pileup	From a pp collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a b -hadron
fromBC	From a c -hadron decay, which itself is from the decay of a b -hadron
fromC	From the decay of a c -hadron
OtherSecondary	From other secondary interactions and decays

Table 3.6: Truth origins which are used to categorise the physics process that led to the production of a track.

different ways to provide the input to the NNs that implement the jet-flavour prediction and the auxiliary tasks. Their parameters are listed in Tab. 3.5.

The jet-flavour prediction is a graph classification task. The representations of all the nodes are combined via a global attention-pooling procedure. Each node is processed through a linear layer $\mathbf{G}$, which provides a scalar output. These outputs serve as attention weights for the tracks, which are combined via a weighted sum. The resulting vector is the global jet representation. It is fed to a NN with three scalar outputs, representing p_b , p_c and p_u .

The track classification task is a node classification one. Each node representation is concatenated to the global jet representation and fed to a NN whose scalar outputs are the probabilities assigned to the track of belonging to one of the classes defined in Tab. 3.6.

The two-track vertex pairing is a link prediction task. The vector representations of each pair of tracks are concatenated with the global jet representation, and fed to a NN. The network assigns a scalar output to the pair, representing the combined probability that the two tracks originate from the same point of space and that none of the two is Fake or from Pile-up.

GN1 is trained to minimize the total loss function

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}, \quad (3.10)$$

where L_{jet} and L_{track} are categorical cross-entropy loss functions over the jet flavours and the track origins respectively, and L_{vertex} is a binary cross entropy loss over the track-pair labels. These labels have a value of 1 if the two tracks originate from the same point and none of them is either Fake or from PileUp, and 0 otherwise. The values $\alpha = 1.5$ and $\beta = 0.5$ are chosen to ensure that L_{jet} converges to a value larger than L_{vertex} and L_{track} . This choice reflects the priority given to the graph classification task above the auxiliaries.

Performance GN1 b and c -tagging performance largely surpass DL1r's. For instance, at a 70% b -tagging efficiency, GN1's light-flavour and c -jet rejection rates outperform DL1r's of factors of about 1.8 and 2.1 respectively when evaluated on simulated $t\bar{t}$ events, as shown in the ROC curves depicted in Fig. 3.31a. Large improvements are observed in the c -tagging ROC curves as well (Fig. 3.31b).

It is of particular interest to characterize the contribution of the auxiliary tasks to the algorithm's performance. To do so, GN1 was modified by enabling only the track classification task (GN1 TC), only the vertexing task (GN1 Vert), and by disabling both the auxiliary tasks (GN1 No Aux)). The resulting ROC curves are displayed in Fig. 3.32. The performance of (GN1 No Aux)) is very similar to that of DL1r. This means that the superior GNN architecture is, per se, sufficient to compensate for the absence of low-level algorithms. The improvements brought about by the auxiliary tasks are of the same scale, and both are significant. The fact that GN1 outperforms both GN1 TC and GN1 Vert demonstrates that the two auxiliary tasks provide different kinds of information to the model during its training.

3.4.2 The GN2 algorithm:

GN2 is an evolution of GN1, sharing a similar architecture. The main difference is in the attention mechanism, which is inspired by transformer encoders.

After initialization through a NN, the tracks are used to build a fully connected graph. A GNN then operates using a $n=8$ multi-head attention mechanism updating each node representation h_i by aggregating its features and those of its $\mathcal{N}_i$ neighbouring nodes. The vectors representing each node are normalized. Then, the receiver node h_i is fed to a fully connected linear layer $\mathbf{W}_q$, while each of the sender nodes is fed to two distinct linear layers $\mathbf{W}_k$ and $\mathbf{W}_v$ ¹³, that generate the updated representations $\mathbf{W}_q(h_i)$, $\mathbf{W}_k(h_j)$ and $\mathbf{W}_v(h_j)$. These vectors are split in 8 parts, $\mathbf{w}_q(h_i)^l$, $\mathbf{w}_k(h_j)^l$ and $\mathbf{w}_v(h_j)^l$, with l indexing the 8 heads of the multi-head attention mechanism. The receiver $\mathbf{w}_q(h_i)^l$ and the senders $\mathbf{w}_k(h_j)^l$ are used to compute edge scores $e^l(h_i, h_j)$ according to the formula

$$e^l(h_i, h_j) = \frac{\mathbf{w}_q(h_i)^l \cdot \mathbf{w}_k(h_j)^l}{\sqrt{s}} \quad (3.11)$$

¹³The suffixes q for the receiver node and k and v for the sender nodes are chosen to indicate the terms *queries*, *keys* and *values*, usually employed to name them in transformers.

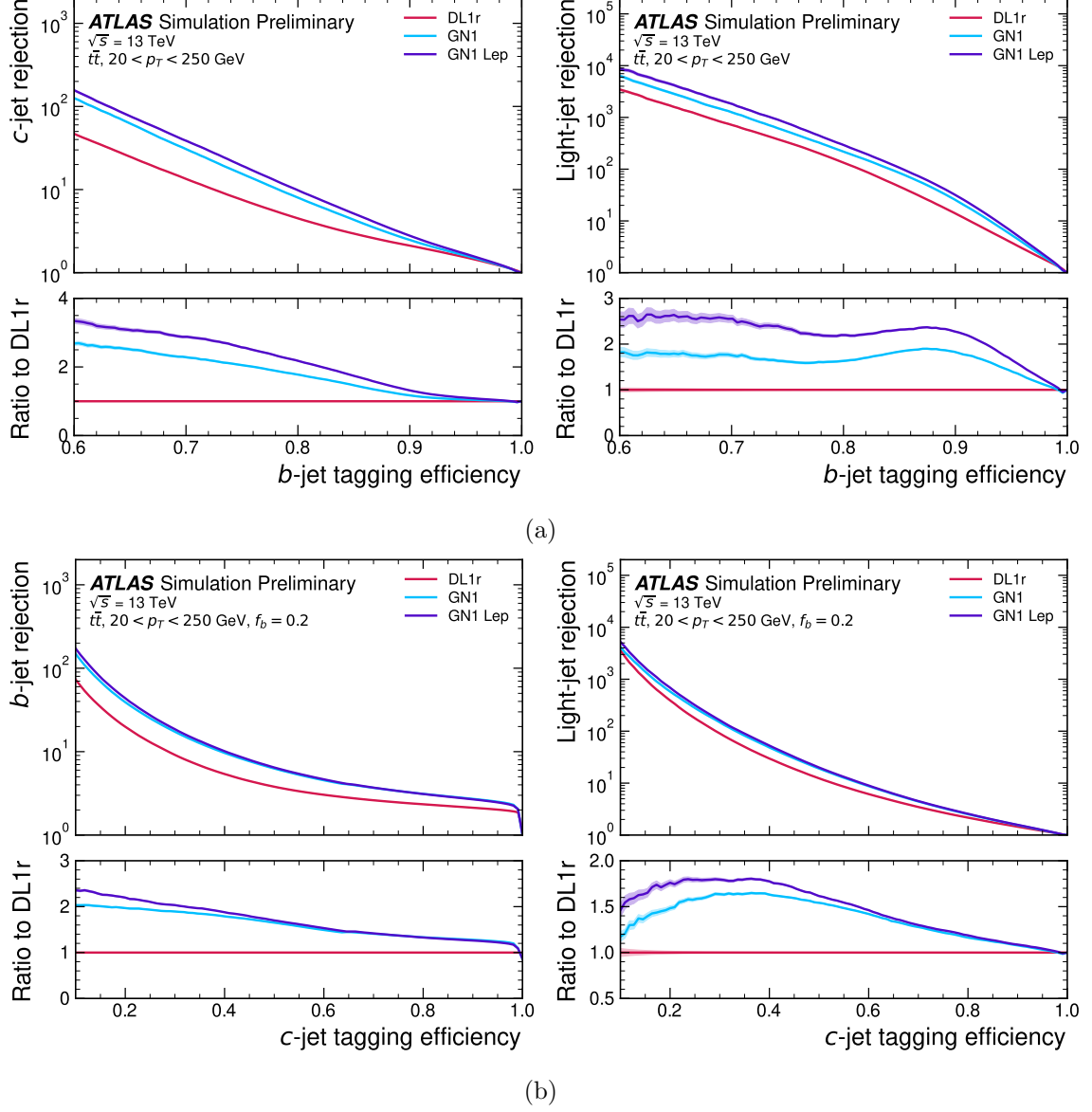


Figure 3.31: Flavour-tagging ROC curves of GN1, GN1Lep and DL1r evaluated on a simulated $t\bar{t}$ sample. The GNN-based algorithms largely outperform DL1r for all the signal efficiencies. From [77].

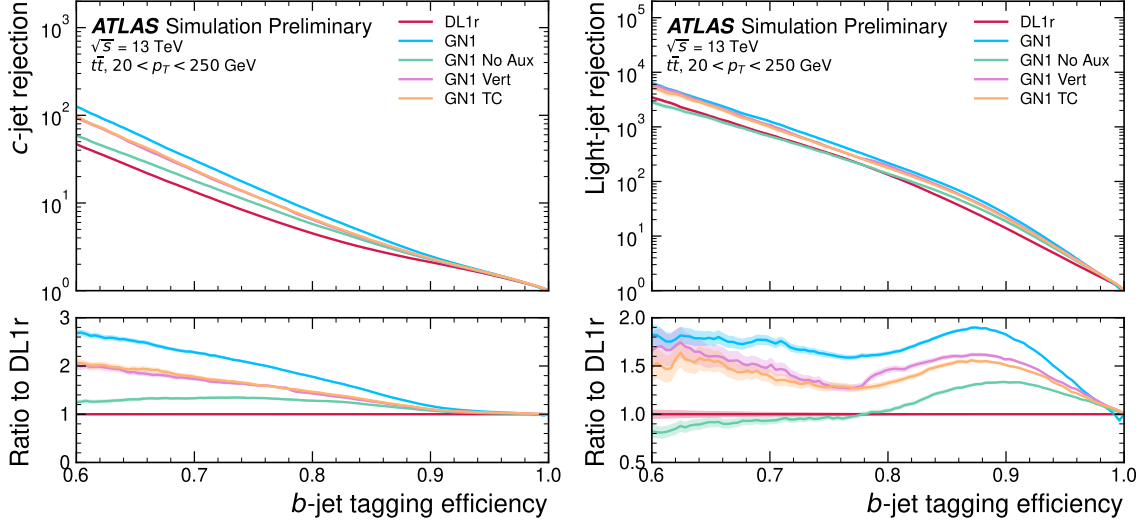


Figure 3.32: b -tagging ROC curves of DL1r and GN1 compared to three modified versions of the GN1 algorithm in which both the auxiliary tasks are disabled (GN1 No Aux), in which only the track classification task is enabled GN1 TC, and in which only the vertexing task is enabled GN1 Vert. From [77].

where the scale s is the dimension of the $\mathbf{w}$ vectors. The edge scores $e^l(h_i, h_j)$ of each node i are combined into a vector of dimension $\mathcal{N}_i$. They are used to compute attention weights a_{ij}^l through the softmax function, which forces the sum of their components to add up to 1:

$$a_{ij}^l = \text{softmax}_j \{e^l(h_i, h_j)\}. \quad (3.12)$$

The receiver node representation is updated to $h_i'^l$, the sum over the $\mathcal{N}_i$ vector representations $\mathbf{w}_v(h_j)^l$ of the neighboring nodes, weighted by the attention weights

$$h_i'^l = \sum_{j \in \mathcal{N}_i} a_{ij}^l \mathbf{w}_v(h_j)^l. \quad (3.13)$$

h_i' is obtained by concatenating the $h_i'^l$ vectors along the l index. h_i' is then normalized and summed to the normalized h_i . The resulting vector is used as input to a NN, $\mathcal{D}$, whose output vector is then summed with the input vector.

This set of operations defines a single layer of the GNN. GN2 employs six of such layers. After these, the node representations are normalized once more, passed through one last linear layer $\mathcal{L}$, and combined for the auxiliary tasks in the same way used by GN1.

Another innovation introduced by GN2 is the incorporation of stochastic dropout layers.

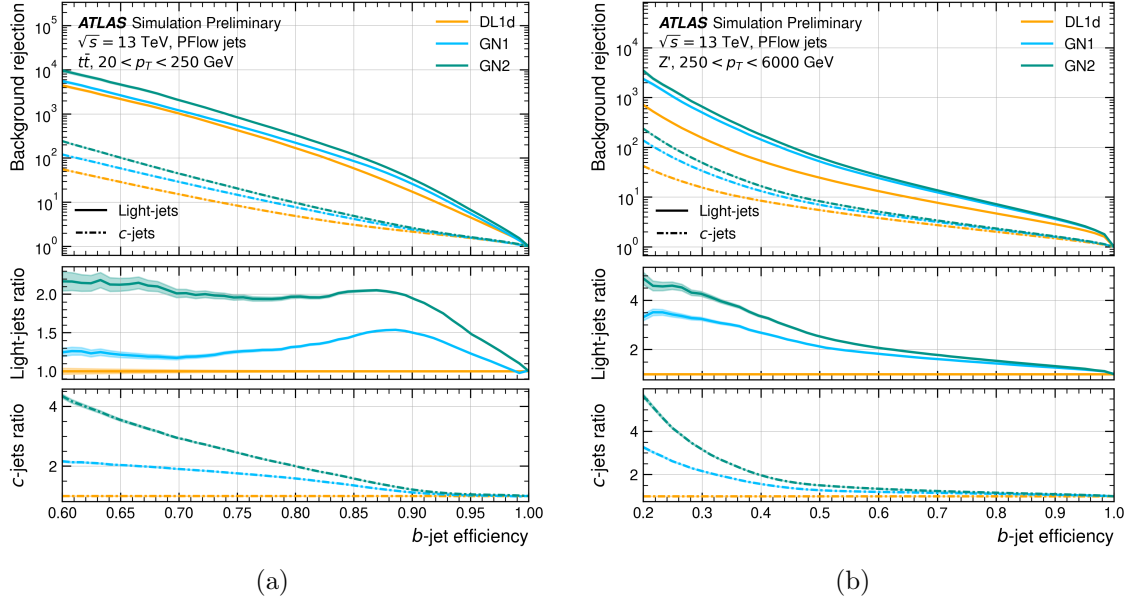


Figure 3.33: Flavour-tagging ROC curves of DL1d, GN1 and GN2 evaluated on a simulated $t\bar{t}$ (a) and Z' (b) sample. From [65].

These layers randomly zero out parameters throughout the architecture, aiming to mitigate the risk of overfitting.

Performance The flavour-tagging performance of DL1d, GN1 and GN2, evaluated on a $t\bar{t}$ simulated sample, are compared in Fig. 3.33. GN2 outperforms GN1 significantly.

Having described throughout this chapter all the algorithms entering the chart displayed in Fig. 3.34, it is now possible to fully appreciate the significant improvements brought about by the successive innovations introduced in the ATLAS jet flavour-tagging algorithms in recent years.

These improvements are measured in simulated MC events. In parallel with the development of new algorithms, calibration studies for characterizing their performance on real jets must be performed.

3.5 Calibration

To use the jet flavour-tagging algorithms in physics analyses, it is crucial to understand their performance on real data. Simulation deviates from data for detector and modelling effects. For instance, the dominating factors provoking the mistagging of light-flavour jets as heavy-flavour jets are the presence of fake tracks and the limited resolution of the detector

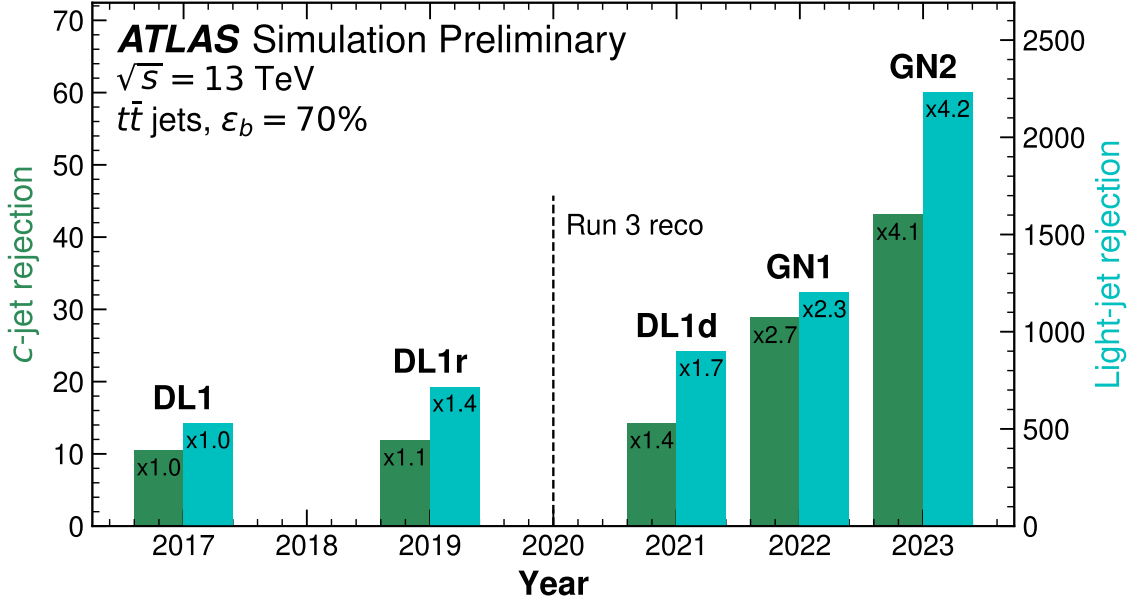


Figure 3.34: b -tagging performance of the subsequent algorithms developed within the ATLAS Collaboration in recent years. The background rejection rates underwent a fourfold improvement in a very limited time span. From [65].

when measuring tracks' IPs. When comparing simulation and data, the amount of fake tracks in the former is underestimated, and the detector's resolution is overestimated [80].

The performance of the jet flavour-tagging algorithms on real data is studied by calibrating the different per-flavour efficiencies ϵ_f on samples enriched in flavour-tagged jets of the given flavour f . The ϵ_f values expected from the simulations are then corrected in the analyses using the Scale Factors (SFs), defined as

$$\text{SF}_f^{\text{data-MC}} = \frac{\epsilon_f^{\text{data}}}{\epsilon_f^{\text{MC}}}, \quad (3.14)$$

extracted from real data via likelihood fits.

For b -jets, $\text{SF}_b^{\text{data-MC}}$ can be extracted in di-lepton $t\bar{t}$ events [81] (Fig. 3.35). Since top-quarks always undergo the $t \rightarrow Wb$ decay, selecting jets from $t\bar{t}$ events where both the W bosons decay leptonically results in a very pure collection of b -jets. To further suppress backgrounds from decays of Z bosons, the two leptons, of opposite charge, are required to have different flavours. Fig. 3.36 displays the $\text{SF}_b^{\text{data-MC}}$ extracted for the Run-2 DL1 algorithm for various p_T ranges. For the reasons detailed in Chap. 4, jet flavour-tagging performance strongly depends on the jets' p_T . This explains why it is important to characterize it on data for various p_T ranges.

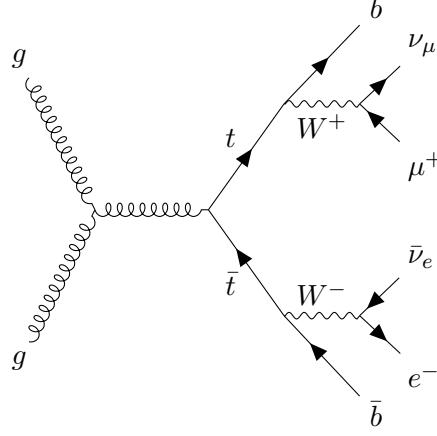


Figure 3.35: Tree-level Feynman diagram of a di-lepton $t\bar{t}$ event.

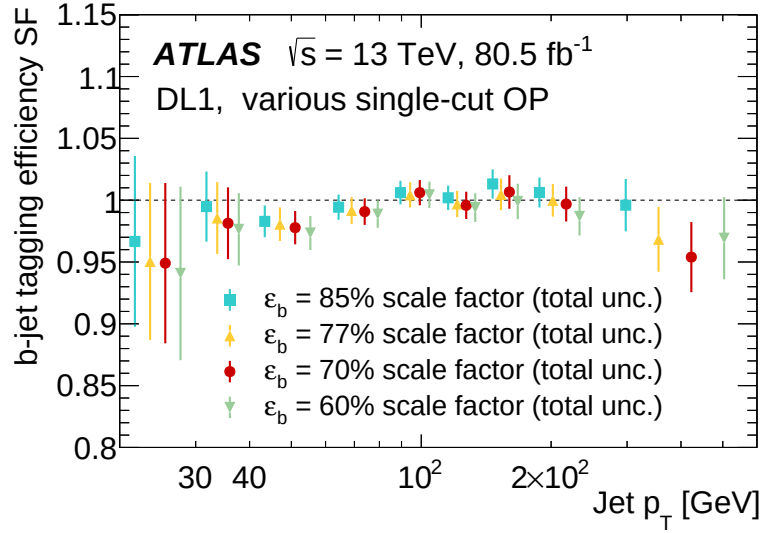


Figure 3.36: $\text{SF}_b^{\text{data-MC}}$ as a function of jet p_T , computed for the 60%, 70%, 77% and 85% WPs of the DL1 algorithm. From [81].

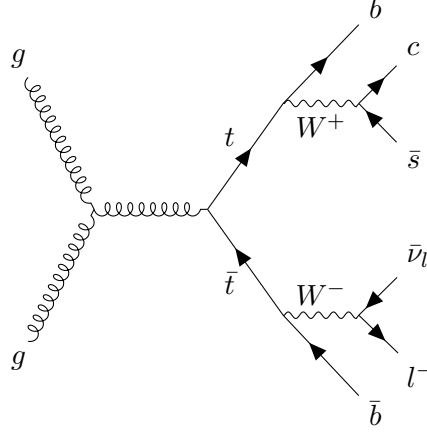


Figure 3.37: Tree-level Feynman diagram of a single-lepton $t\bar{t}$ event.

$\text{SF}_c^{\text{data-MC}}$ is extracted for c -jets using single-lepton $t\bar{t}$ events [82] (Fig. 3.37). These allow constructing a sample enriched in c -jets thanks to the relatively high branching ratio of $W \rightarrow cs$, of $\sim 33\%$. At tree-level, single-lepton $t\bar{t}$ decays contain four jets: two b -jets coming from the $t \rightarrow Wb$ decays, which are termed top-jets in this context, and two jets coming from the hadronic $W \rightarrow q\bar{q}'$ decay, the W -jets. A likelihood-based $t\bar{t}$ event reconstruction is performed to assign the jets to the top decays and to the decay of the W , by minimizing deviations in the invariant masses of the top quarks and W boson. Top jets are required to be b -tagged with the tightest 60% DL1r WP. Then, $\text{SF}_c^{\text{data-MC}}$ is computed in the fit by comparing in data and in the simulation the fraction of events where either of the jets coming from the W is tagged. Since $\epsilon_l \ll \epsilon_c$, these tagged jets are almost always c -jets.

Computing $\text{SF}_l^{\text{data-MC}}$ poses additional challenges due to the very low light-flavour jet mistagging efficiency ϵ_l . For instance, at the 85% and 60% WPs, DL1r rejects 97.52% and 99.96% of the light-flavour jets in a $t\bar{t}$ sample [80]. Hence, it is virtually impossible to collect a statistically significant sample enriched in b -tagged light-flavour jets using the standard techniques: the $\epsilon_b \gg \epsilon_l$ relation causes a strong dominance of b -jets in the sample.

The *Negative Tag Method* was introduced to target this problem. It consists of modifying the tagger to achieve similar light-flavour jets mistagging efficiency ϵ_l , but a much reduced ϵ_b and ϵ_c . The method allows to calibrate ϵ_l , provided that an additional scaling is applied to match its value for the modified tagger to that for the real tagger.

The method builds upon the shapes of the distributions of the signed IPs of tracks in heavy-flavour jets and light-flavour jets (Fig. 3.38). Whereas, in heavy-flavour jets, the distributions display a pronounced positive tail due to the tracks emerging from the displaced decays of the heavy-flavour hadrons, in light-flavour jets the IP values of the associated tracks are typically consistent with zero within the detector resolution. The resulting

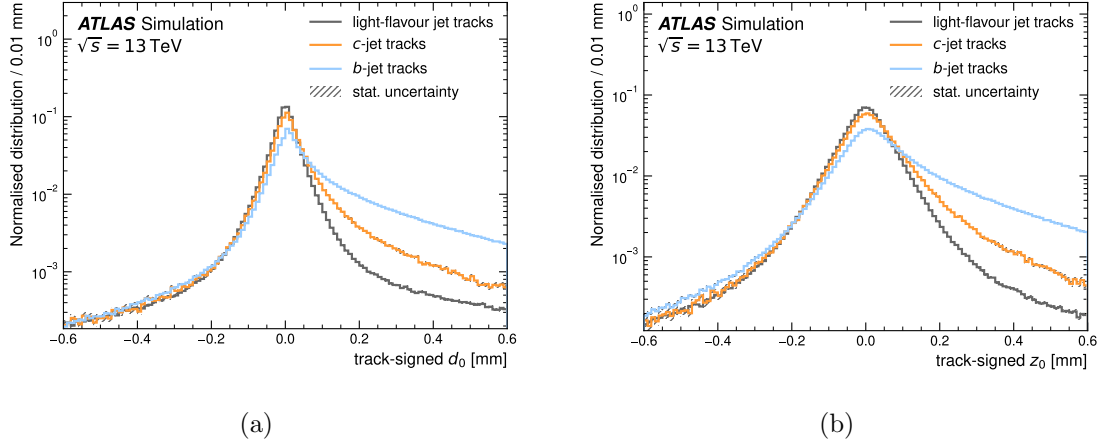


Figure 3.38: Signed- d_0 (a) and signed- z_0 (b) distributions for tracks matched to b -jets, c -jets and light-flavour jets in simulated $t\bar{t}$ events. From [80].

distribution is much more symmetric¹⁴. The Negative Tag Method is based on the assumption that most of the b -tagged light-flavour jets are misidentified due to resolution effects in the reconstruction of the tracks, which result in tracks with positive IPs grouped in vertices significantly displaced. Since the signed IP distributions of tracks associated to light-flavour jets are almost symmetric, inverting their sign results in an almost identical distribution. Hence, inverting the signs of all the IPs of tracks and all the decay lengths of the secondary vertices in light-flavour jets should not change their probability of being mistagged.

Flipped taggers can be obtained by inverting these signs when training and using the algorithms described in Sec. 3.2. To calibrate ϵ_l , a flipped version of DL1r, called DL1rFlip, was trained using the same hyperparameters as defined in Tab. 3.2. It achieves an ϵ_l value similar to DL1r, but much reduced ϵ_b and ϵ_c , hence meeting its design goals. Fig. 3.39 compares the D_b outputs of DL1r and DL1rFlip, evaluated on a $Z + \text{jets}$ sample (Fig. 3.40).

DL1rFlip is calibrated via a fit to the distribution of the invariant mass of SV1 secondary vertices obtained on a sample of jets constructed by selecting the leading- p_T jets originating from $Z + \text{jets}$ events. Such a sample is expected to contain a 7%–8% proportion of b -jets and a 4%–6% proportion of c -jets. Given the reduced DL1rFlip ϵ_b value, such contamination is small enough to calibrate the light-flavour jets mistagging efficiency.

¹⁴The residual positive tail in light-flavour jets is due to the presence of long-lived particles such as K_s and Λ .

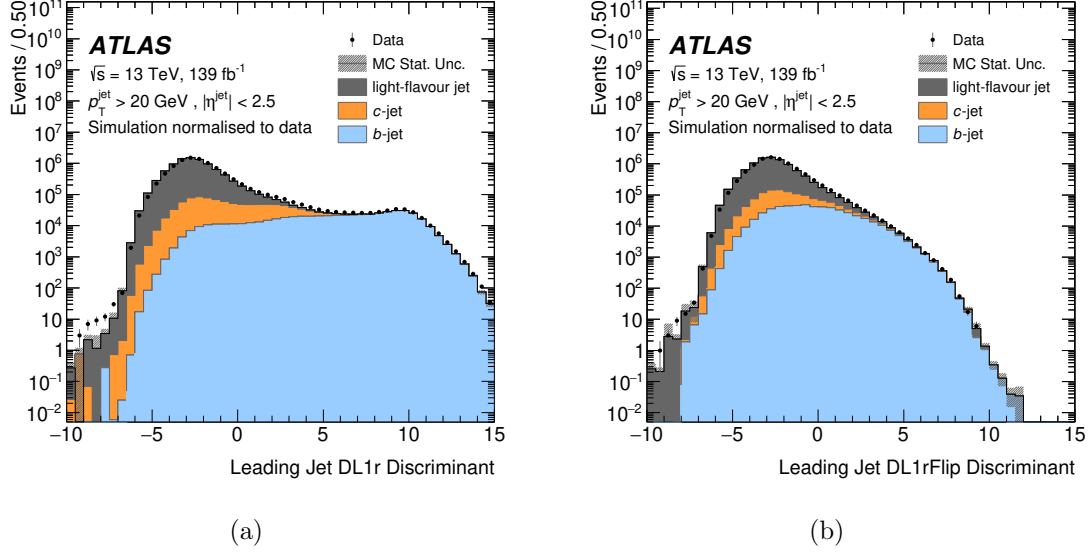


Figure 3.39: Comparisons between data and simulation for the distributions of the DL1r discriminant (a) and DL1rFlipped discriminant (b). From [80].

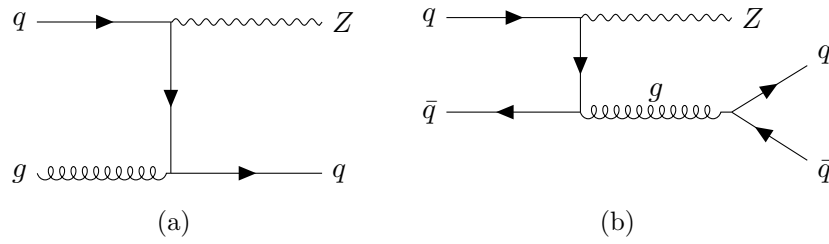


Figure 3.40: tree-level Feynman diagrams of Z plus one jet (a) and Z plus two jets (b) events.

3.6 Offline jet flavour-tagging DQ

I have been coordinating the activities of the offline jet flavour-tagging DQ group, together with a colleague, since the beginning of Run-3. The group's scope is to provide fast feedback about the status of the detector components by looking at the aspects of their combined performance related to jet flavour-tagging.

The quality assessment is made by means of 1D and 2D distributions of the most relevant objects for the jet flavour-tagging operations. In the 1D case, the distributions obtained in a given run are compared to those from a reference run already included in a GRL. The integral of the reference distribution is normalized to that of the new data, and a Kolmogorov test is performed to quantify how the two samples are likely to be drawn from the same underlying distribution. 2D distributions are not compared to any reference. Rather, they are monitored to identify local deviations from the surrounding yields in specific zones of the detector.

The events and jets to monitor are subject to a selection that aims at isolating as many non pile-up PFlow jets with good tracking content as possible. The goal of the selection is to build a dataset on which a precise assessment of the jet flavour-tagging algorithm's behaviour is possible.

Object selection: Events are required to have at least a PV, with at least 4 associated tracks. Additionally, they are required to contain at least an isolated electron or muon. This second requirement helps reduce pile-up dependency, as events with very high pile-up activity are less likely to contain isolated objects.

Jets in the selected events are then filtered based on their kinematic features, via a $p_T > 20$ GeV and $|\eta| < 2.5$ cut. They must have a sufficient JVT score, to reduce those produced in pile-up collisions. Finally, jets containing isolated muons or muons carrying more than 90% of the jet energy are discarded. Muons in b -jets are usually soft muons from semileptonic decays of b -hadrons. Also, the environment inside hadronic jets is rich in activity because of the hadronisation cascade of the colour charges. The cuts aim at excluding from the dataset prompt leptons reconstructed as jets. Selected jets are finally divided into three categories: *good*, *suspect* and *bad* jets, based on the quality of their tracking content. *Good* jets are those used for standard DQ assessment, whereas distributions of *suspect* and *bad* jets are rarely scrutinized for diagnostic. Fig. 3.41 depicts the number of events and jets discarded at every subsequent cut in the Run-3 run number 436550.

DL1d output: Among the monitored distributions, the D_b distribution of the DL1d tagger stands out for its importance. An example is showcased in Fig. 3.42, where the tagger evaluated on the *good* jets inclusively. The same distribution is monitored in bins of p_T , η , ϕ , and average pile-up $\langle\mu\rangle$. Given the energy of colliding protons and the pile-up conditions, the relative amount of jets produced with the different flavours is expected to be stable from

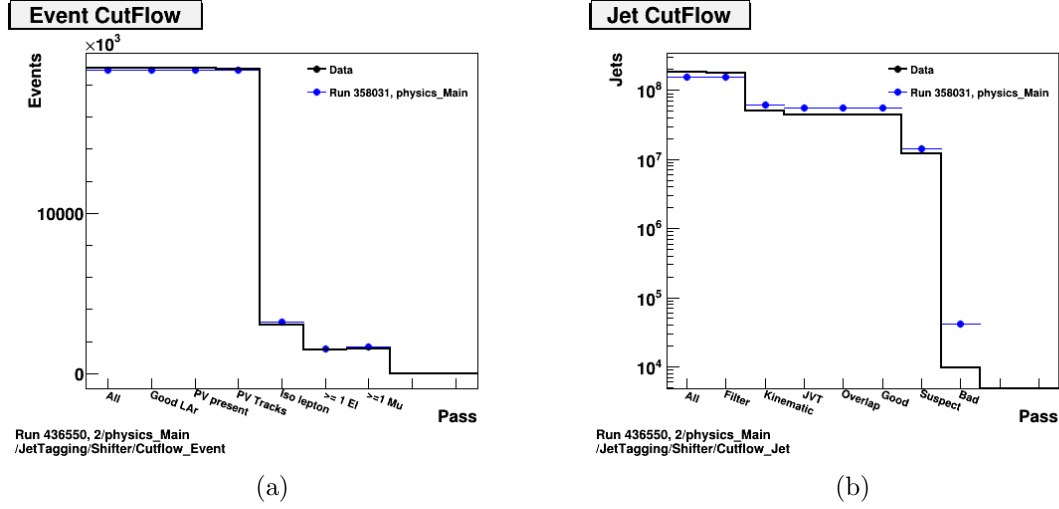


Figure 3.41: Offline jet flavour-tagging DQ sequence of cuts on events (a) and jets (b).

run to run.

2D maps (η, ϕ) distributions of reconstructed physics objects can help identify modules of the detector temporarily faulting. Fig. 3.43 shows a set of 2D distributions affected by the faulting of the L0.B2 pixel module of the ID B-layer¹⁵ during run 436550. CP and sub-detector groups continuously interact to identify issues of this kind and understand how they affect the reconstruction of physics objects.

$t\bar{t}$ events To construct a sample of events enriched in b -jets in each run, an additional selection is applied, requiring the presence of an isolated muon and an isolated electron with opposite signs in the event. Distributions built from these events are particularly stable with respect to issues arising from unusual noise in the calorimeters. For instance, Fig. 3.44 shows the D_b distribution of DL1d evaluated on the *good* jets (Fig. 3.44a) and on the *good* jets from the $t\bar{t}$ sample (Fig. 3.44b) collected in run 429606. Such run was affected by a misconfiguration in the setup of the LAr electromagnetic calorimeter¹⁶ which resulted in an abundance of light-like reconstructed jets. This is reflected in Fig. 3.44a, which shows a distribution leaning more towards low values, hence towards light-like jets, in the data with respect to the reference. This effect is absent in Fig. 3.44b.

Soft Muon Tagged jets Another sample enriched in b -jets is built collecting those with an associated soft muon. The D_b distribution of the DL1d tagger evaluated on these jets

¹⁵The ID B-layer is the second layer of the pixel detector starting from its centre.

¹⁶The issue was later corrected in a reprocessing of the run.

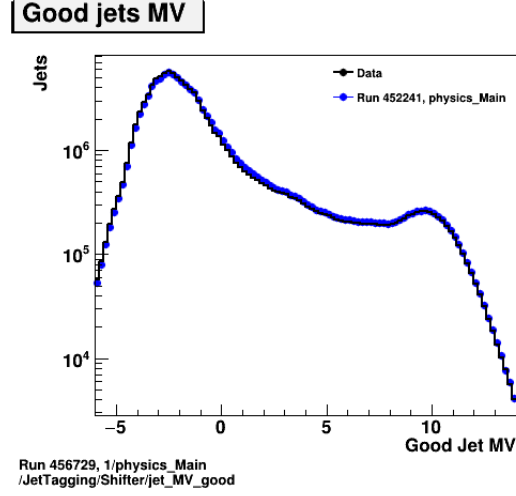


Figure 3.42: D_b distribution of the DL1d tagger evaluated on the *good* jets collected by the ATLAS detector during the Run-3 run number 456729. The data collected in run number 456729, in black, is compared to that collected in run number 452241, in blue.

is shown in Fig. 3.45. The relative amount of heavy-flavour jets in the collection is large enough to discern the two peaks of the c -jets and b -jets in the data.

Tracking and vertexing quality distributions Distributions monitoring the quality of tracks and vertices are included to help correlate issues in the detector with issues reflected on the higher-level jet distributions. Useful tracking distributions are showcased in Fig. 3.46. They display the distributions of the number of hits used for reconstructing the tracks. These variables are involved in the classification of jets as *good*, *suspect*, and *bad*. Discrepancies from the reference can help in understanding the reason for unusual relative populations of the three classes, as well as identifying issues in the ID subsystems.

Information about the number and quality of reconstructed vertices is monitored via the distributions of the amount of reconstructed primary vertices per event, normalized to the average number of interactions per bunch crossing $\langle\mu\rangle$; and by looking at the distributions of d_0 and z_0 (Fig. 3.47).

3.6.1 A harsh beginning for the Run-3 DQ: exemplifying how monitoring DQ can help recover problematic data

The first distributions monitored when Run-3 started in 2022 differed greatly from their Run-2 reference. The D_b distribution of DL1d did not display the two peaks of the b and light-like jets (Fig. 3.48a), the d_0 of tracks had two peaks right above and below zero (Fig. 3.48b), the z_0 distribution was wider than expected (Fig. 3.48c), and the number

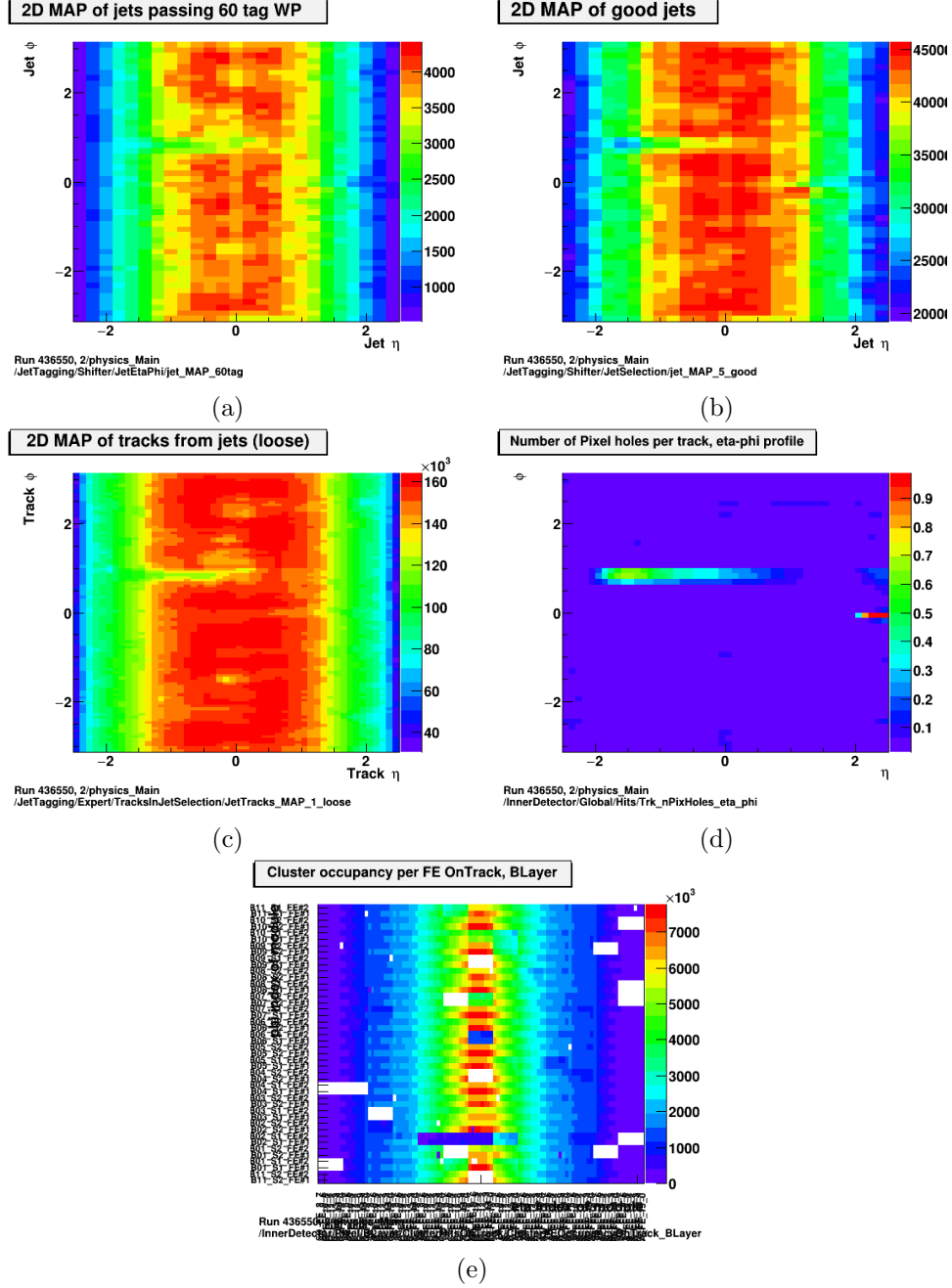


Figure 3.43: Effects on the jets tagged with the DL1d 60% efficiency WP (a), on the *good* jets (b), on the tracks associated to the *good* jets (c) and on the number of pixel holes in the reconstructed tracks (d) provoked in the $-2 < \eta < 0$ zone of the detector by the faulting of the L0.B2 module of the B-layer (e) in run 436550.

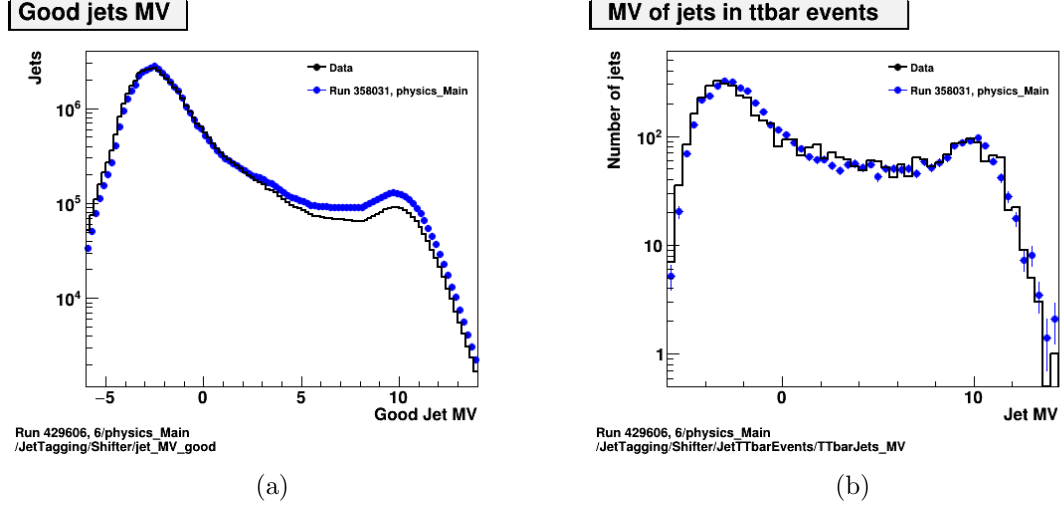


Figure 3.44: DL1d D_b distribution evaluated on *good* jets (a) and on jets belonging to $t\bar{t}$ events (b) collected in run 429606.

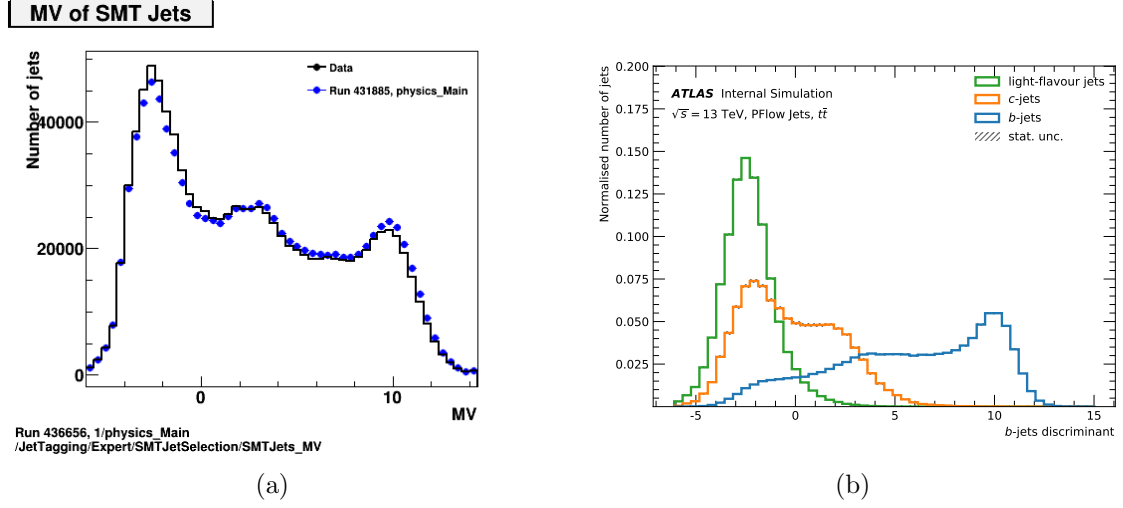


Figure 3.45: DL1d D_b distribution evaluated on *good* jets containing soft muons (a) compared to that evaluated on jets belonging to simulated $t\bar{t}$ events (b). The three peaks of the b , c and light-flavour jets are identifiable in the data.

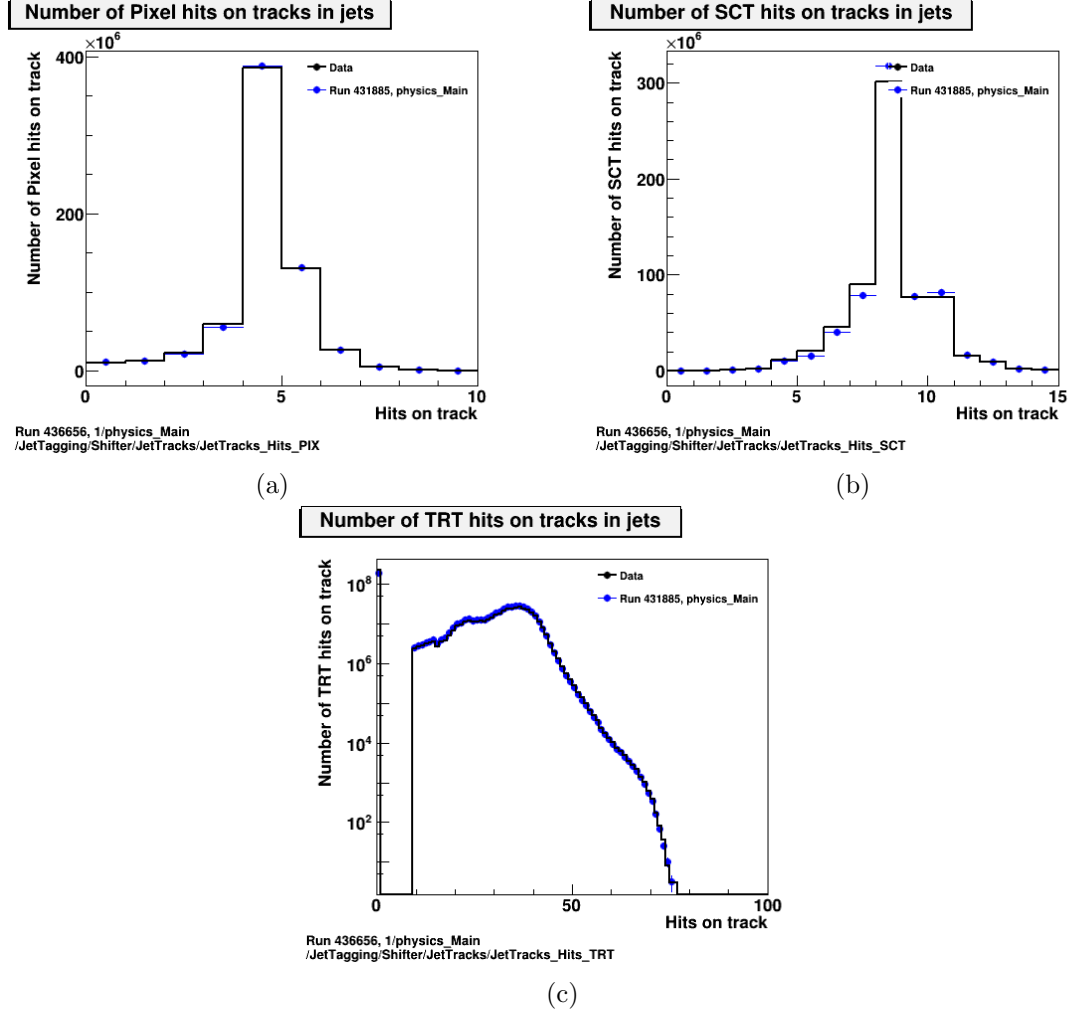


Figure 3.46: Number of pixels (a), SCT (b) and TRT (c) hits used to reconstruct tracks associated to jets selected by the offline jet flavour-tagging DQ cut flow.

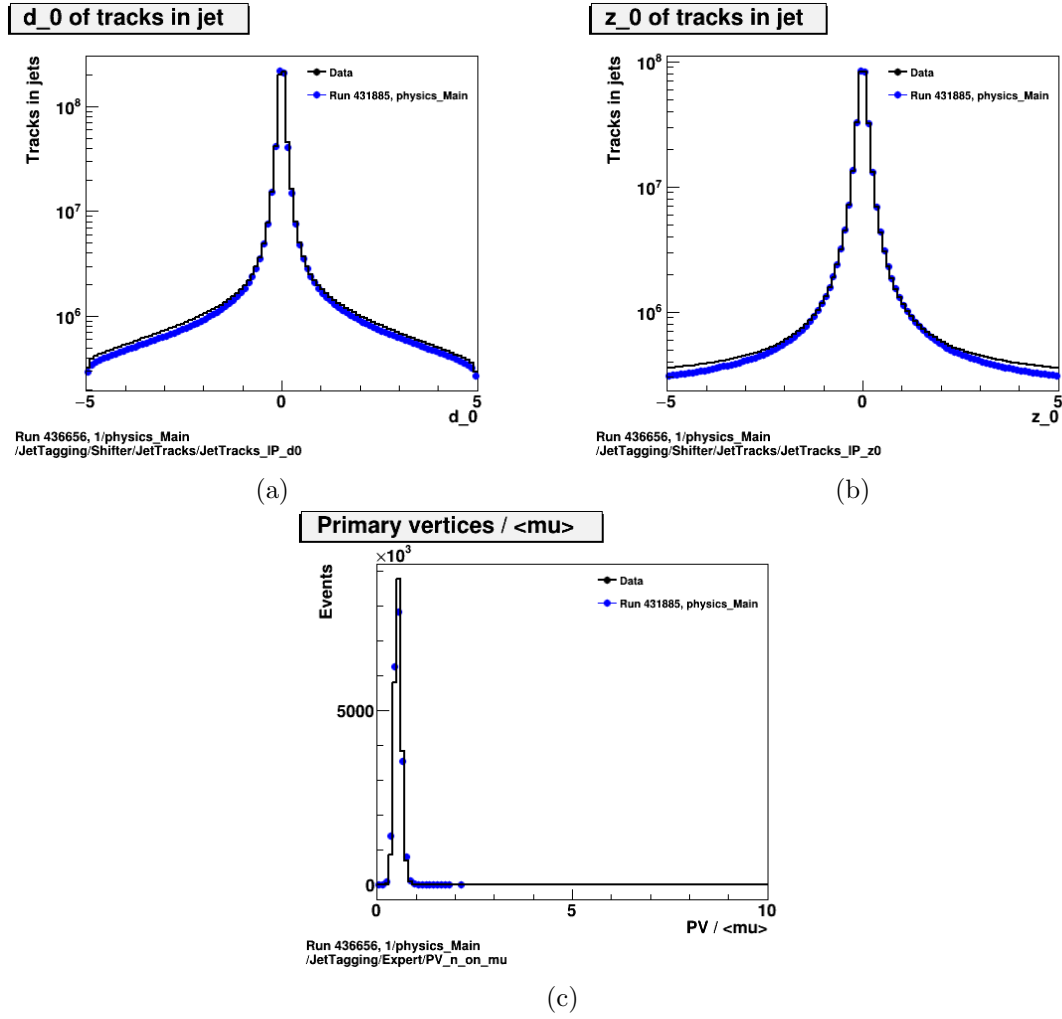


Figure 3.47: d_0 (a) and z_0 (b) of tracks associated to jets selected by the Offline jet flavour-tagging DQ cut flow in run 436656. Number of reconstructed PVs normalized to the average number of interactions per bunch crossing $\langle \mu \rangle$ in the same run (c).

of selected *good* jets had a sinusoidal dependence on ϕ (Fig. 3.48d and 3.48e), apparently violating the symmetry of the detector.

The scrutiny of these features helped identify the problem from the jet flavour-tagging perspective. A bug in the code responsible for fitting the position of the beam spot in the run was causing the failure of the fit. As a result of this failure, the beam spot was assigned the default location of the centre of the detector. The consequence of the misplacement is reflected in the plots. For instance, the sinusoidal shape in the *good* jets distributions is a geometrical effect of the misplacement. The effect is illustrated in Fig. 3.49. Jets produced in directions orthogonal to the line connecting their origin to the geometrical centre of the detector are more difficult to reconstruct if, during reconstruction, the wrong assumption is made of them originating from the geometrical centre of the detector. This is because their constituents are not well contained in cones originating from the centre (Fig. 3.49a). Such difficulty does not arise for jets produced in the same direction as the line connecting their origin and the detector's centre. (Fig. 3.49b).

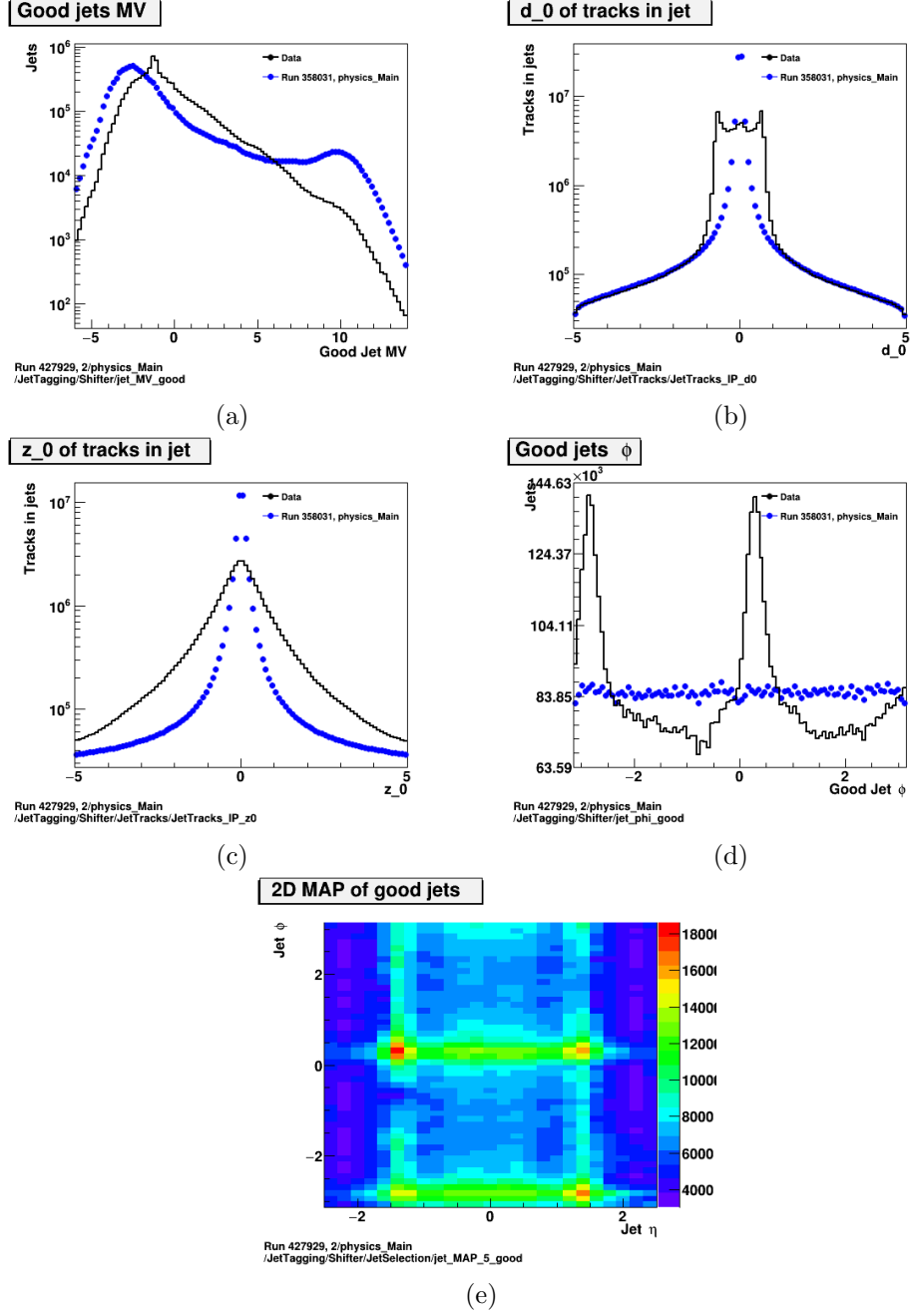
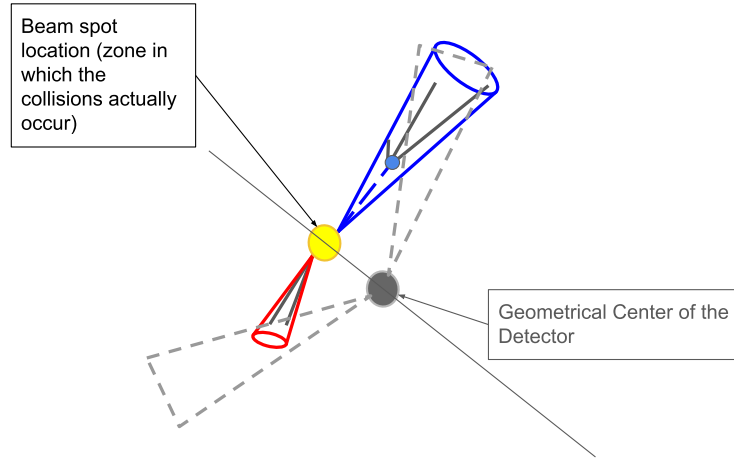
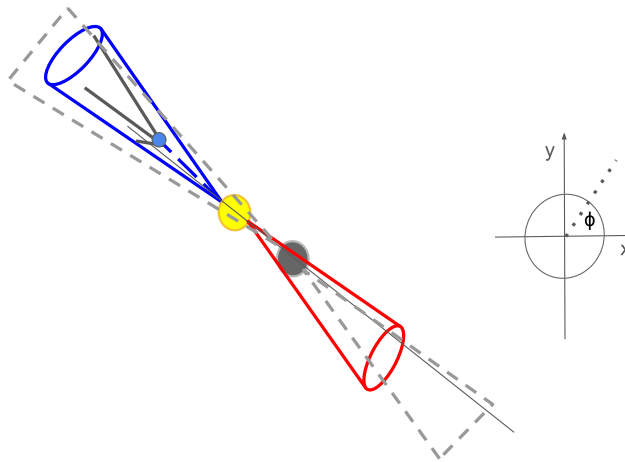


Figure 3.48: D_b distribution of *good* jets (a) d_0 (b) and z_0 (c) of tracks associated to jets, ϕ distribution of *good* jets (d) and (η, ϕ) 2D map of *good* jets selected by the offline jet flavour-tagging DQ cut flow in run 427929, at the beginning of Run-3.



(a)



(b)

Figure 3.49: The sinusoidal shape in the number of reconstructed jets (Fig. 3.48d and 3.48e) when the beam spot location erroneously defaults at the center of the detector is explained by these two situations. When jets are in the direction orthogonal to the line connecting their origin and the center of the detector (a), it is difficult to include their constituents in cones with vertices at the detector's center. If the direction of the jets is parallel to such line instead, their reconstruction is not affected.

Chapter 4

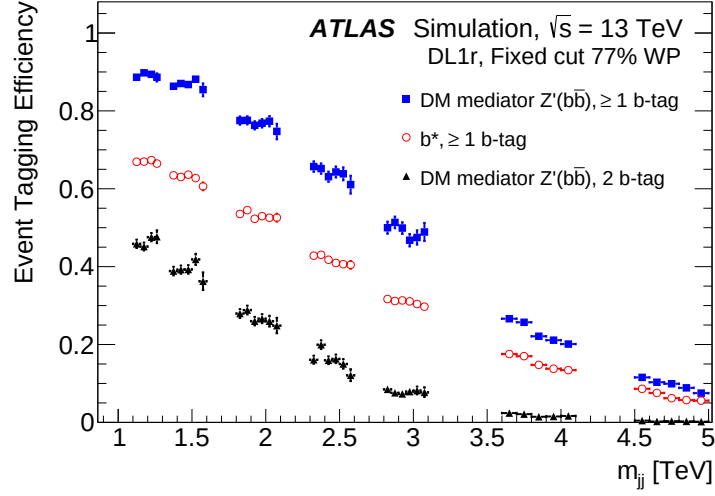
Jet flavour-tagging @ high p_T

Notwithstanding the significant improvements of the jet flavour-tagging algorithms achieved in recent years, some long-standing challenges persist, rooted in the physical processes underlying the formation of the jets and in the limitations of the detector. One of these challenges is identifying the flavour of jets produced with very large transverse momenta.

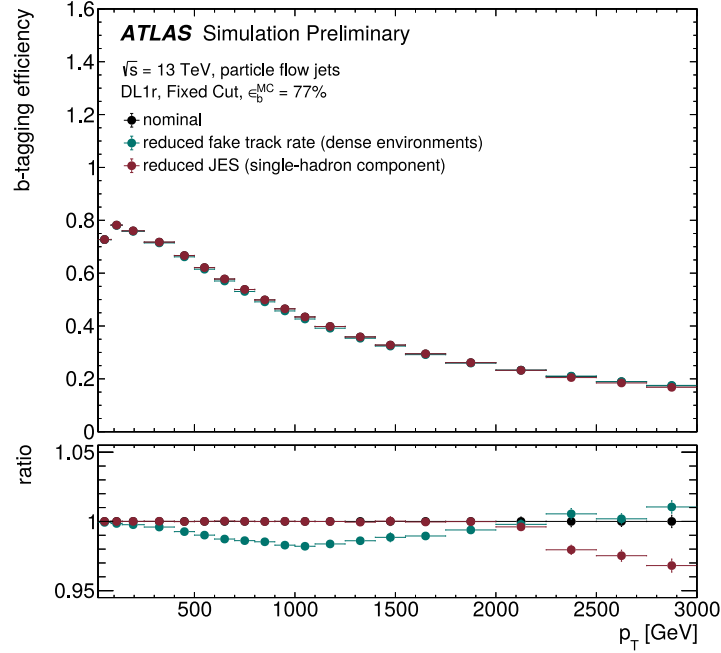
Fig. 4.1a illustrates the fraction of heavy-mass resonances decaying into b -jets correctly identified in the analysis described in [59], employing the 77% b -tagging WP of the DL1r algorithm. The resonance identification efficiency declines with increasing mass, as heavier resonances tend to produce jet pairs with high p_T , whose flavour is more difficult to determine. This difficulty is demonstrated in Fig. 4.1b, showing the DL1r b -tagging efficiency as a function of the jets' p_T when using the 77% WP¹. For high jet- p_T values, the b -tagging performance decreases significantly.

The decreasing jet flavour-tagging performance at high jet- p_T values, ubiquitous across all the jet flavour-tagging algorithms, is caused by the interplay of multiple concurring factors. The process that begins with the creation of a free quark or gluon during a collision, proceeds through its QCD radiation, and ends with the formation of a colourless jet due to the hadronization of the emitted colour charges is called *fragmentation*. The number of charged particles produced in a jet during fragmentation is proportional to the transverse momentum of the initial quark or gluon. This effect is evident in Fig. 4.2, which displays the average number of tracks reconstructed and associated to a heavy-flavour jet in the simulated sample. As the p_T of the b -jet (Fig. 4.2a and 4.2b) and of the c -jet (Fig. 4.2c and 4.2d) increases, the amount of tracks from fragmentation grows. Fig. 4.3 shows how the fraction of tracks produced in different processes evolve as a function of p_T . Focusing on the b -jets without loss of generality, Fig. 4.3a and 4.3b show that the fraction of tracks

¹As a reminder, the 77% WP is defined as a fixed threshold on the D_b discriminant variable (Eq. 3.3), corresponding to a 77% proportion of the total number of b -jets correctly identified. Hence, the proportion of b -jets identified with the 77% WP as a function of p_T deviates from 77% for bins containing jets with a D_b distribution that differs from the distribution computed on all the jets.



(a)



(b)

Figure 4.1: (a) In the search for new resonances decaying in jet pairs, described in [59], events originated by the decay of a hypothetical Z' boson from a leptophobic Z' dark-matter mediator model (blue and black dots) and of an excited b -quark from a compositeness model (red dots) are required to contain at least one (1 b-tag) or exactly two (2-btag) b -tagged jets to be selected. The fraction of those correctly identified by means of the b -tagging selection decreases as a function of the mass of the resonance m_{jj} . This happens because heavy-mass resonances decay in high p_T b -jets, more difficult to identify. (b) The b -tagging efficiency of the DL1r algorithm for a fixed threshold on the D_b variable, corresponding to a 77% proportion of the total number of b -jets correctly identified, as a function of the jet p_T [83]. The blue and red markers correspond to the efficiency in the presence of uncertainty components affecting the track reconstruction and the jet momentum calibration, respectively.

from fragmentation grows from 10% at 50 GeV to 30% at 250 GeV. In the same range, the number of tracks emerging from the decay of the b -hadron peaks at 30% around 250 GeV, and is as low as 15 % at 1 TeV. As detailed in Chapter 3, this second category of tracks holds most of the information about the flavour of the jet. Its relative reduction at high p_T blurs the difference between heavy-flavour and light-flavour jets.

Another category of tracks whose average number and relative share of the total grows significantly with the p_T of the jets is fake tracks. Fake tracks are defined by having less than 50% of the hits produced by a single simulated particle. The higher their fraction, the more diluted the overall discriminant information available to the jet flavour-tagging algorithms for jet flavour identification. Their increase, as well as the decrease in the number of reconstructed tracks emerging from the decay of the heavy-flavour hadron, can be explained in terms of the challenges raised by tracking in dense environments [84].

Dense environments refer to areas within the detector where numerous charged particles are separated by small angles. High- p_T hadronic jets, which consist of tightly collimated particles, are prime examples. The density of particles within jets is directly linked to their p_T for two main reasons. Firstly, as seen in the previous paragraphs, there is an increase in the number of particles resulting from fragmentation. Secondly, due to Lorentz boosts, the jets' cones shrink proportionally to their momentum. These combined effects lead to a reduction of the track reconstruction efficiency in high- p_T jets. This phenomenon is illustrated for particles generated before the IBL in Fig. 4.4a, showcasing the track reconstruction efficiency decreased of $\sim 5\%$ when measured in jets with $p_T \sim 1.5$ TeV compared to jets with $p_T \sim 200$ GeV.

A further complication arises when reconstructing tracks emerging from very displaced decays. If the heavy-flavour hadron's p_T is high enough, it may traverse one or more layers of the ID. This is illustrated in Fig. 4.5, which shows that the fractions of 1 TeV b -hadrons traversing the first and second layer of the ID before decaying are respectively $\sim 40\%$ and $\sim 60\%$. In such cases, the efficiency of track reconstruction is affected in two ways. First, the decay products produce fewer signals in the detector since they do not interact with the ID layers situated more internally than the decay point. Second, when the decay occurs at a point of the detector located close to the next active layer, the decay products reach it still closely packed. This can produce clusters shared by multiple particles if their separation is smaller than the resolution of the detector, and inefficiencies in associating clusters to track. These two factors concur to the drop in track reconstruction efficiency for particles with long production radius, *i.e.* produced distant from the beam line in the transverse plane, shown in Fig. 4.4b. The track reconstruction efficiencies for particles associated to jets with $|\eta| < 1.2$ decreases from $\sim 94\%$ at a 1mm production radius to $\sim 80\%$ at 100mm. A direct result of these complications is displayed in Fig. 4.6 [73]. It shows that the number of b -jets with a reconstructed secondary vertex decreases at high p_T , whereas it increases for c -jets and light-flavour jets. The latter are to be considered as fakes.

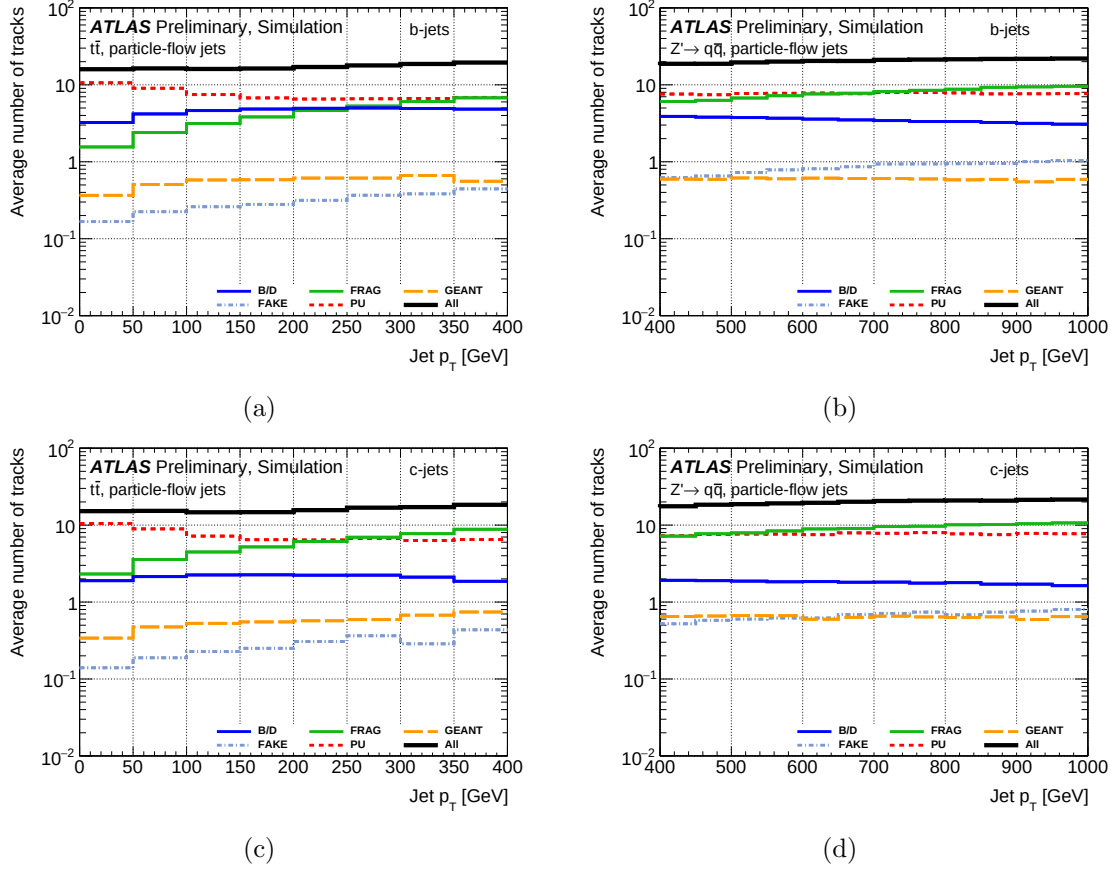
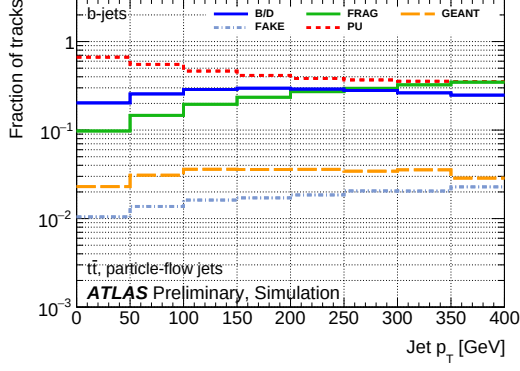
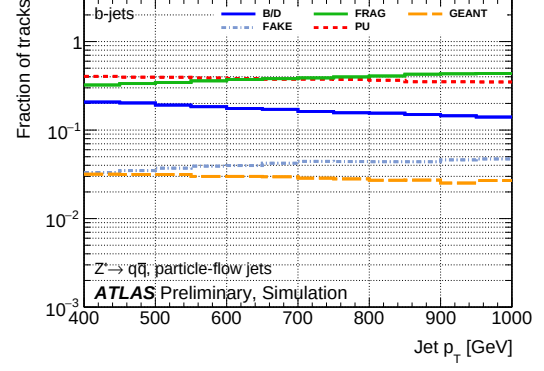


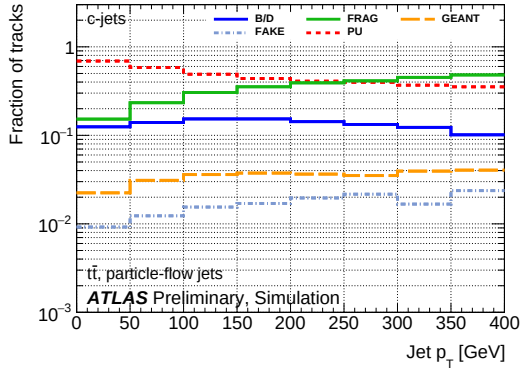
Figure 4.2: Average number of tracks reconstructed and associated to a b -jet (a and b) or to a c -jet (c and d) in the simulated $t\bar{t}$ and Z' samples as a function of the jet p_T . Tracks are classified according to their origin: tracks from the decay chains of heavy-flavoured hadrons (B/D), other particles produced in the jet development and fragmentation (FRAG), secondary charged particles from interactions with the detector materials or from decays of long-lived hadrons (GEANT), and tracks from pileup (PU). Tracks with less than 50% of the hits produced by a single simulated particle are classified as fake tracks (FAKE). From [85].



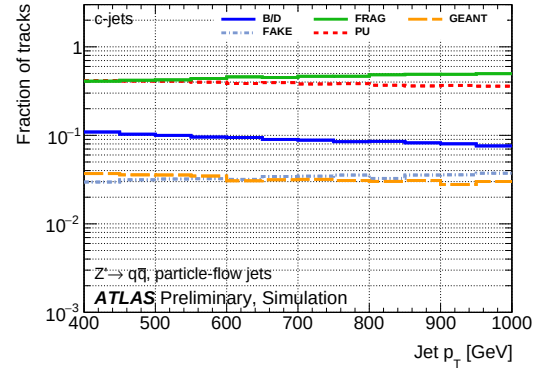
(a)



(b)



(c)



(d)

Figure 4.3: Fraction of tracks reconstructed and associated to a b -jet (a and b) or to a c -jet (c and d) in the simulated $t\bar{t}$ and Z' samples, as a function of the jet p_T . From [85].

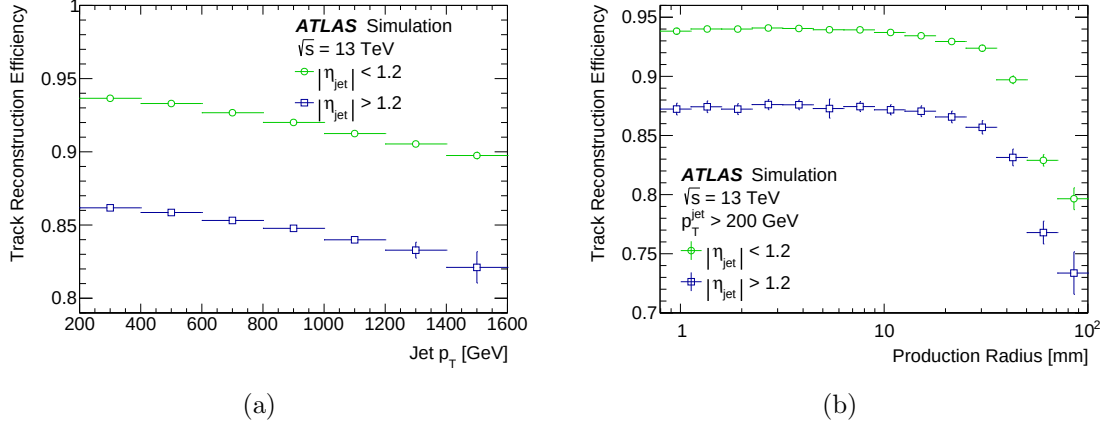


Figure 4.4: The track reconstruction efficiency is compared for charged particles produced before the IBL in jets with $|\eta| < 1.2$ ($|\eta| > 1.2$) as a function of the jet p_T (a) and of the production radius of the charged particles for simulated dijet MC events, where charged particles are not required to be created before the IBL (b) [84].

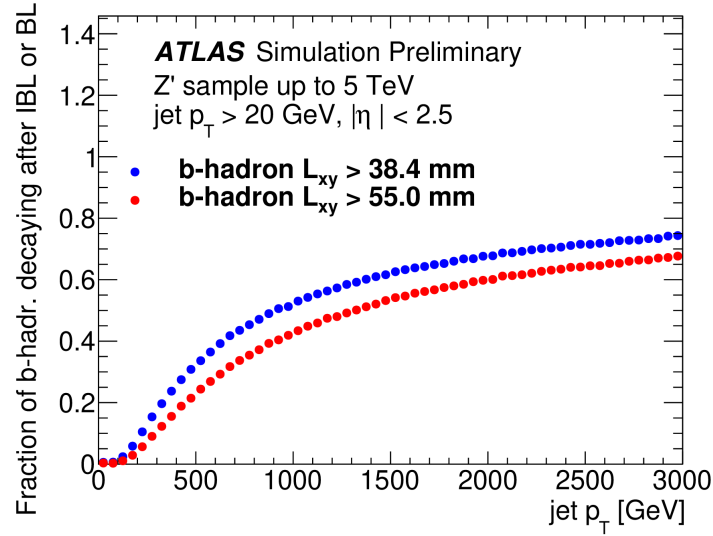


Figure 4.5: Fraction of b -hadrons with transverse decay length greater than the IBL and the subsequent B-Layer (BL) as a function of the jet p_T . The blue and red points correspond to the scenarios where b -hadrons decay respectively beyond the IBL and BL. From [86].

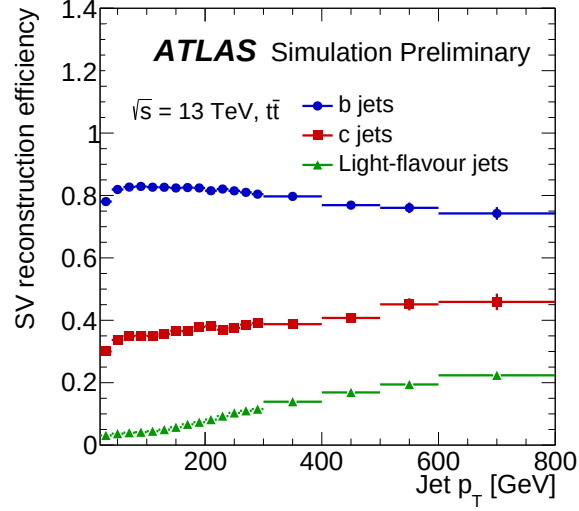


Figure 4.6: The secondary vertex reconstruction efficiencies for three jet flavours, as a function of the jet p_T . From [73].

4.1 Jet flavour information in the distribution of clusters in the ID

In [87], distinctive patterns have been identified in the distribution of clusters within b -jets containing b -hadrons that decay beyond the initial layers of the ID. In Fig. 4.7, a simplified representation illustrates these distributions for high and low p_T b -jets and for a light-flavour jet.

When a b -hadron decays between two pixel layers, the charged particles it produces do not interact with the layers preceding the decay point. Consequently, an increase in the number of clusters is observed in the first layer beyond the decay point. This increase is not evident in the simplified b -jet where the b -hadron decays before reaching the first pixel layer, nor in the idealized light-flavour jet.

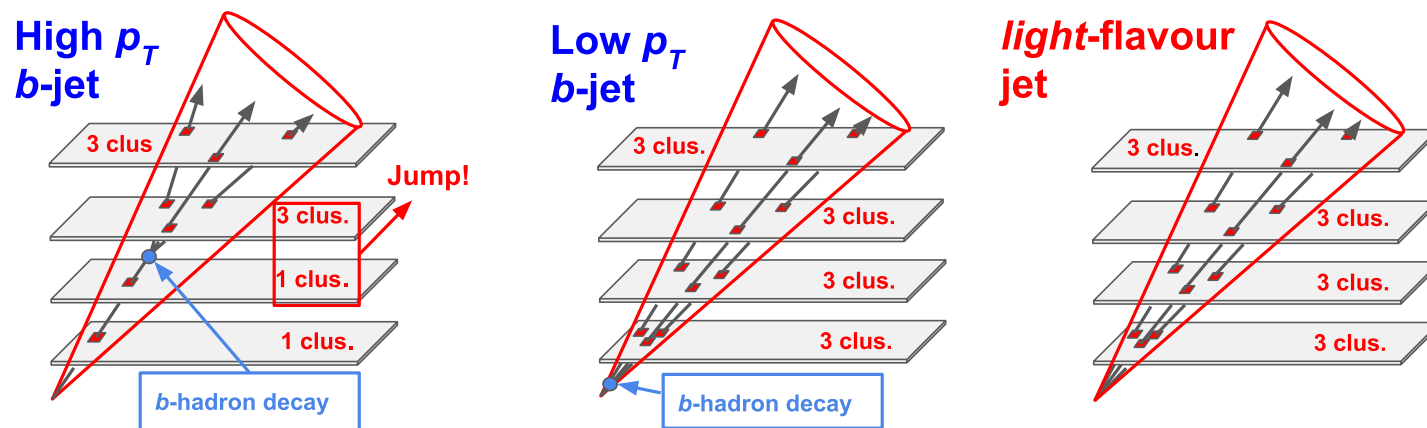


Figure 4.7: Distribution of clusters in idealized high and low p_T b -jets and in a light-flavour jet. The number of clusters increases in the first pixel layer beyond the displaced decay of the charged b -hadron in the high p_T scenario. This feature can be used to boost the performance of jet flavour-tagging algorithms in the p_T region in which it decreases the most.

Basic calculations can provide insights into the momentum range at which b -hadrons are expected to decay beyond the first layer of the ID. With b -hadron masses typically around 5 GeV, assuming an average decay length l of 33.25 mm (the radial distance from the first pixel layer to the beam-line) in the formula $l = \gamma\beta c\tau = p\tau/m_0$, and rearranging it to find the momentum p , we find that b -hadrons have momenta of about 370 GeV for an average $\langle c\tau \rangle \sim 450 \mu\text{m}$.

Taking into account that b -hadrons' momenta represent approximately 85% of the total momenta of the jets they are associated with [88], we can deduce that their average transverse decay length is equal to the IBL radius when they are associated to jets with $p_T \sim 400 \text{ GeV}^2$. This value is not far from where the performance of **DL1r** begins to decline (Fig. 4.1b). Therefore, leveraging the increase in the number of clusters per pixel layer as an additional feature for identifying b -jets seems particularly promising where track-based and secondary-vertexing algorithms are challenged.

The visual depictions of jets in Fig. 4.7 are simplified in the sense that, modulo the products of the displaced b -hadron decay in the high p_T scenario, they show a constant number of clusters per pixel layer. This deviates from what happens in real jets. In reality, multiple charged particles can be generated due to interactions with detector materials. Some of these particles might be absorbed by the pixel layers. Additionally, long-lived particles such as kaons or other mesons could be present in the jets and decay between two layers of the ID. Local variations in the number of particles produced in pile-up proton collisions, which are mistakenly associated with the jets, can impact the number of clusters recorded within them. Furthermore, particles may exit the reconstructed jet cone, causing the clusters they activated in the outer layers to be lost to the jet.

Notwithstanding these sources of background, [87] indicates that the patterns in the distributions of clusters inside jets can be used to identify those containing b -hadrons.

4.2 Preliminary studies on cluster-based jet flavour-tagging within the ATLAS Collaboration

I had a crucial role in performing and promoting studies to exploit the potential of the clusters of silicon hits for jet flavour-tagging in ATLAS. These studies have been performed in parallel with the evolution of the main track-based algorithms. They tested the impact of the idea introduced in [87] when incorporated in the architecture of the ever-improving cutting-edge algorithms, from BDTs to NNs to GNNs.

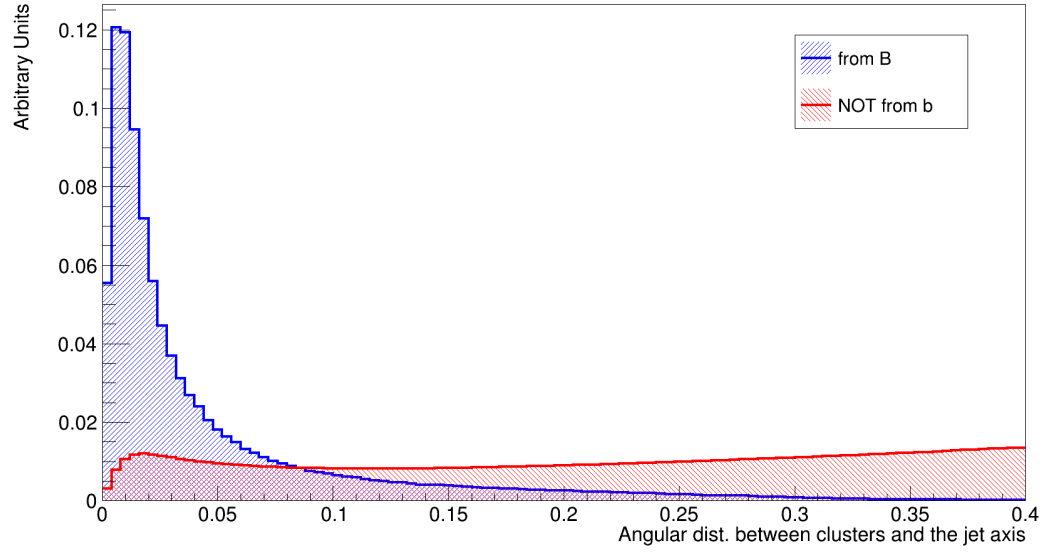
²To identify the jet p_T value above which b -hadrons *most likely* decay beyond the IBL, it would be necessary to compute that above which b -hadrons *median* transverse decay length is equal to the IBL radius. Such value is $\sim 900 \text{ GeV}$, deduced from Fig. 4.5.

4.2.1 BDT using IP3D and clusters

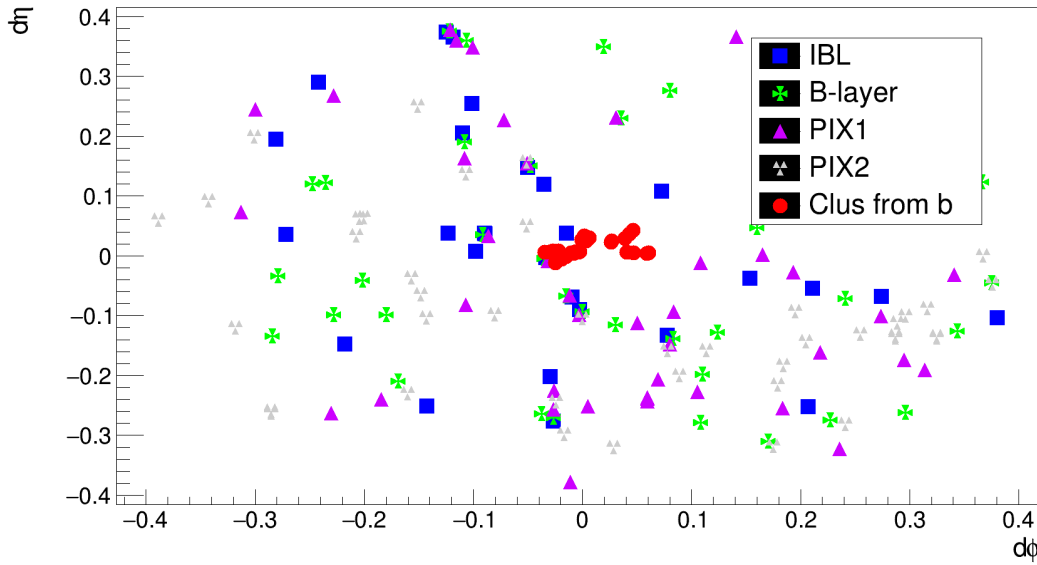
First, while DL1r was still being developed, I trained a BDT to identify the flavour of jets belonging to the Z' simulated sample [89, 90]. This study tested the impact of clusters on the performance of a jet flavour-tagging algorithm using for the first time a MC sample faithfully simulating the conditions realized within the ATLAS detector during pp collisions. The BDT was trained using as input variables the jet p_T , $|\eta|$, the IP3D p_b and p_l scores, the IP3D $_l$ log-likelihood ratio as defined in Sec.3.2.1, and the four numbers of clusters per pixel layer in a cone with an angular opening $\Delta R = \sqrt{d\eta^2 + d\phi^2} = 0.04$ around the jet axis.

The Z' sample was chosen as it was designed purposely for developing the jet flavour-tagging algorithms at high p_T . The value $\Delta R = 0.04$ was chosen based on Fig.4.8a, which shows the distribution of the angular separation between the clusters and the jet axis within the simulated jets. The cut at $\Delta R = 0.04$ ensures that approximately 90% of the clusters produced by charged particles belonging to the b -hadron decay chain are used in the training while discarding a significant amount of the other types of clusters. Fig. 4.8b shows the $d\eta$ and $d\phi$ values, with respect to the jet axis, of the clusters of pixel hits associated to a single b -jet. It showcases that the decay chain of the b -hadron develops at the core of the jet's cone.

The main result of the study is given in Fig. 4.9, which showcases the light-flavour jet rejection rates of the BDT trained with and without the number of clusters per pixel layer as a function of the jet p_T . Adding the information on the clusters improves the light-flavour jet rejection rates of up to 30% for jets with p_T of 500 GeV.



(a)



(b)

Figure 4.8: (a) Distribution of the angular distance to the jet axis ΔR of the clusters originated by charged particles belonging and not belonging to the decay chain of the b -hadron in simulated b -jets. (b) $d\eta$ and $d\phi$ of clusters belonging to a single b -jet in the various layers of the pixel detector. The clusters belonging to the decay chain of the b -hadron, displayed in red, accumulate at the core of the jet.

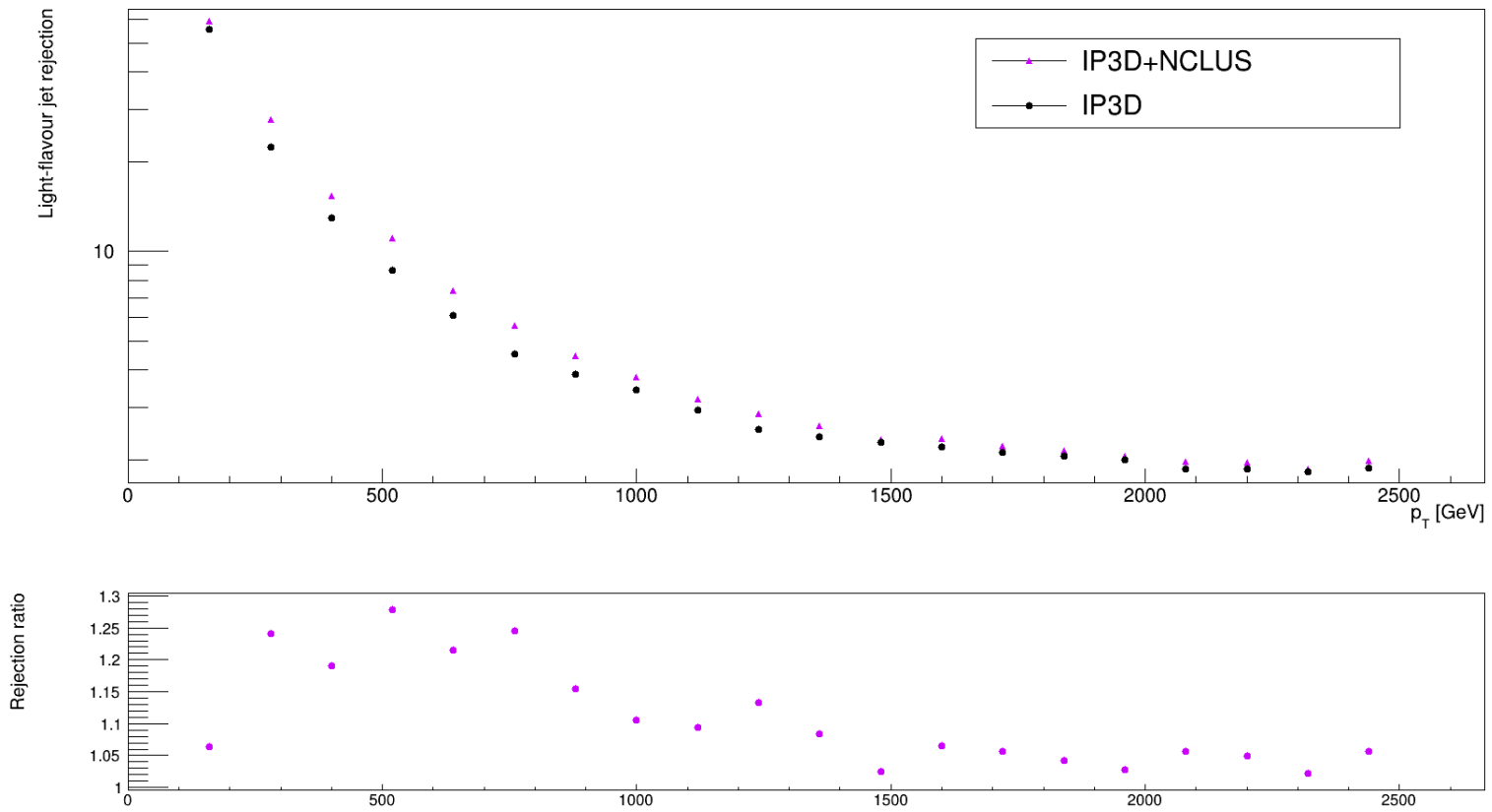


Figure 4.9: light-flavour jet rejection rates of the BDT trained with and without the number of clusters per pixel layer as a function of the jet p_T . Adding the information on the clusters improves the light-flavour jet rejection rates of up to 30% for jets with p_T of 500 GeV.

DL1d using clusters The impact of the information about the clusters of silicon hits was also studied when incorporated in the DL1d algorithm. The training was performed on the Z' sample, using, in addition to the track-based input variables described in Sec. 3.2.4, the number of clusters per pixel layer and the average angular separation of the clusters from the jet axis. Different openings of the cone around the jet axis containing the clusters were considered, and preliminary improvements of the order of 10% in the rejection of the background jets were observed. This study never went beyond the preliminary stages because of the promising performance of the new GN1 algorithm, which shifted the focus of the jet flavour-tagging group towards GNN-based algorithms.

4.3 Jet flavour-tagging with graphs and clusters

I trained the algorithms described in this section while the GN1 and GN2 algorithms were being developed. While extensively building upon the code written by other members of the jet flavour-tagging group, who created the core of the frameworks utilized to release the main taggers, I contributed to improving their flexibility as a byproduct of adapting them for my research. Incorporating the clusters of silicon hits in the GN2 architecture eventually improves the b -tagging rejection rates of significant factors over a wide set of WPs, peaking at 80% when focusing on jets with approximately $1 \text{ TeV} < p_T < 2 \text{ TeV}$ from the Z' sample. The improvement in c -tagging is less pronounced, of the order of 10%.

4.3.1 GNNs trained solely on clusters: the GN1ClusOnly and GN2ClusOnly algorithms

GN1ClusOnly and GN2ClusOnly are two GNN-based jet flavour-tagging algorithms that I developed to operate on representations of jets as fully connected graphs of clusters of pixel hits. They were trained on jets with $p_T > 300 \text{ GeV}$ from a simulated Z' sample, saving only clusters in a cone with an angular opening of $\Delta R = 0.04$ around the jet axis, for the reasons motivated in previous sections (Fig. 4.10). Albeit, from my research standpoint, these algorithms' scope was mainly to develop a platform upon which the subsequent algorithms utilizing both tracks and clusters would be built; their promising performance in identifying the flavours of jets based solely on clusters opens the door to additional applications. For instance, algorithms not relying on tracks could prove useful for fast trigger pre-selection on jets, considering that, in such a context, tracking algorithms drain most of the available computational resources.

Preprocessing the training sample The training sample was subject to a preprocessing procedure, consisting of an undersampling step and a scaling and shifting step. For undersampling, jets are distributed in bins according to their p_T and η . Jets of the two most represented flavours in each bin are randomly removed until the populations of the

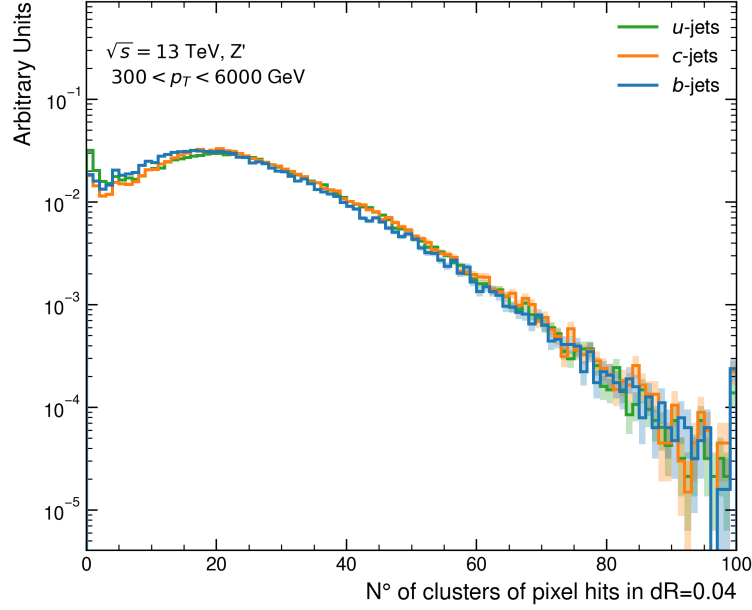


Figure 4.10: Normalized distributions of the number of clusters of pixel hits saved in a cone with angular opening $\Delta R = 0.04$ around the jet axis. The jets belong to a simulated Z' sample.

three flavour classes are equalized. Thus, two goals are achieved in one go: uniforming the kinematic distributions of the jets of the various flavours, and uniforming the amount of jets of each class shown to the network during training.

The first achievement ensures that the network is not biased towards learning the kinematic differences between the jet classes in the training sample, while still learning the correlations between the kinematic features of jets and the other training variables. Jet flavour-tagging algorithms must be optimized to identify flavour on a jet-by-jet basis, and the relative amount of jets with different flavours in the different p_T and η bins is process-dependent. If the network is exposed to the global kinematic features of a given training sample, it is biased when evaluating jets belonging to different samples, with different p_T and η distributions for the flavour classes. In other words, the network must be optimized to recognize intrinsic differences between individual jets of different flavours, rather than learning how the collections of jets of different flavours populate the phase space in the specific training sample. Still, the p_T and η of each individual jet are important information to provide to the network, because it can exploit their correlation with the other variables.

The second achievement, namely balancing the populations of the flavour classes in the training sample, ensures that the network is not biased towards favouring the majority class, the light-flavour jets, with respect to the others. With such bias, the model would be specialized in identifying jets of the majority background, disavouring the performance on

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Cluster Input	Description
x, y, z	Cartesian coordinates of the cluster with respect to the PV
$d\eta$	Pseudorapidity of the cluster, relative to the jet η
$d\phi$	Azimuthal angle of the cluster, relative to the jet ϕ
layer	Pixel layer measuring the cluster (0 is the IBL, 3 is the outermost pixel layer)
splitProbability1	Probability that the cluster is shared by two charges (Sec. 2.3.1)
splitProbability2	Probability that the cluster is shared by three charges

Table 4.1: Input features to train the `GN1ClusOnly` and `GN2ClusOnly` models, namely basic jet kinematics and information about the clusters of pixel hits.

the other classes. Fig. 4.11 shows the p_T and η distributions in the sample before and after the undersampling procedure.

After undersampling, input variables are shifted to be centred in 0 and scaled to have a standard deviation of 1. This step fastens the training procedure and helps avoid cases of *saturation* when randomly initializing the network weights, *i.e.*, cases in which the model is initialized with weights corresponding to a region of the loss landscape with vanishing gradients, where gradient descent is challenged.

Finally, the sample is split into training and validation samples. To summarize, the output of the preprocessing procedure is a training and a validation sample composed of 23M and 500k jets respectively, balanced in b -jets, light-flavour jets and c -jets, identically distributed in p_T and η , with input variables scaled and shifted to be centred in zero with unitary standard deviation.

Input Variables Clusters features are represented as vectors of variables, listed in Tab. 4.1. Their distributions in the training sample, after undersampling but before scaling and shifting, are shown in Fig. 4.12. The p_T and η of the jets are concatenated to each of the vectors representing the cluster features along with the other variables.

GN1ClusOnly: Architecture and Training `GN1ClusOnly` and `GN2ClusOnly` have similar architectures to those detailed in Sec. 3.4 for GN1 and GN2 respectively, but they do not implement the auxiliary tasks.

The hyperparameters shaping the architecture of `GN1ClusOnly` are listed in Tab. 4.2. The

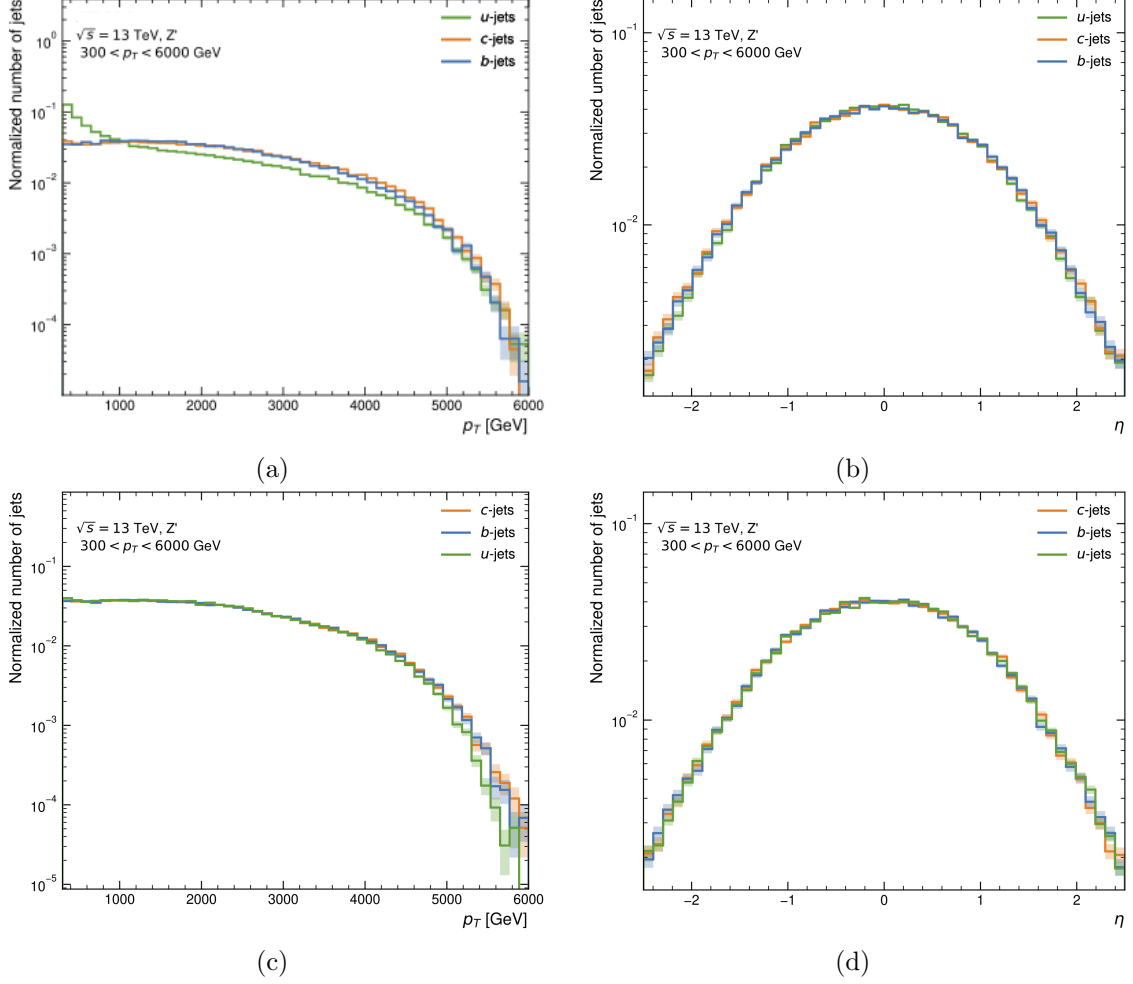


Figure 4.11: Normalized p_T and η distributions of the simulated jets belonging to the Z' sample used to train GN1ClusOnly and GN2ClusOnly, before (a and b) and after (c and d) the undersampling procedure. u -jets abbreviates light-flavour jets.

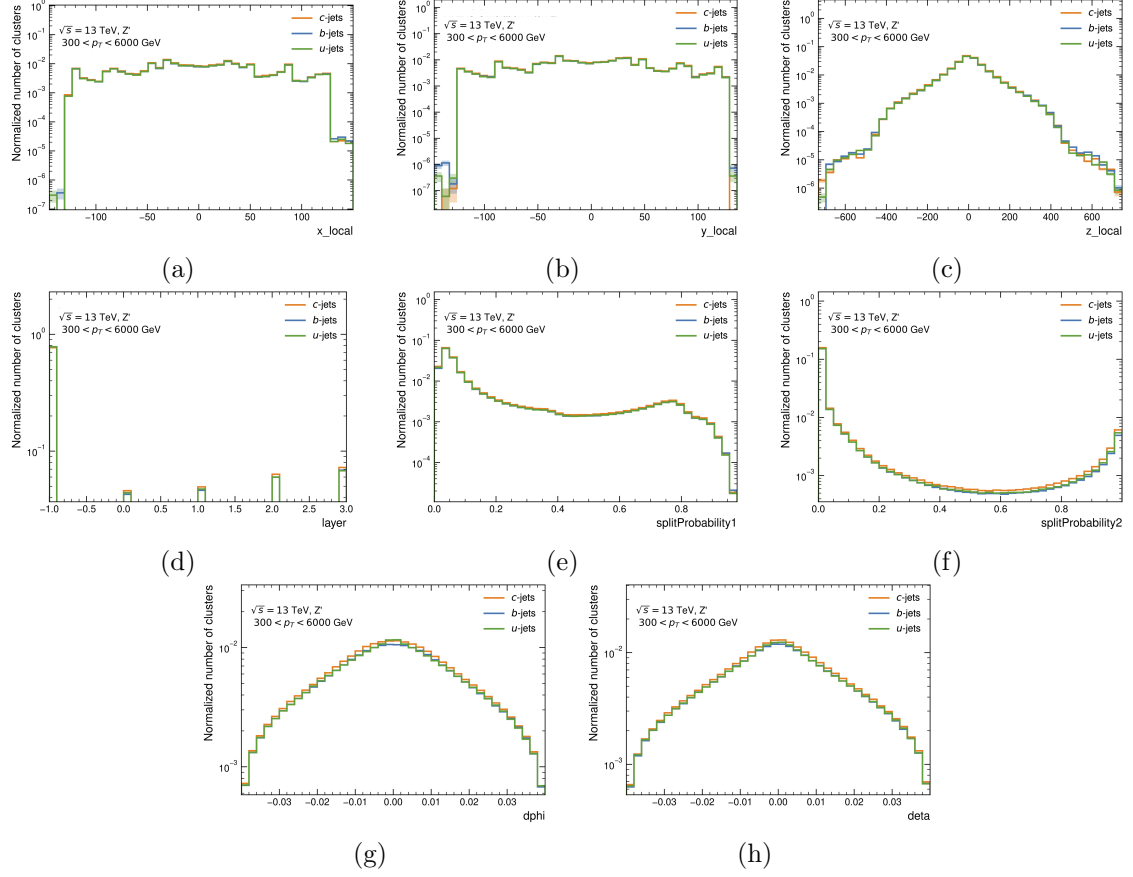


Figure 4.12: Input variables used by the GN1ClusOnly and GN2ClusOnly algorithms to describe the clusters' features.

Neural Networks	Input layer	Hidden layers	Output size
Cluster initializer network	10	64, 64, 64	64
Graph classification network	192 (= 128+64)	128, 64, 32, 16	3
1st (2nd and 3rd) GNN layer		Description	
$\mathbf{W}_s$ and $\mathbf{W}_r$	Linear layers, input size 64 (256), output size 256		
$\mathbf{a}^k$	Vectors of size 128		
Graph Pooling		Description	
$\mathbf{G}$	Linear layer input size 128, output size 1		

Table 4.2: Hyperparameters shaping the architecture of the **GN1ClusOnly** algorithm. Each parameter was defined when describing the architecture of the **GN1** algorithm in Sec. 3.4.1.

input vectors representing the clusters are processed through a NN that outputs a higher dimensional vector representation for each cluster (*initialization*). The resulting vectors serve as nodes to build a fully connected graph³, which is processed by three GNN layers. These update the representation of the nodes with the procedure detailed in Sec. 3.4.1 (*message-passing*). The resulting vectors are then concatenated to their representations before the message-passing step. Each node is then processed through a linear layer $\mathbf{G}$, which provides a scalar output. These outputs serve as attention weights for the nodes, which are combined via a weighted sum (*pooling*). The obtained vector, which is the global jet representation, is finally processed through a NN that outputs the p_b , p_c and p_u values assigned to the jet (*graph classification*).

GN1ClusOnly was trained for 75 epochs on 2 NVIDIA V100 GPUs. Each epoch took approximately one hour to complete over a training sample comprising 23M jets. This time is significantly longer than the approximate 30 minutes needed to train **GN1** with the same software on the same machines. Since up to 100 clusters in a cone with angular opening $\Delta R = 0.04$ around the jet axis are saved per jet, to be compared to the maximum of 40 tracks used when training **GN1**, graphs used by **GN1ClusOnly** tend to have more nodes, resulting in a more computationally expensive training. The learning rate was scheduled using an Adam optimiser [91], with an initial value of 1e-3, dividing the dataset into batches of 1000 jets per GPU. Fig. 4.13 shows the validation loss, defined as the L_{jet} term in Eq. 3.10 and computed at the end of each epoch over the 500k jets composing the validation sample. Epoch 72, the one minimizing this loss, was used to evaluate the algorithm’s performance.

GN2ClusOnly: Architecture and Training The parameters shaping the architecture of **GN2ClusOnly** are listed in Tab. 4.3. The differences distinguishing **GN2ClusOnly** from **GN1ClusOnly**, mainly the attention mechanism inspired by transformer encoders, mirror those between **GN1** and **GN2** (Sec. 3.4). Thus, they will not be described again here.

³Edges between the nodes are not assigned features.

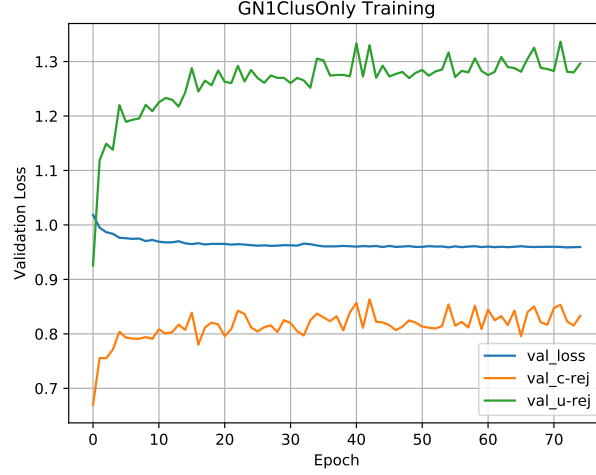


Figure 4.13: Validation loss of the **GN1ClusOnly** algorithm and logarithms of the c -jet and light-flavour jet rejection rates for a 70 % b -jet efficiency WP during its training. The epoch minimizing the loss is epoch 72.

Neural Networks	Input layer	Hidden layers	Output size
Cluster initializer network	10	256	192
Graph classification network	128	128, 64, 32	3
GNN layers		Description	
$\mathbf{W}_q$, $\mathbf{W}_k$ and $\mathbf{W}_v$	Linear layers, input size 192, output size 192		
$\mathcal{D}$	NN, input size 192, hidden layer of size 256, output size 192		
$\mathcal{L}$	Linear layer, input size 192, output size 128		
Graph Pooling		Description	
$\mathbf{G}$	Linear layer input size 128, output size 1		

Table 4.3: Hyperparameters shaping the architecture of the **GN1ClusOnly** algorithm. Each parameter was defined when describing the architecture of the **GN1** algorithm in Sec. 3.4.1.

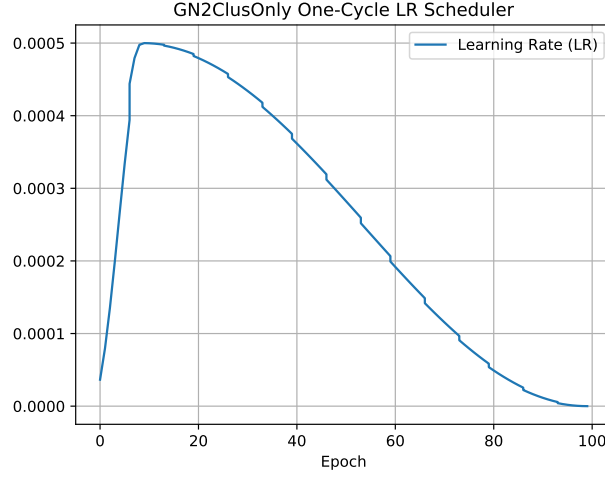


Figure 4.14: One-cycle learning rate scheduler used to train **GN2ClusOnly**. The scope of this scheduler is to begin the gradient descent procedure with short steps in the loss landscape, to identify the direction towards a minimum. Longer steps are then taken to approach the minimum. The pace is reduced towards the end of the training to avoid escaping the local minimum identified.

GN2ClusOnly was trained for 100 epochs on 2 NVIDIA V100 GPUs. The learning rate was scheduled using a one-cycle scheduler (Fig. 4.14) with initial and final values of $1e-7$, and max value of $5e-4$. Fig. 4.15 shows the validation loss, computed at the end of each epoch over the 500k jets composing the validation sample. Epoch 68, the one minimizing this loss, was used to evaluate the performance of the algorithm.

Performance The performance of the **GN1ClusOnly** and **GN2ClusOnly** algorithms are evaluated using the D_b and D_c discriminant variables computed as in Eq. 3.3 with the same f_c and f_b values adopted by the official **GN1** and **GN2** algorithms. These are $f_c=0.05$ and $f_b=0.2$ for **GN1**, and $f_c=0.1$ and $f_b=0.2$ for **GN2**. Fig. 4.16 shows how the mistagging efficiencies on background jets change when varying f_x ⁴ for a given WP.

Throughout this section, when the jet flavour-tagging WP needs to be fixed, it is set at 30% for b -tagging and at 20% for c -tagging. This choice is based on the computation of the standard WP thresholds using the $t\bar{t}$ sample. In the case of b -tagging, four standard WPs are optimized and calibrated by the jet flavour-tagging group, corresponding to 60%, 70%, 77% and 85% b -jet efficiencies when the thresholds are used to identify b -jets belonging to the $t\bar{t}$ sample. The Z' sample, however, comprises jets with a much harder p_T spectrum. Therefore, the flavour of these jets is more difficult to identify. Consequently, the signal

⁴ f_x indicates f_c when discussing b -tagging, and f_b when discussing c -tagging.

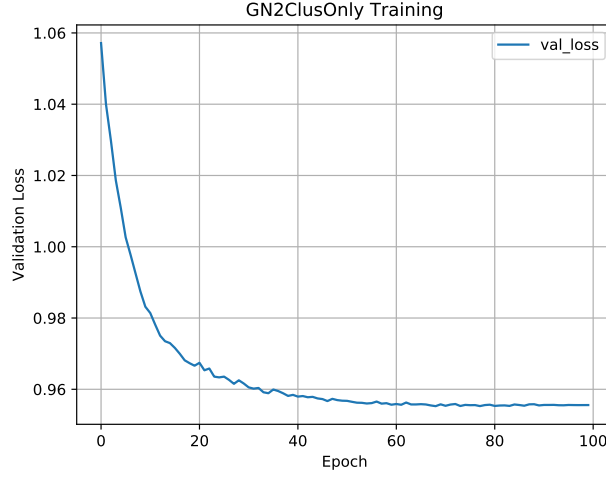


Figure 4.15: Validation loss of the `GN2ClusOnly` algorithm during its training. The epoch minimizing the loss is epoch 68.

efficiencies resulting from applying the thresholds that define the standard WPs are reduced when used on jets belonging to the Z' sample, reaching values as low as $\sim 30\%$. In Fig. 4.16, the `RNNIP` algorithm is also compared for scale, choosing values of the f_x parameter that strike a balance similar to that achieved by the `GN1ClusOnly` and `GN2ClusOnly` algorithms in the rejection rates of the two classes of background jets.

Fig. 4.17 shows the D_x distributions for the `GN1ClusOnly` and `GN2ClusOnly` algorithms computed on the jets belonging to the Z' sample. The distributions are almost identical for the two taggers. As expected, the effectiveness of D_b in separating b -jets from the other types of jets surpasses the separation achieved by D_c between c -jets and the other jet types. This follows from the fact that the c -jets identifying features are less pronounced than the b -jets', as explained in Sec. 3.1.

Fig. 4.18 shows the ROC curves of the `GN1ClusOnly`, `GN2ClusOnly` and `RNNIP` algorithms. Considering that `RNNIP` has access to track-based information, much more suitable for representing the physics processes unfolding inside the jets, the close performance achieved by the `GN1ClusOnly` and `GN2ClusOnly` algorithms, trained only on the clusters of pixel hits in the ID, expresses the potential of their GNN-based architecture.

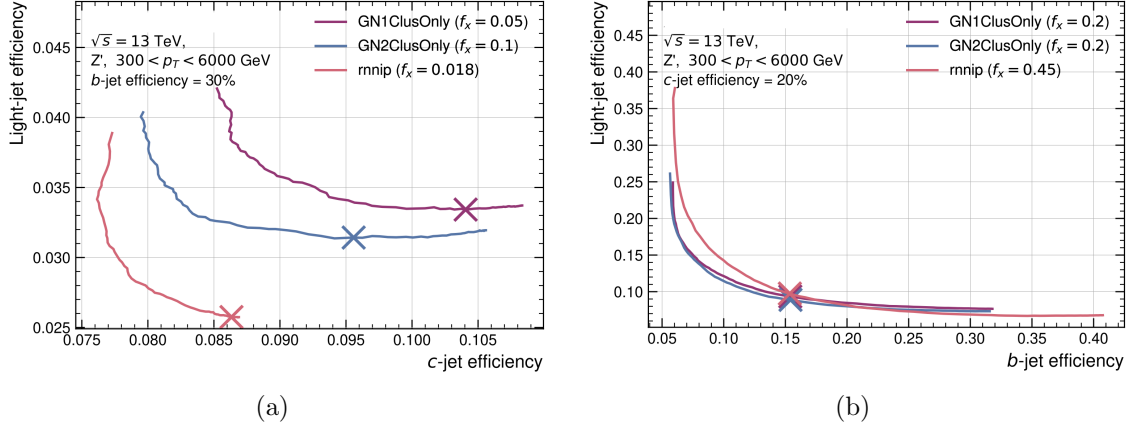


Figure 4.16: Background jet mistagging efficiencies of the GN1ClusOnly, GN2ClusOnly and RNNIP algorithms as a function of the f_c (a) and f_b (b) parameters. The cross markers indicate the values adopted when computing the D_b and D_c variables for the algorithms.

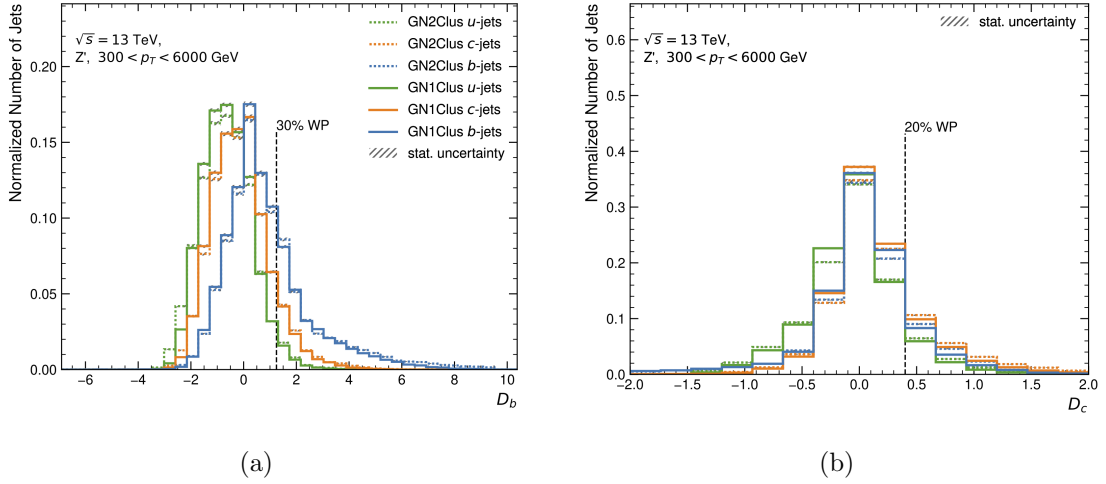


Figure 4.17: D_b (a) and D_c (b) distributions of the GN1ClusOnly and GN2ClusOnly algorithms when evaluated on a simulated Z' sample. As expected, D_b better separates b -jets from the other types of jets compared to the separation achieved by D_c between c -jets and the other jet types.

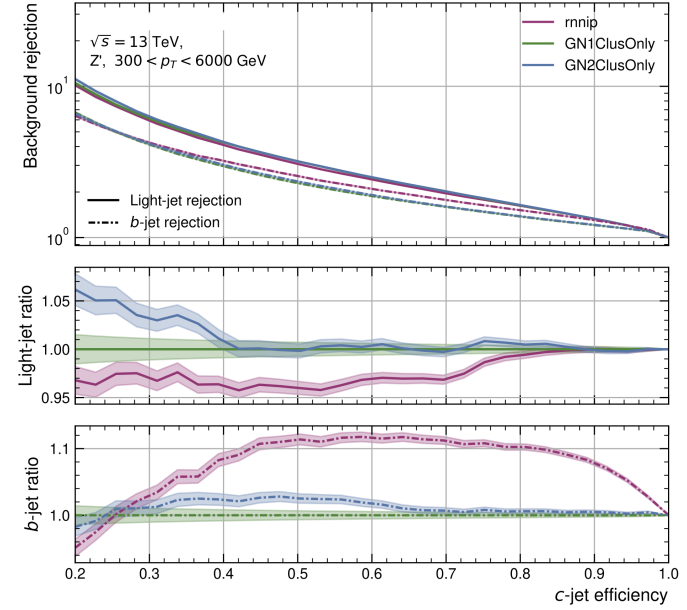
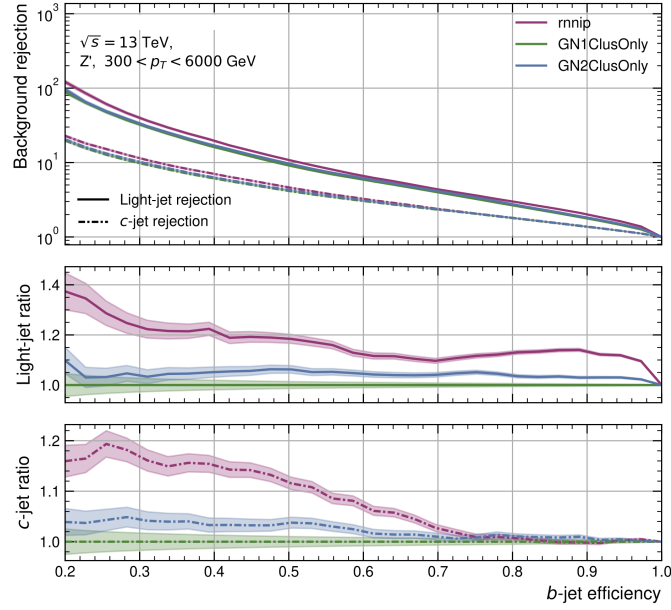


Figure 4.18: b -tagging (left) and c -tagging (right) ROC curves of the $GN1ClusOnly$, $GN2ClusOnly$ and $RNNIP$ algorithms when evaluated on the high- p_T jets of a simulated Z' sample. Considering that $RNNIP$ has access to track-based information, the close performance achieved by the $GN1ClusOnly$ and $GN2ClusOnly$ algorithms expresses the potential of their GNN-based architecture.

It is particularly interesting to study the performance of the `GN1ClusOnly` and `GN2ClusOnly` algorithms when evaluated as a function of the jets p_T . Fig. 4.19 displays the background rejection rates of `GN1ClusOnly`, `GN2ClusOnly`, and `GN2` as a function of the jets p_T for the 30% b -tagging and 20% c -tagging WPs. For this comparison, `GN2` was retrained on the same dataset used to train `GN1ClusOnly` and `GN2ClusOnly`. Albeit the performance of `GN2` largely surpasses that of the other two algorithms, the ratio among their performance approaches one for high values of the jets p_T . This corroborates the hypothesis that exploiting clusters could succeed in recovering part of the information about the flavour of high p_T jets that is lost to the algorithms relying solely on tracks.

4.3.2 `GN2+Clus_No_Aux`: improving `GN2` high p_T performance with clusters

To further test the hypothesis that concluded the previous section, I developed an evolution of the `GN2` algorithm, named `GN2+Clus_No_Aux`, capable of processing together tracks and clusters. `GN2+Clus_No_Aux` was trained over the same training sample and applying the same preprocessing procedures used for the `GN1ClusOnly` and `GN2ClusOnly` training campaigns. Also the architecture of the algorithm is unchanged with respect to `GN2ClusOnly`, with the addition of a NN for the initialization of the vectors representing the features of each track. Tracks and clusters are represented by the same input variables used by `GN2` and `GN2ClusOnly` algorithms respectively (Tabs. 3.4 and 4.1), and treated as equal types of nodes in the GNN layers. This means that, when building the graph, each cluster and track node is connected to all the other nodes in the graph. The input variables representing the clusters are shown in Fig. 4.12. Fig. 4.20 displays the variables representing the tracks.

I retrained the `GN2` and `GN2_No_Aux` algorithms on the same sample used to train `GN2+Clus_No_Aux`, to allow for a fair comparison of the performance of the models by factoring out effects unrelated to the introduction of the clusters. As opposed to `GN2+Clus_No_Aux` and `GN2_No_Aux`, which were trained for 100 epochs, `GN2` was trained for 200 epochs. This, to ensure a full convergence of the model, since the loss at epoch 100 was still slowly decreasing. The evolution of the losses of the algorithms during their training is displayed in Fig. 4.21. For `GN2`, the loss is computed as in Eq. 3.10. Since the architectures of `GN2+Clus_No_Aux` and `GN2_No_Aux` do not include the auxiliary tasks, the two algorithms are trained to minimize only the categorical cross-entropy term L_{jet} . Hence, while the lower validation loss of `GN2+Clus_No_Aux` anticipates the improved performance of the algorithm over `GN2_No_Aux`, comparing the magnitude of the loss of `GN2` with the others is misleading since the additional terms of the loss formulated in Eq. 3.10 are positive definite.

Performance The performance of the `GN2+Clus_No_Aux`, `GN2_No_Aux` and `GN2` algorithms are evaluated using the D_b and D_c discriminant variables computed as in Eq. 3.3 with the same f_c and f_b values adopted by the official `GN2` algorithm, namely $f_c=0.1$ and $f_b=0.2$.

Fig. 4.22 shows how the mistagging efficiencies on background jets change when varying

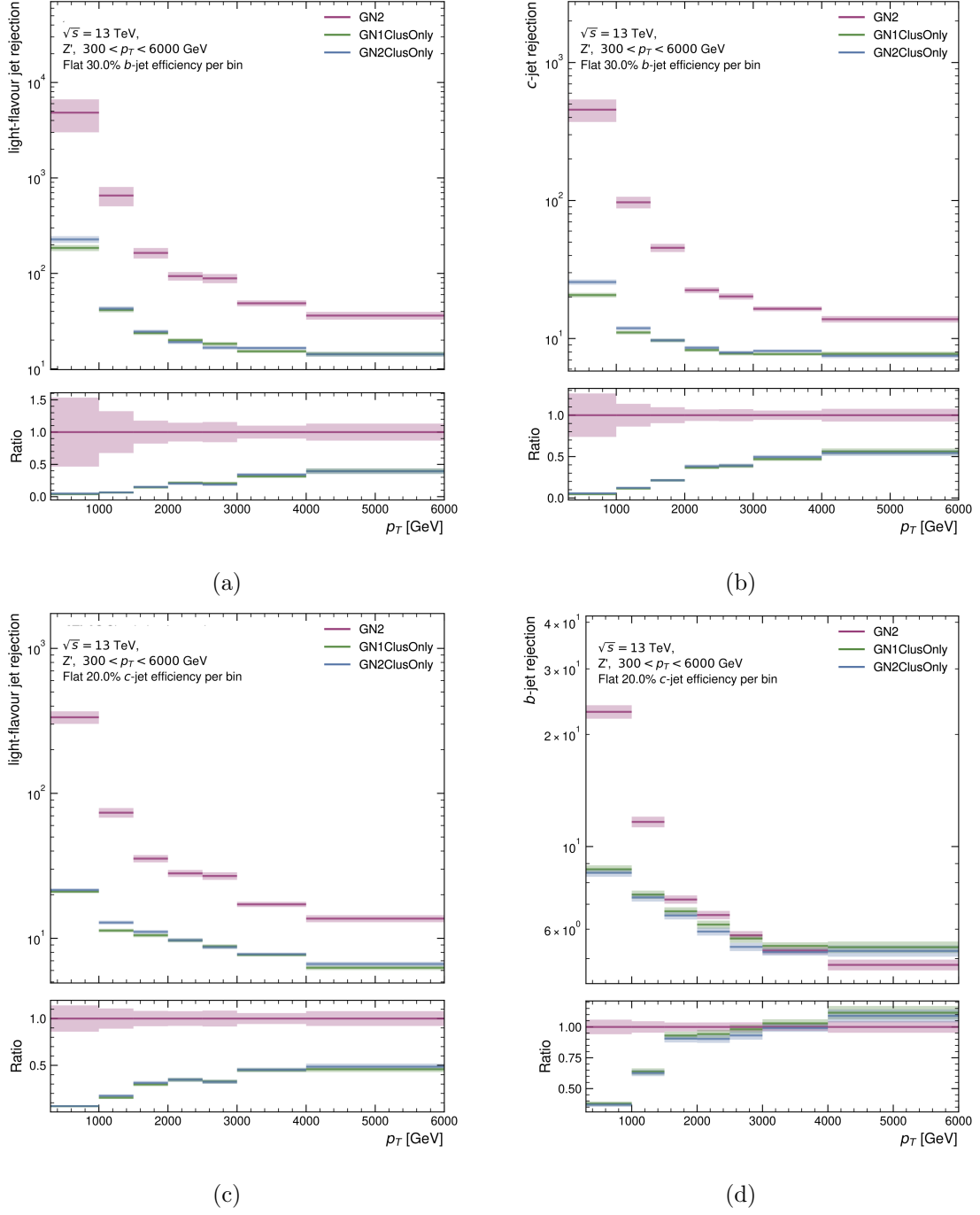
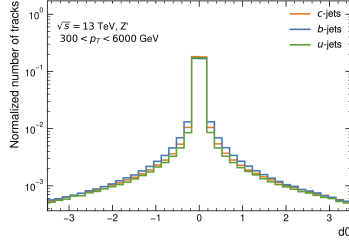
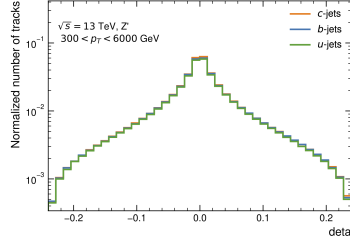


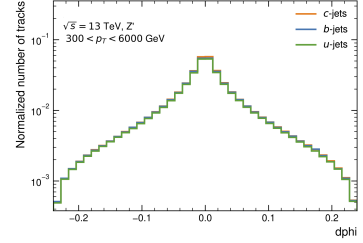
Figure 4.19: b -tagging (a and b) and c -tagging (c and d) background rejection rates of the GN1ClusOnly, GN2ClusOnly and GN2 algorithm as a function of the jets p_T . The WP cut values are recomputed in every p_T bin to ensure a flat signal efficiency. The performance of GN2 largely surpasses that of the other two algorithms, nevertheless, their background rejection rates decrease less steeply for high values of the jets p_T .



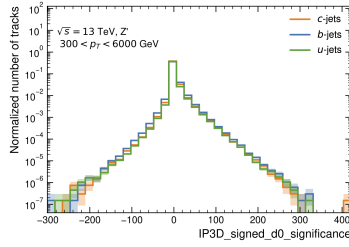
(a)



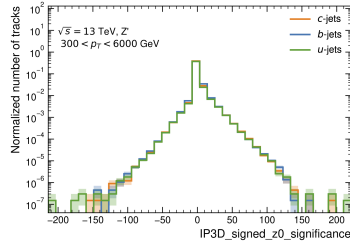
(b)



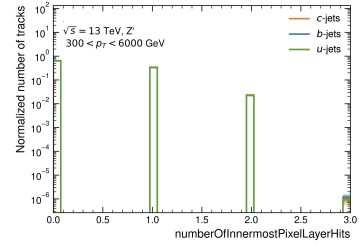
(c)



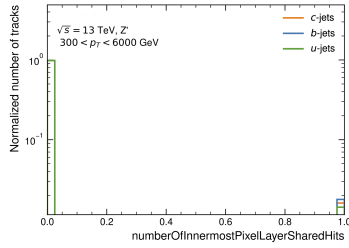
(d)



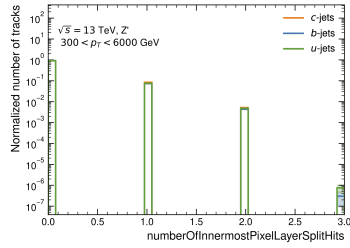
(e)



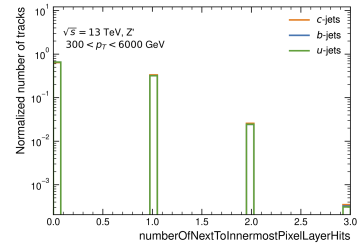
(f)



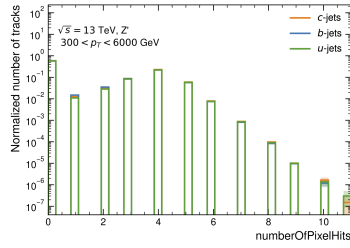
(g)



(h)



(i)



(j)

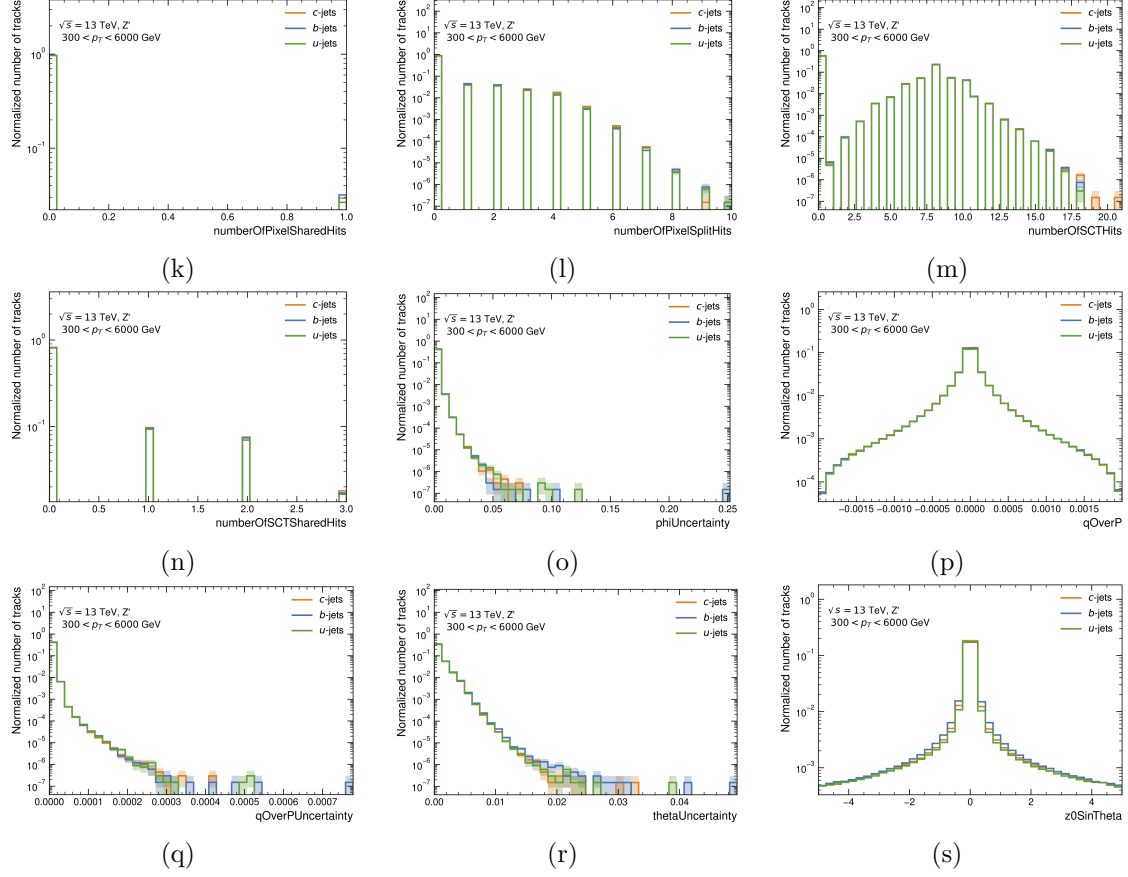


Figure 4.20: Input variables used by the GN2+C1us_No_Aux algorithm to describe the tracks' features.

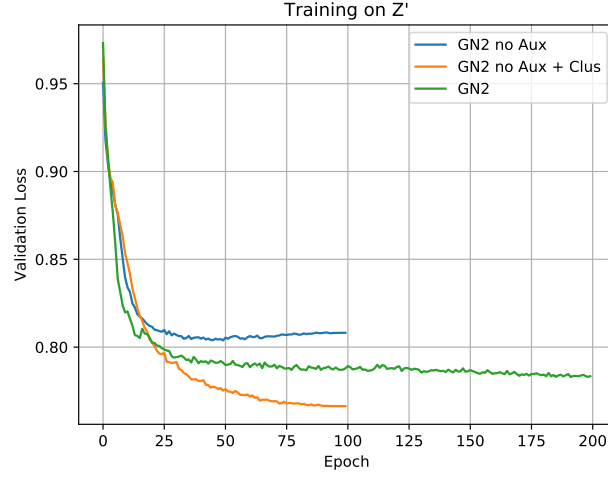


Figure 4.21: Validation losses of the GN2_No_Aux, GN2+Clus_No_Aux and GN2 algorithms.

f_x using the 30% b -tagging WP and for the 20% c -tagging WP, and Fig. 4.23 compares the ROC curves of the algorithms. These comparisons reveal the significant improvements brought about by introducing the clusters. When it comes to b -tagging (Fig. 4.22a and 4.23 left), GN2+Clus_No_Aux not only outperforms GN2_No_Aux almost twofold in terms of background rejection rates: it is also significantly better than GN2, most noticeably in rejecting light-flavour jets. These observations suggest that, at least at very high p_T , the distinctive information contained within the clusters of silicon hits near the core of the jets holds greater significance than the information obtained by influencing the model through auxiliary tasks, which are rooted in the physics of heavy-flavour hadron decays. These observations can be understood in terms of the decreased capability of the tracks to capture the physics processes unfolding in jets with high p_T , for the reasons detailed in Sec 4. The improvement is slightly less pronounced when it comes to c -tagging. In this case, GN2+Clus_No_Aux and GN2 have similar performance. In any case, they both outperform significantly GN2_No_Aux, of factors of up to $\sim 80\%$.

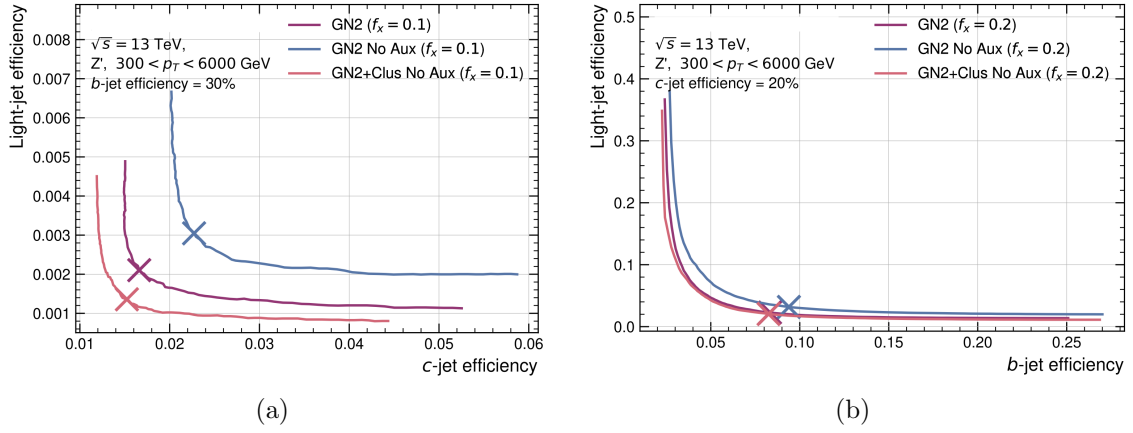


Figure 4.22: Background jets mistagging efficiencies as a function of the f_c (a) and f_b (b) parameters. The cross markers indicate the values that optimize the performance of the GN2 algorithm on a $t\bar{t}$ sample. They were selected for evaluating the performance of GN2+Clus.No.Aux, GN2.No.Aux and of the GN2 algorithm retrained on the simulated Z' sample.

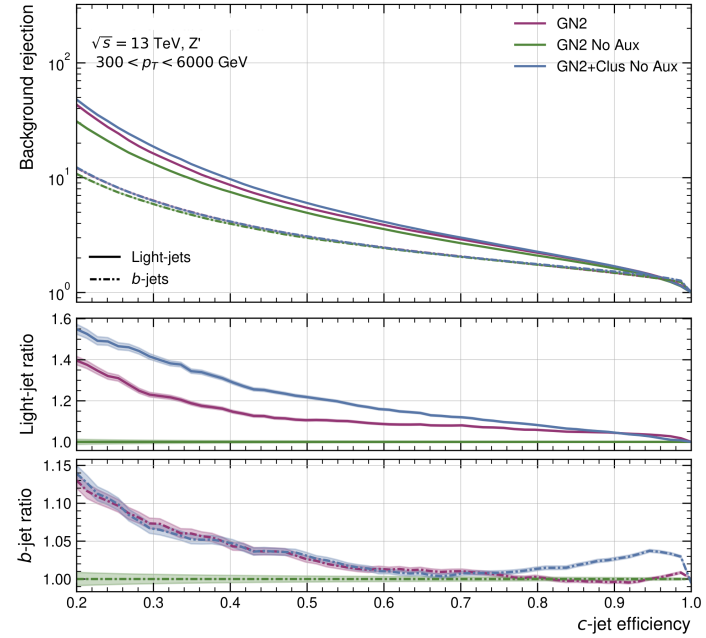
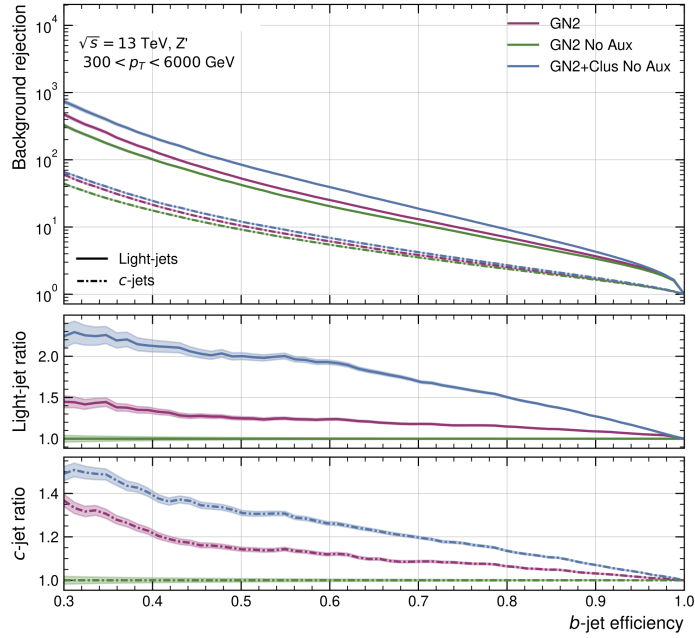


Figure 4.23: b -tagging (a) and c -tagging (b) ROC curves of the GN2, GN2_No_Aux and GN2+Clus_No_Aux algorithms when evaluated on the high p_T jets of a simulated Z' sample.

Fig. 4.24 displays the background rejections of the algorithms as a function of the jets' p_T . While the improvements of GN2 over GN2_No_Aux are mostly located in the first p_T bins, the ratio of the rejection rates of GN2+Clus_No_Aux over GN2_No_Aux is more stable in the higher p_T bins, testifying the p_T reach of the additional rejection power resulting from the introduction of the clusters in the algorithm.

Fig. 4.25 displays the jet flavour-tagging efficiencies of GN2+Clus_No_Aux, GN2_No_Aux and GN2 as a function of the transverse decay length L_{xy} of the heavy-flavour hadrons contained in the heavy-flavour jets. As opposed to the plots in Fig. 4.24, which were produced by recalculating the cuts on the D_x distributions bin by bin to maintain a consistent signal efficiency in every bin; in this case, a single threshold is used for all jets, corresponding to a given signal efficiency inclusive of the set of all jets. Hence, the signal efficiency in any specific L_{xy} bin can vary. When it is above the inclusive value, it indicates that the jets falling in the bin are more signal-like than the average. The b -tagging b -jet efficiency curve shown in Fig. 4.25a is of particular interest if analyzed from the perspective of the hypothesis put forward in [87] and detailed in Sec. 4.1. According to this hypothesis, the significance of the information introduced by the clusters lies in their increase across two consecutive layers when a heavy hadron decays between them. The plot shows that, while the b -jet efficiencies of the GN2 and GN2_No_Aux algorithms couple in the same way with the b -hadrons' L_{xy} , GN2+Clus_No_Aux favours b -jets where the decay happens between the pixel layers of the ID, right where the clusters of pixel hits are provided to the network.

4.3.3 GN2+Clus: outperforming the official GN2 with clusters

Since clusters are not preserved by default in the final data format employed across ATLAS for both analyses and jet flavour-tagging studies, a special sample had to be prepared *ad-hoc* to train the models described in the previous sections of this chapter. Information about the clusters was initially introduced to improve the performance of the jet flavour-tagging algorithms at high p_T . Thus, the first dataset containing the clusters was chosen to be composed of Z' events. Compared to the statistics used to train the official GN2 model, described in Sec. 3.4.2 and referred to as GN2v00 henceforth, which involved 213 million jets from a $t\bar{t}$ and Z' hybrid sample, the 23 million jets in the new sample were a relatively small number. The promising performance of the GN2+Clus_No_Aux algorithm prompted the group to produce an additional, relatively small $t\bar{t}$ sample to be used in concert with the Z' sample to train a GN2+Clus model on a hybrid sample. In the meantime, the framework was developed to enable the use of auxiliary tasks with track nodes in models that process multiple types of inputs. These two improvements enable an almost fair comparison of the performance of GN2+Clus and GN2v00. Additionally, GN2 was retrained on the same sample used to train GN2+Clus, to compare the two models when trained on the same amount of jets as well.

To build the hybrid training sample, jets with $20 < p_T < 300$ GeV are picked from the $t\bar{t}$ events, and jets with $300 < p_T < 6000$ GeV are picked from the Z' events. Clusters are

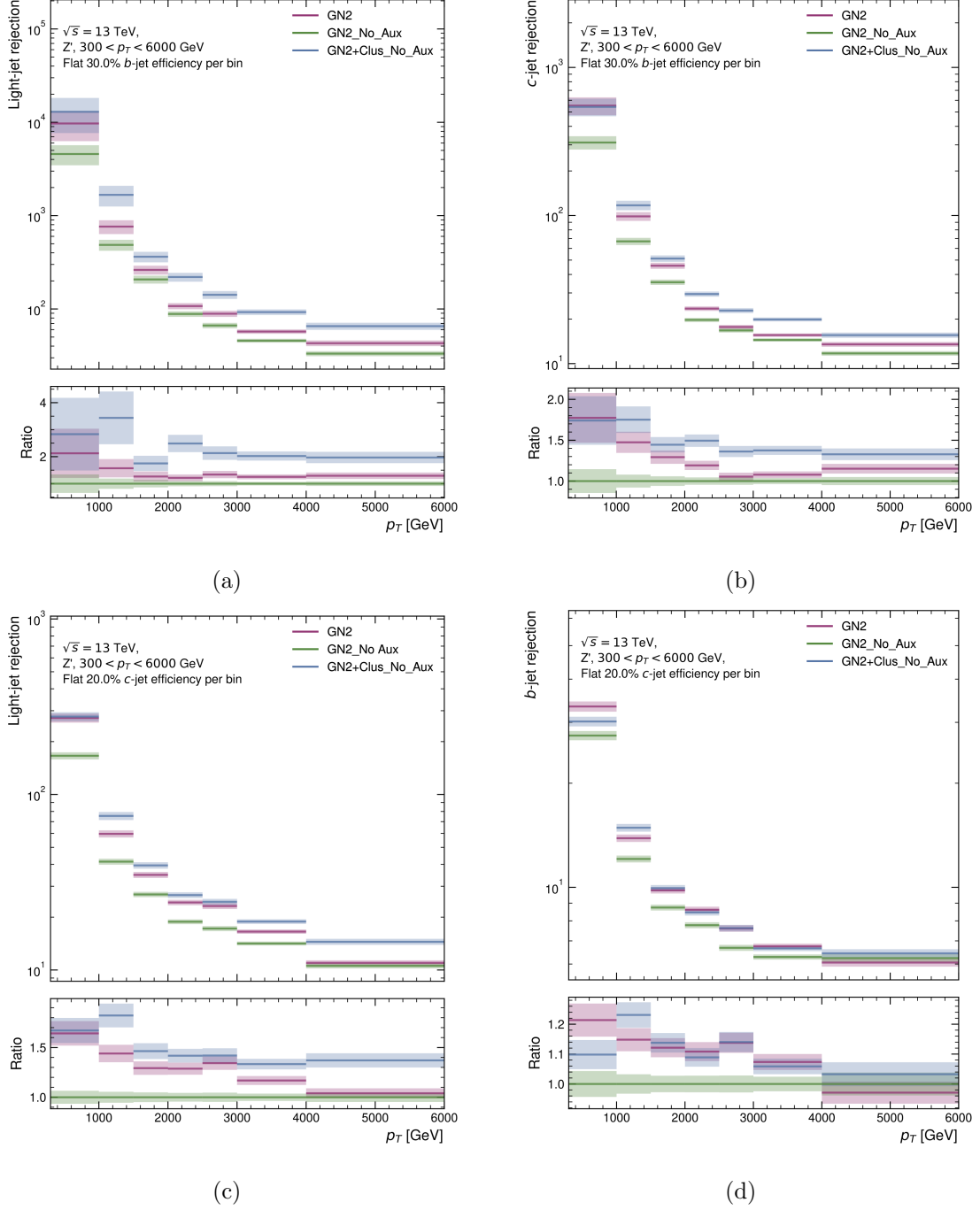


Figure 4.24: b -tagging (a and b) and c -tagging (c and d) background rejection rates of the GN1ClusOnly, GN2ClusOnly and GN2 algorithm as a function of the jets p_T . The WP cut values are recomputed in every p_T bin to ensure a flat signal efficiency.

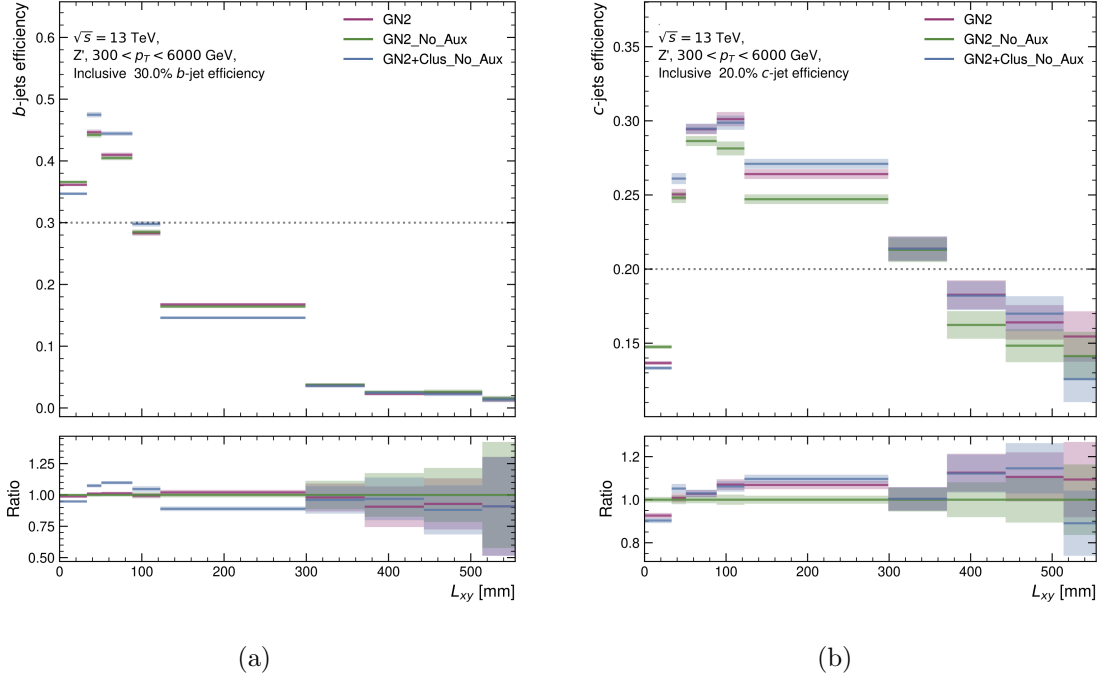


Figure 4.25: b -tagging (a) and c -tagging (b) signal efficiencies as a function of the transverse decay length of the heavy-flavour hadron associated to the heavy-flavour jet for a fixed cut on the D_x distribution. The edges of the bins correspond to the locations of the active layers in the ID (Fig. 2.4), *i.e.*, the IBL is located at the edge between the first and the second bin, the B-layer between the second and the third, and so on.

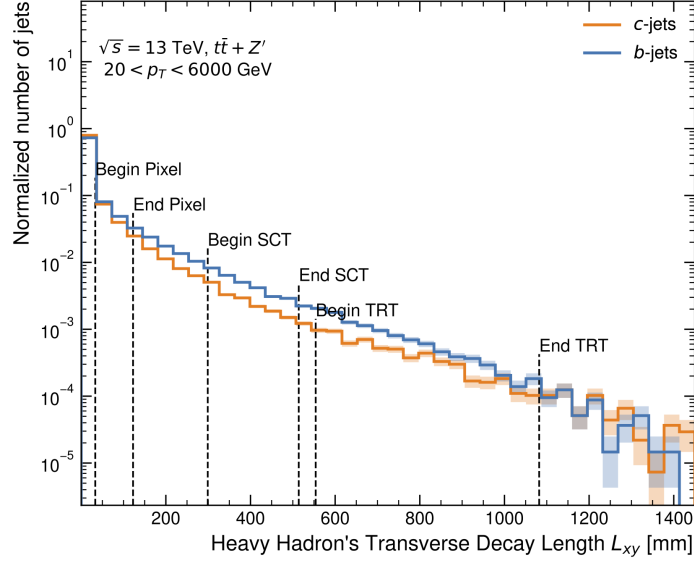


Figure 4.26: Transverse decay length L_{xy} of heavy-flavour hadrons associated with heavy-flavour jets in the hybrid $t\bar{t} + Z'$ sample used to train GN2+Clus.

saved in a cone with an angular opening $\Delta R = 0.1$ around the jet axis. The value $\Delta R = 0.04$, used to train the models focusing only on the very high- p_T jets (Sec. 4.3.1 and 4.3.2), was chosen by studying the angular distribution of clusters produced during b -hadron decays in the Z' sample (Fig. 4.8a). Since the decay products of heavy-flavour hadrons with lower momenta are less collimated, ΔR was increased for this training including jets from the $t\bar{t}$ sample.

Another novelty is the introduction of clusters of hits of silicon strips in the SCT layers of the ID (Sec. 2.2.1). Whereas most of the heavy-flavour hadrons decay before the outermost layer of the pixel detector, some have Lorentz boosts large enough to reach the SCT layers and beyond, as demonstrated in Fig. 4.26 by showing the transverse decay length L_{xy} of the heavy-flavour hadrons associated with heavy-flavour jets belonging to the $t\bar{t} + Z'$ hybrid sample. Including the SCT clusters enables exploiting the increase in the number of clusters of silicon hits schematized in Fig. 4.8 also for these very displaced decays.

The training sample was preprocessed with the procedure outlined in Sec. 4.3.1, resulting in a hybrid jet collection comprising 25.6M jets, with jets of the different classes equally distributed in p_T and η . The input variables used to represent the jets and tracks are the same as listed in Tab. 3.4. Variables to represent the clusters are as of Tab. 4.1, with the addition of the transverse distance to the PV $r = \sqrt{x^2 + y^2}$ and with the exclusion of splitProbability, which is not defined for clusters of silicon strips⁵. Distributions of jet,

⁵This is due to the worse spatial resolution of the SCT strips compared to the pixels', which worsens the accuracy of such information; and because such variable is mostly used to improve tracking in dense

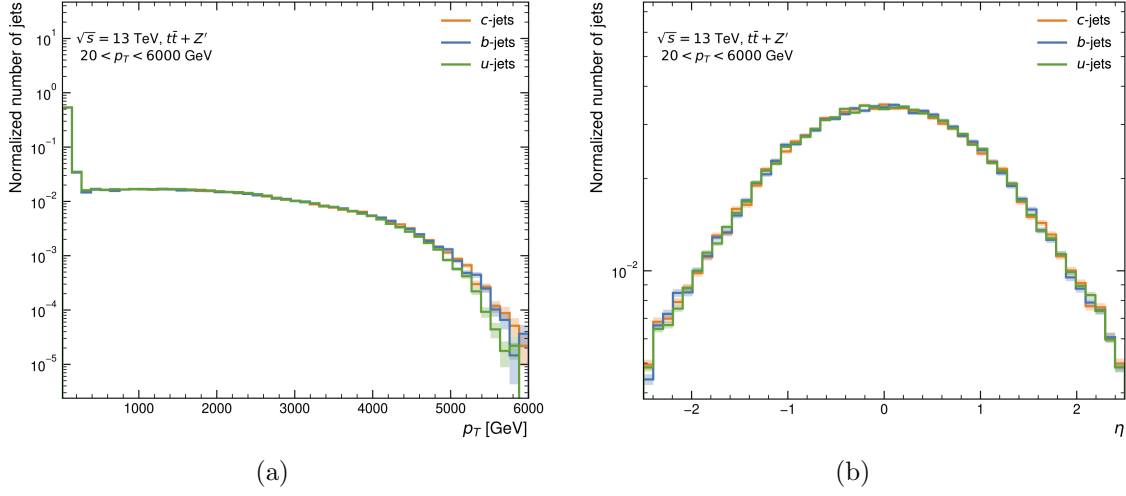


Figure 4.27: Input variables used by the GN2+Clus algorithm to describe the jets' features.

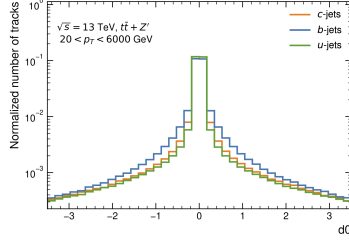
track, and cluster variables after resampling but before scaling and shifting are displayed in Fig. 4.27, 4.28 and 4.29 respectively.

GN2+Clus was trained with the same procedure outlined in Sec. 4.3.1. In addition, a re-training of GN2 was performed on the same sample, to allow a fair comparison of the models in order to understand the impact of introducing the clusters on the jet flavour-tagging performance. Fig. 4.30 shows the validation losses of the algorithms during training, computed as in Eq. 3.10. The epochs that minimize the loss are epochs 99 and 88 for GN2+Clus and GN2, respectively.

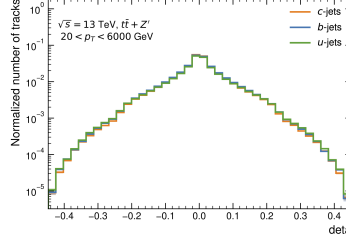
Performance on the $t\bar{t}$ sample The performance of the algorithms is studied on jets belonging to separate $t\bar{t}$ and Z' samples, allowing for a characterization of the models along a wide range of p_T values. This paragraph focuses on the $t\bar{t}$ sample.

Fig. 4.16 shows how the mistagging efficiencies on background jets change when varying f_x . Fig. 4.31 shows how the mistagging efficiencies of the background jets change when varying f_x using the 70% b -tagging WP and for the 50% c -tagging WP. Whereas for b -tagging (Fig. 4.31a) the value $f_c = 0.1$ adopted by the GN2v00 algorithm strikes a good balance between the c -jet and light-flavour jet rejection rates, for c -tagging (Fig. 4.31b) the nominal $f_b=0.2$ value results in a different trade-off between the classes of background jets for GN2v00 compared to GN2 and GN2+Clus. To ensure a fair comparison of the performance of the models, *optimal* f_b values are recomputed for all the algorithms, choosing the points of the curves in Fig. 4.16 that minimize the distance to the origin. Fig. 4.32 compares the ROC curves of the algorithms. Given the ~ 10 times bigger statistics used to train it, GN2v00

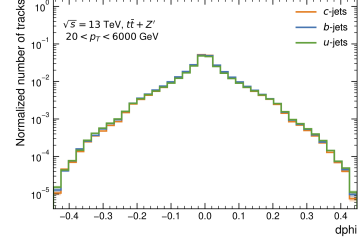
environments, and the distance of the SCT from the collision generally allows the particles to spread more before producing signals in the SCT sub-system.



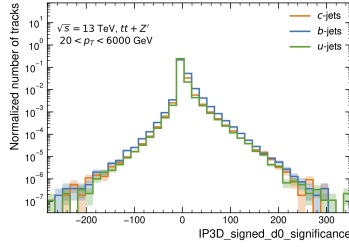
(a)



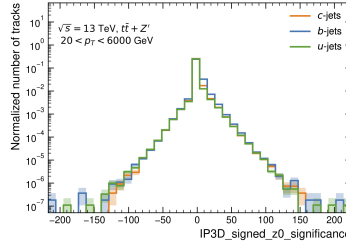
(b)



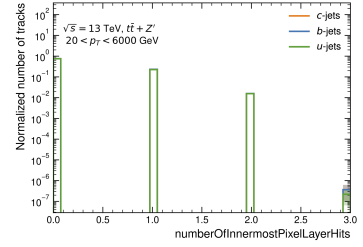
(c)



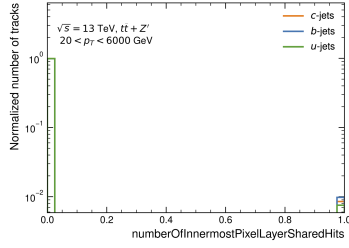
(d)



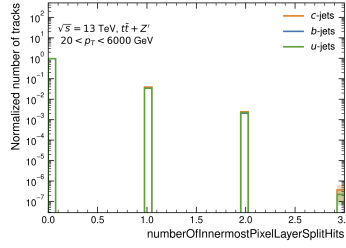
(e)



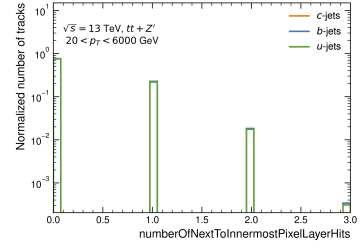
(f)



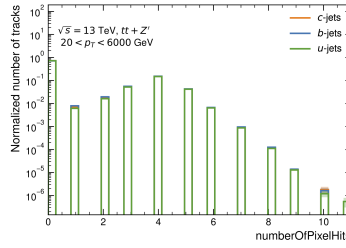
(g)



(h)



(i)



(j)

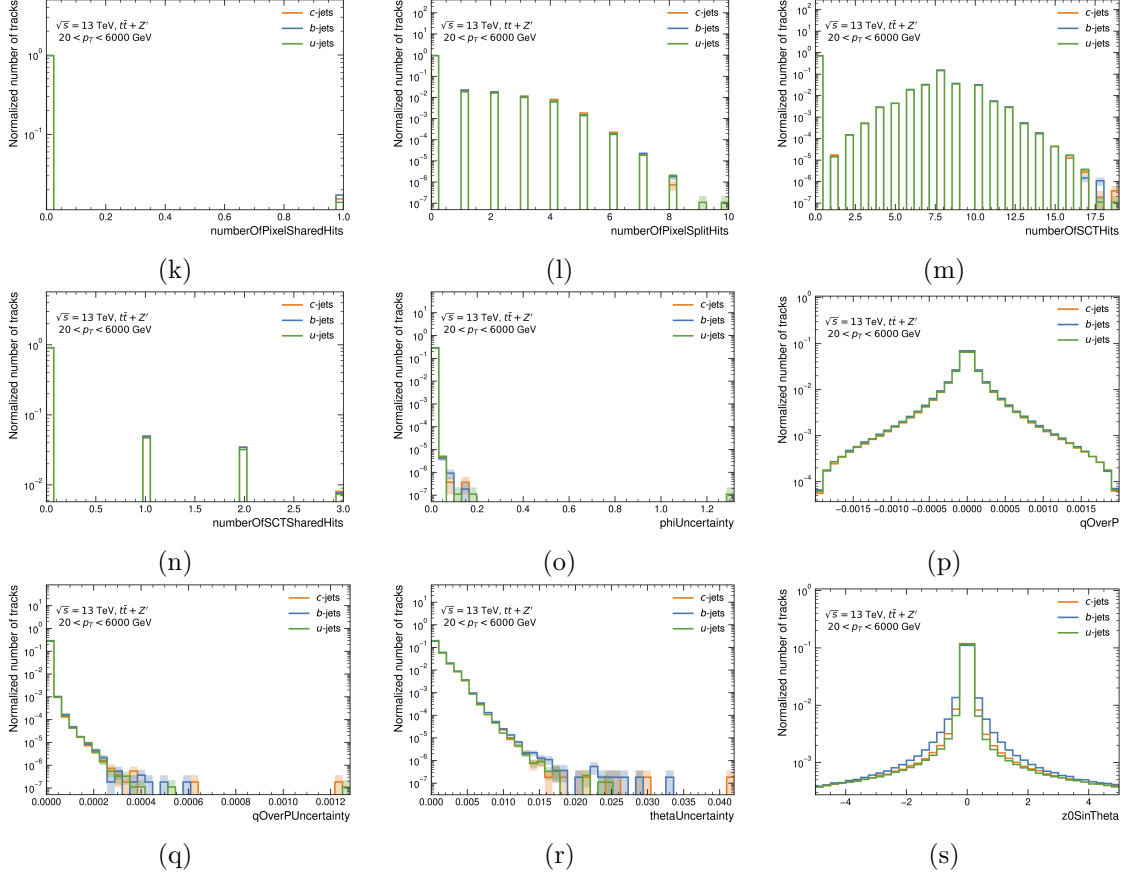


Figure 4.28: Input variables used by the GN2+C1us algorithm to describe the tracks' features.

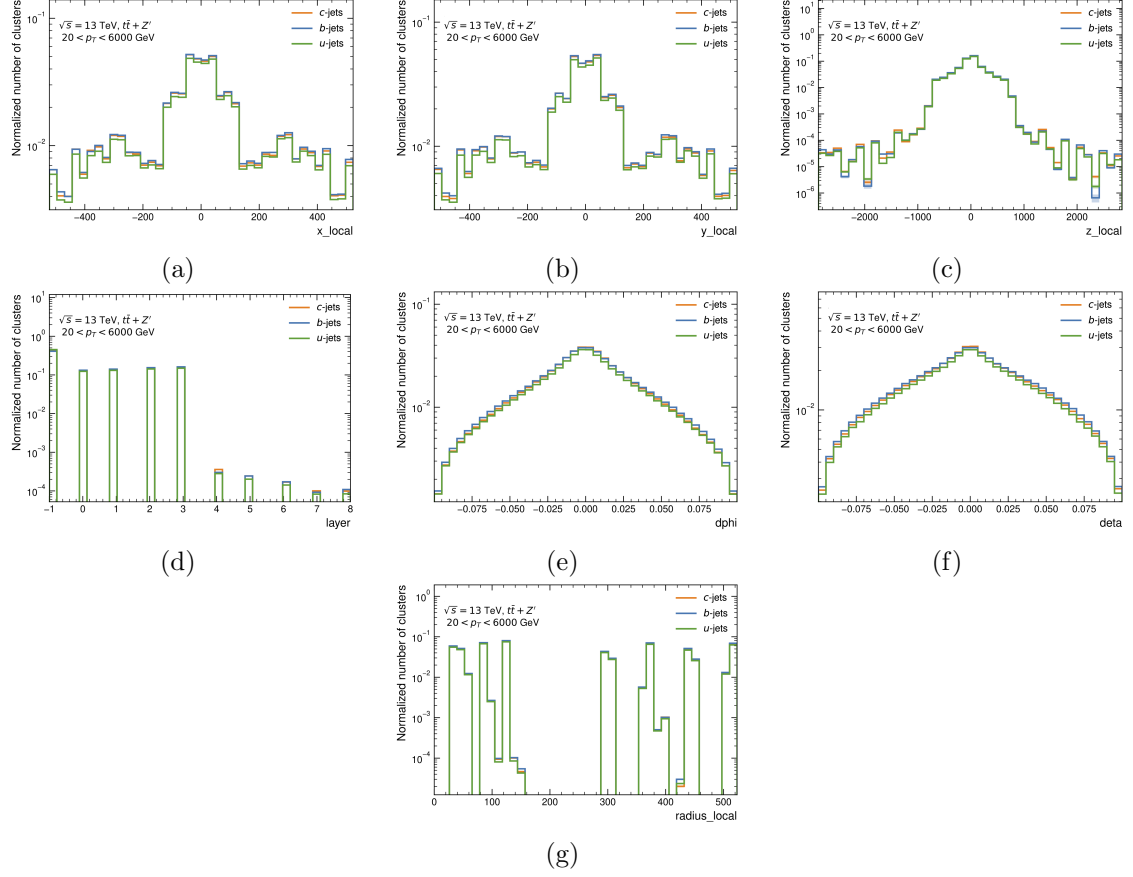


Figure 4.29: Input variables used by the GN2+Clus algorithms to describe the clusters' features.

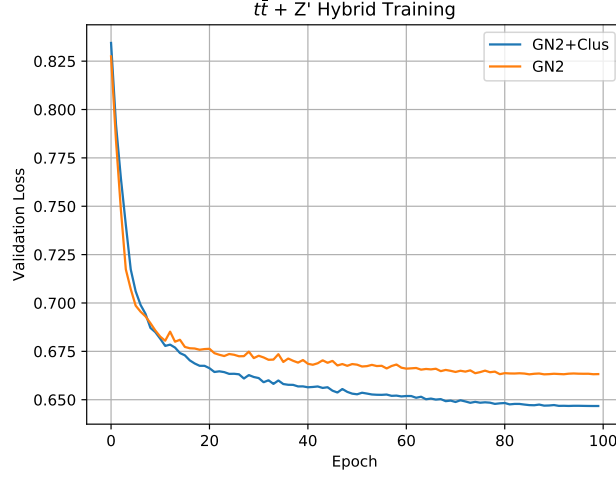


Figure 4.30: Validation losses of the GN2 and GN2+Clus algorithms.

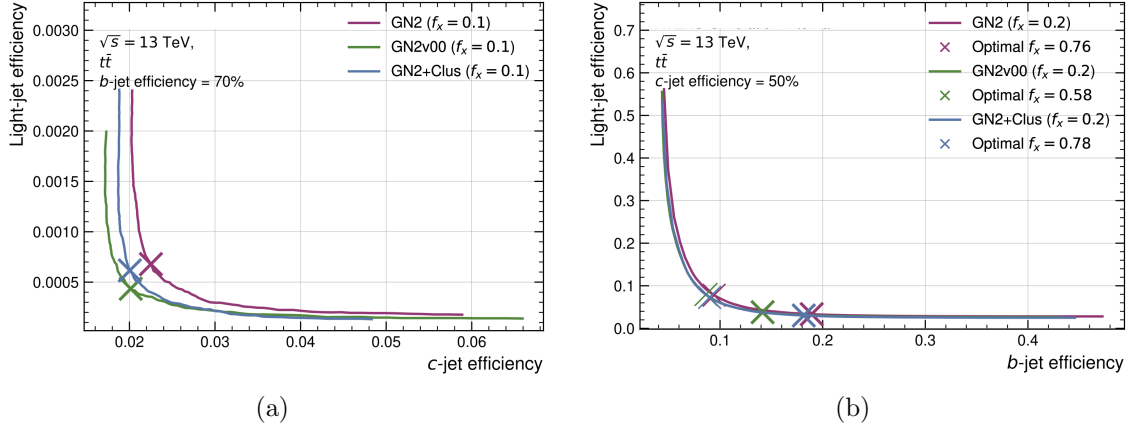


Figure 4.31: Background jets mistagging efficiencies of GN2v00, GN2+Clus and of the re-trained GN2 algorithm as a function of the f_c (a) and f_b (b) parameters. The thick cross markers indicate the values adopted for the nominal GN2v00 algorithm. To guarantee a fair comparison of the c -tagging performance, given the different trade-offs between the backgrounds struck by the nominal f_b value, an optimal value was adopted, defined as the point on the signal isoefficiency curve minimizing the distance from the origin and indicated by the thin markers. Values are computed on the $t\bar{t}$ validation sample.

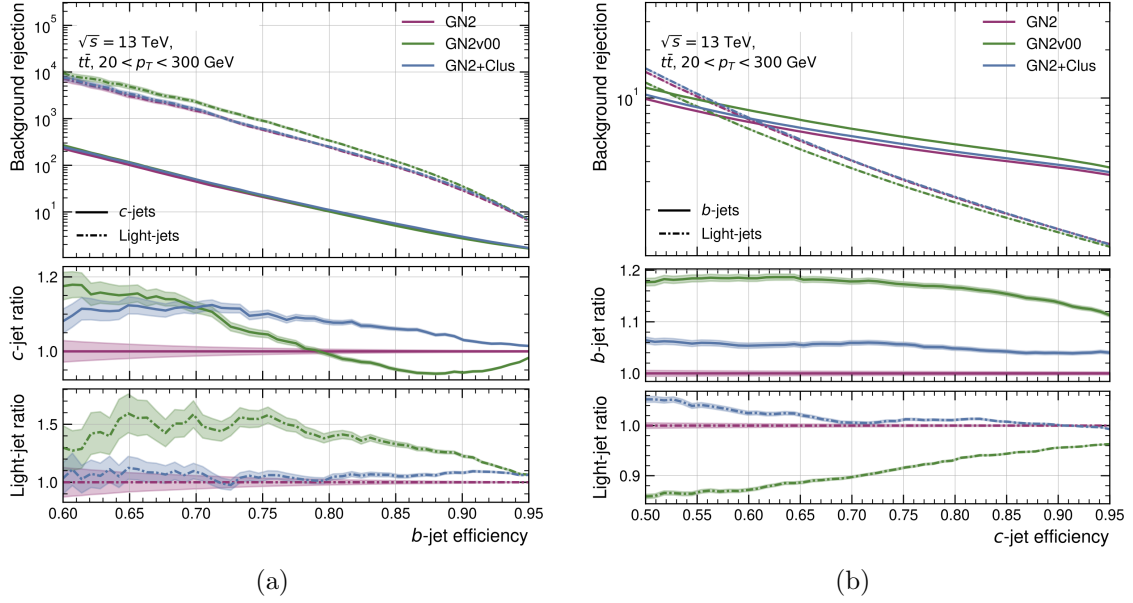


Figure 4.32: b -tagging (a) and c -tagging (b) ROC curves of the GN2v00, GN2+Clus and of the retrained GN2 algorithms when evaluated on the jets of the $t\bar{t}$ validation sample.

stands out in the b -tagging task ⁶. When focusing on GN2+Clus and GN2, trained over the same jets, the former shows some improvement, as it is also provided with information about the clusters.

A closer look at the p_T distribution of the jets in the $t\bar{t}$ validation sample (Fig. 4.33) allows to refine the characterization of the performance of the models. Most of the jets populate the first bin, which has the lowest p_T . That bin is where the smallest improvement brought about by introducing the clusters is expected. The isoefficiency and ROC curves that were shown in Fig. 4.31 and 4.32 mostly represent performance on this low p_T , highly populated bin. All those possible improvements in higher p_T bins, gained by introducing the clusters, are there washed away: lost in time, like tears in the rain. To serve them justice, it is necessary to study the performance of the algorithms as a function of p_T . This raises a problem, given the extreme light-flavour jet rejection power held by these models in this p_T range. The problem is clear in Fig. 4.34a: very few background jets are misidentified, hence, the rejection rate is statistically very uncertain in all the bins. This, even though many bins have been merged with respect to the similar plots shown in the previous sections. To better compare the algorithms, the looser 85% b -jet efficiency WP is adopted henceforth. Its corresponding threshold on the D_b distribution is displayed in Fig. 4.35. Choosing a looser WP increases both the signal efficiency and the background mistagging efficiencies.

⁶The different trade-off between b -jet and light-flavour jet rejection rates visible in the ratios of the c -tagging ROC curve (Fig. 4.32b) is caused by the chosen f_b values. The number of background jets rejected by the algorithms is nonetheless very similar.

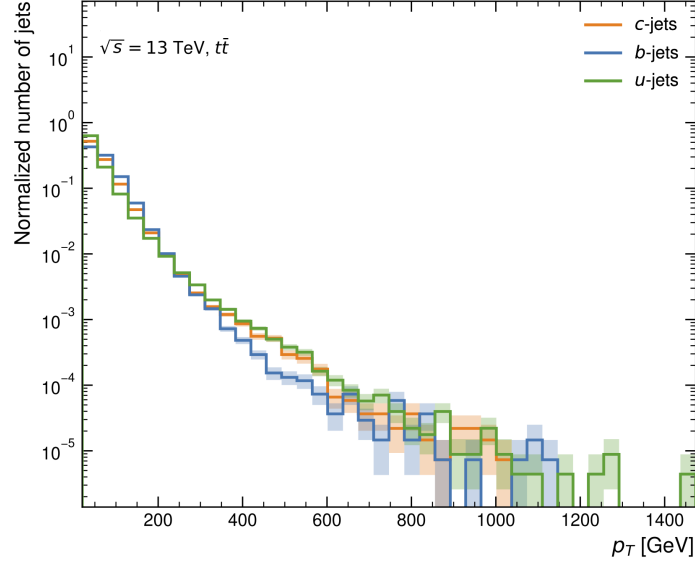


Figure 4.33: p_T distributions of jets in the $t\bar{t}$ validation sample used to evaluate the performance of the GN2, GN2v00, and GN2+Clus algorithms. Most of the jets populate the first bin.

This results in an assessment of the rejections less affected by statistical uncertainties. The optimal f_c value adopted for this WP is displayed in Fig. 4.36.

Fig. 4.34b and 4.34d show the light-flavour jets and c -jets rejection rates of the algorithms for the 85% b -tagging WP as a function of the jets p_T . GN2v00 excels in the first p_T bin, but GN2+Clus catches up as the transverse momentum increases. This means that, above the first, lowest p_T bins, the clusters can provide as much additional information as what is gained with GN2v00 over GN2 by training on a dataset almost ten times as big. This happens already for bins covering relatively low p_T values. Similar considerations hold when considering Fig. 4.37, which shows the b -jet and light-flavour jet rejections for the 50% c -tagging WP as a function of the jets p_T ⁷.

Performance on the Z' sample Fig. 4.38 shows how the mistagging efficiencies of the background jets change when varying f_x using the 60% b -tagging WP⁸ and for the 20% c -tagging WP. The adopted f_x values are those optimized on the $t\bar{t}$ sample. Fig. 4.39 showcases the ROC curves of the algorithms. As far as b -tagging is concerned, GN2+Clus improves the rejection rates of the background jets of significant factors over a wide range of WPs, with peaks of $\sim 80\%$. The improvement in c -tagging is less pronounced, of the

⁷The different balance struck by the GN2v00 algorithm between b -jet and light-flavour jet rejection rates is caused by the chosen f_b values.

⁸The 60% b -tagging WP on the Z' sample uses the same threshold corresponding to the 85% WP on the $t\bar{t}$ sample.

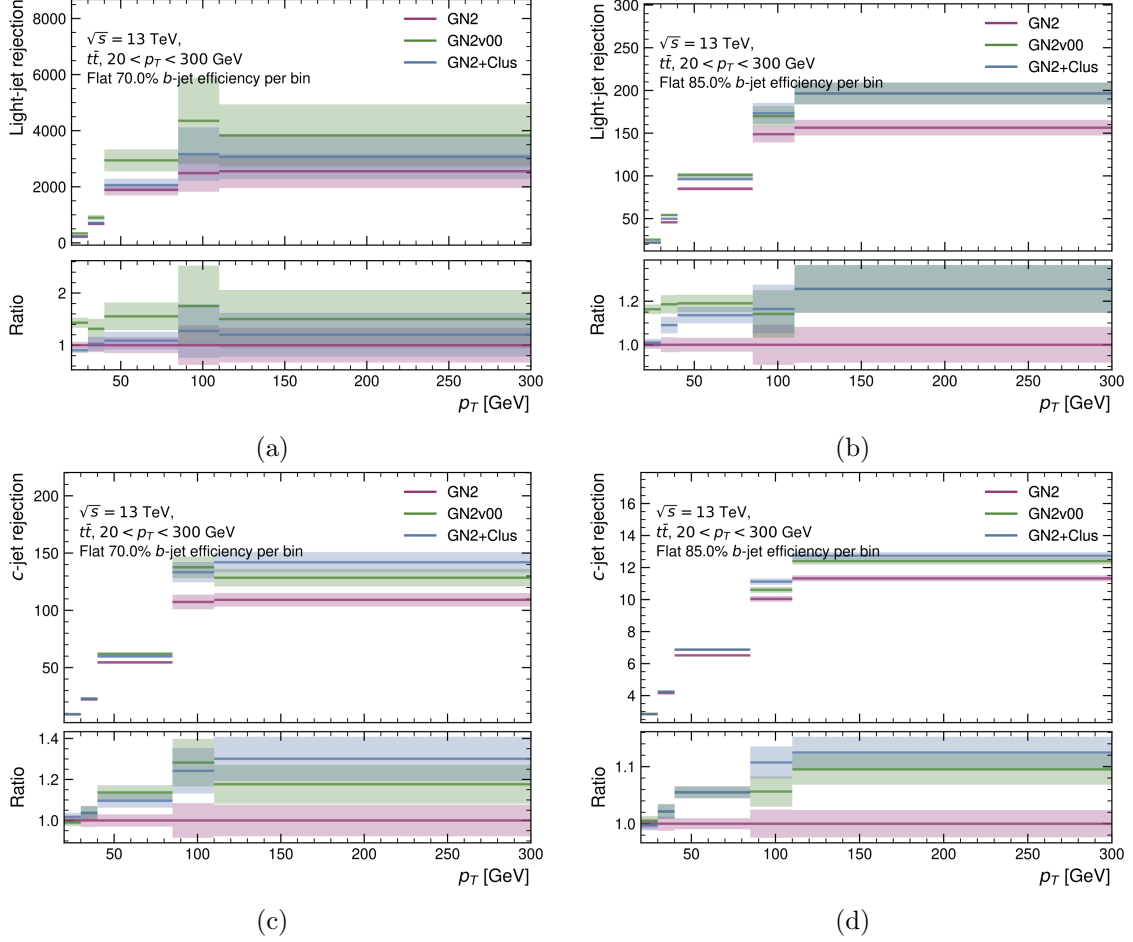


Figure 4.34: (a) light-flavour jets rejections as a function of p_T for at the 70% b -tagging WP of the GN2v00, GN2+Clus and of the retrained GN2 algorithm. The WP cut values are recomputed in every p_T bin to ensure a flat b -jet efficiency. The first bin has enough statistics to tell the rejections apart, whereas, in the others, the rejection rates are compatible within the (very large) statistical uncertainties. (b) The b -tagging WP is moved to 85%. In this case, the light-flavour jets rejection rates, albeit smaller, are assessed with smaller uncertainty. (c) and (d) c -jets rejections as a function of p_T for the 70% and 85% b -tagging WP respectively.

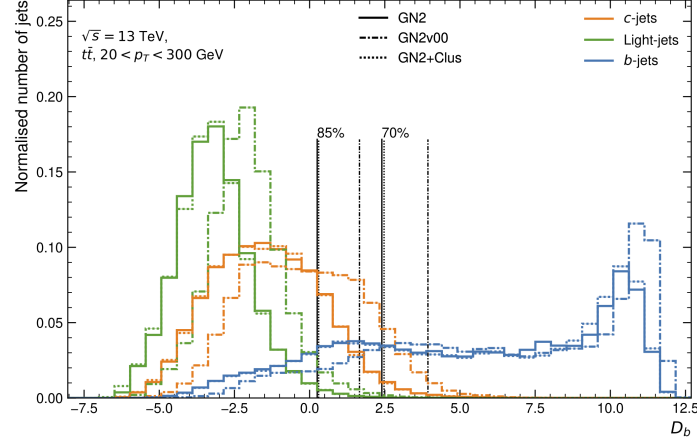


Figure 4.35: D_b distribution of the GN2v00, GN2+Clus and of the retrained GN2 algorithm, computed on jets belonging to the simulated $t\bar{t}$ sample. The vertical bars display the threshold corresponding to the 85% and 70% b -tagging WPs. With the looser 85% WP, more background jets are misidentified as b -jets. This makes the assessment of the rejection performance of the algorithms less affected by statistical uncertainties.

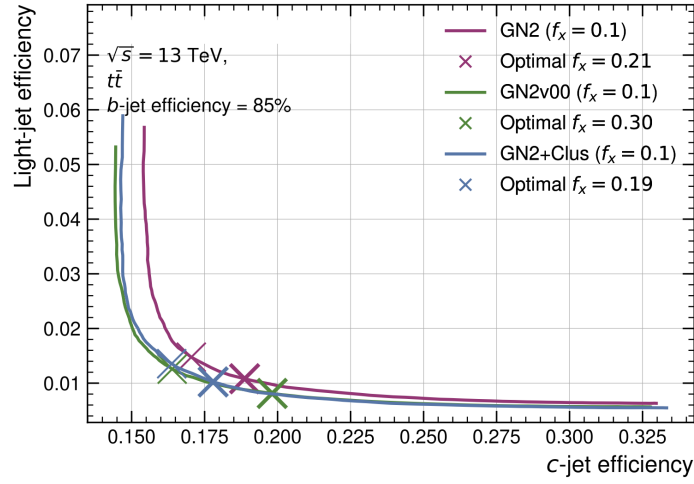


Figure 4.36: Background jets mistagging efficiencies of GN2v00, GN2+Clus and of the retrained GN2 algorithm as a function of the f_c parameter. Values are computed on the $t\bar{t}$ validation sample.

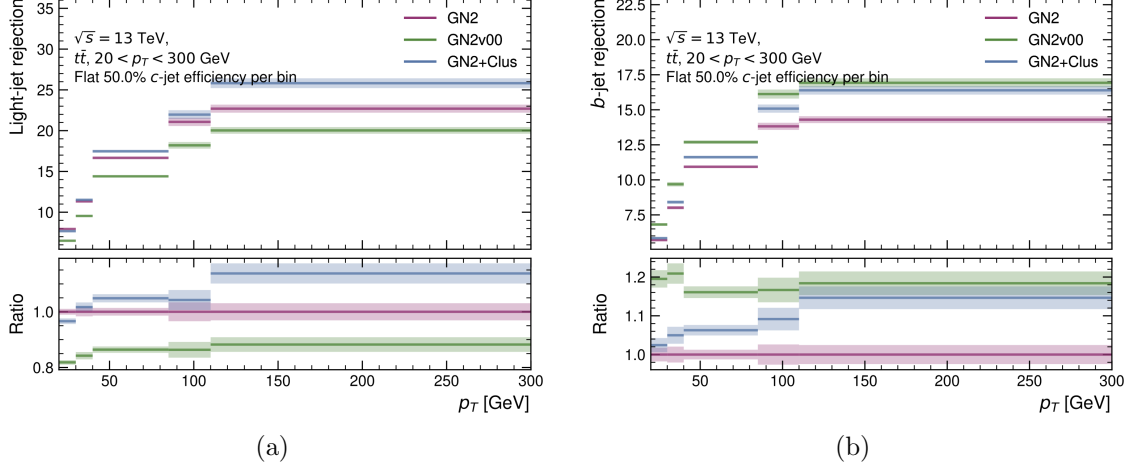


Figure 4.37: c -tagging background rejection rates of the GN2v00, GN2+Clus and of the retrained GN2 algorithm as a function of the p_T of jets belonging to the simulated $t\bar{t}$ sample. The WP cut values are recomputed in every p_T bin to ensure a flat 50% c -jet efficiency.

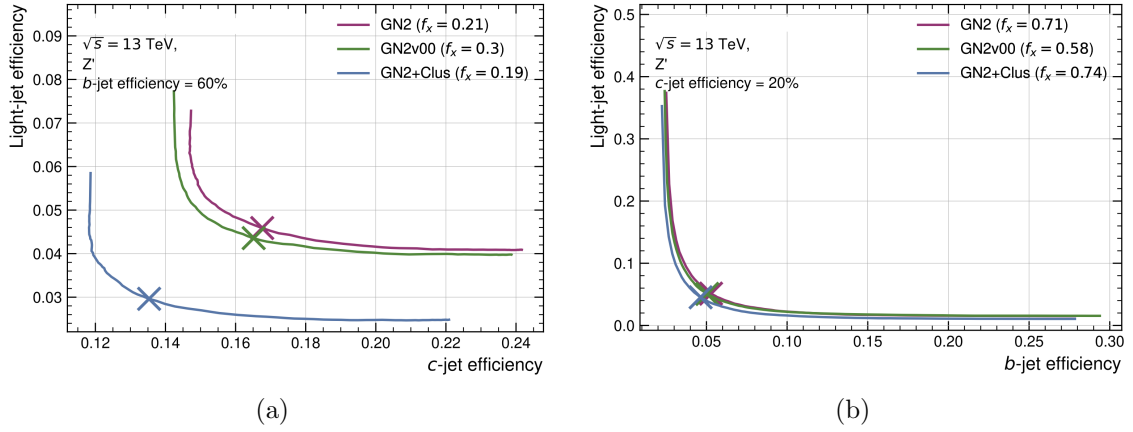


Figure 4.38: Background jets mistagging efficiencies of GN2v00, GN2+Clus and of the retrained GN2 algorithm as a function of the f_c (a) and f_b (b) parameters. Values are computed on the Z' validation sample.

order of 10%. Fig. 4.40 shows how this improvement is distributed across the p_T spectrum. The improvement is concentrated in the 300-1800 GeV bins. This aligns with what was observed in the preliminary studies within the ATLAS collaboration detailed in Sec. 4.2.

Finally, the improvement brought about by including the clusters of SCT hits is studied by comparing the performance of two **GN2+Clus** models trained with and without these clusters. Fig. 4.41 shows the background jet rejection rates of the resulting **GN2+Clus** algorithms as a function of the jets p_T . The model trained only using clusters of pixel hits has slightly decreased b -tagging rejection rates. It is particularly interesting to compare the rejection rates of the two models as a function of the transverse decay length of the heavy-flavour hadrons within the heavy-flavour jets, shown in Fig. 4.42. Most of the improvement brought about by the SCT hits is concentrated in the two bins corresponding to heavy-flavour hadrons decays happening between the two last pixel layers and between the pixel and SCT subsystems. This observation indicates that retaining more clusters produced by particles stemming from the decay of the heavy-flavour hadron is beneficial for identifying the heavy-flavour jets. For decays happening in the outermost parts of the pixel subsystem or even beyond its reach, few or none of these clusters are available to the model utilizing only the pixel hits. Thus, its performance is decreased.

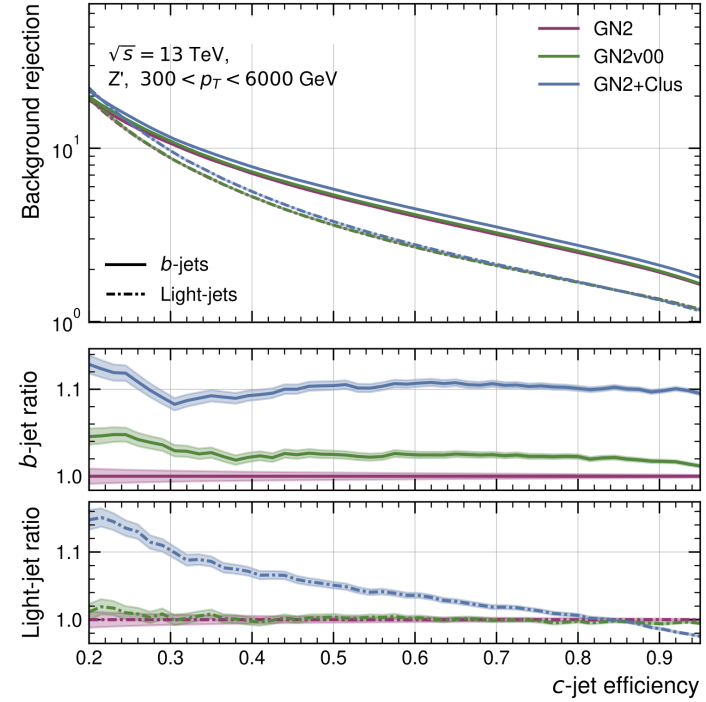
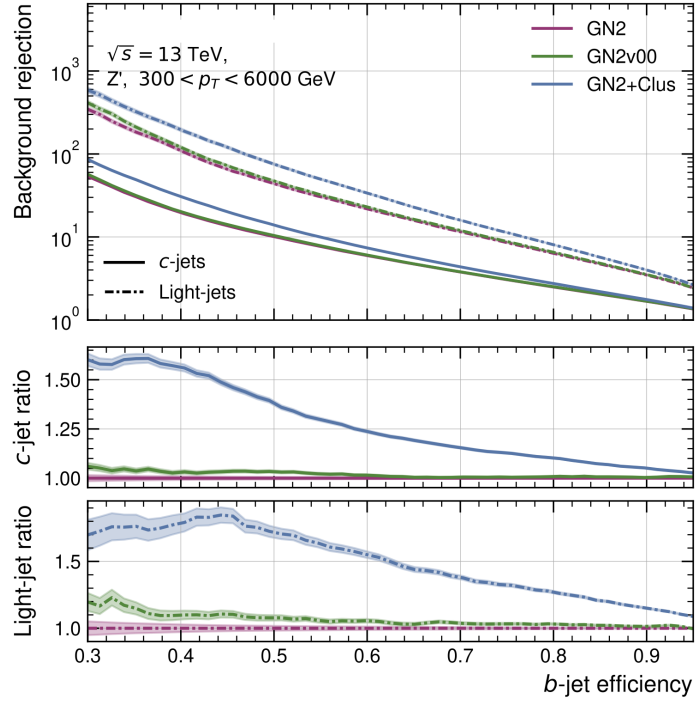


Figure 4.39: b -tagging (left) and c -tagging (right) ROC curves of the GN2v00, GN2+Clus and of the retrained GN2 algorithms when evaluated on the jets of the Z' validation sample. GN2+Clus outperforms the background rejection rates of the other algorithms of significant factors over a wide range of WPs, with peaks of $\sim 80\%$ when b -tagging and $\sim 10\%$ when c -tagging.

Outlook The studies presented in this section highlight that incorporating information about the distribution of silicon hit clusters into the jet flavour-tagging algorithms used by the ATLAS Collaboration results in considerable performance improvements across a broad range of transverse momenta when evaluated on simulated jet samples. Looking ahead, two main research directions are outlined to advance these studies to integrate algorithms such as **GN2+Clus** into physics analyses.

The first direction consists of searching for ways to improve the performance of these algorithms. Possible improvements include testing additional input variables that provide new, previously unused information, such as the involvement of a cluster in track reconstruction. Modifying the internal structure of the GN block of the algorithm could also be beneficial. For example, instead of fully connecting cluster nodes and track nodes as vectors embedded into the same vector space, they could be treated as separate graphs and interconnected at a subsequent stage. More intricate architectures might also be investigated, where clusters and tracks are connected according to principles based on physical intuition: for instance, connecting cluster nodes with the node of the closest track, or connecting clusters used to reconstruct a track with that track's node. Furthermore, leveraging truth information about the clusters could help develop auxiliary tasks that guide the algorithm towards a more profound, or at least more interpretable, understanding of the physical processes that influence the structure of the jets. Similarly to how **GN2** is trained for 2-track vertexing, **GN2+Clus** could be adapted to perform 2-cluster tracking; and just as **GN2** classifies the origins of tracks, **GN2+Clus** could be adapted to classify the origins of clusters (Sec. 3.4.1).

The second research line focuses on validating the performance of these algorithms using samples simulated with different MC generators and calibrating them on real data. This step is critical, as using clusters of silicon hits for jet flavour-tagging is a novel approach, and the extent to which MC generators accurately model those aspects that are used by **GN2+Clus** needs to be thoroughly examined. Distributions of clusters of silicon hits in jets are shaped by the interplay of various physics processes, from prompt hard decays to soft QCD radiation, and unlike tracks, clusters do not individually convey the energy of the generating process. It is therefore essential to ensure that the observed performance enhancements are not merely due to MC-dependent characteristics of the simulated jets but are firmly rooted in real physics. Consequently, the performance of algorithms like **GN2+Clus** also needs to be monitored when evaluated on jets simulated with different MC generators. In terms of calibrating these algorithms with real data, adjusted-MC methods such as that outlined in [92] can be used to ascertain the impact of this new type of input as a way of validating the results of traditional calibration methods. These methods consist in adjusting the cluster variables in the MC to match the data, and examining how these changes affect the performance of the jet flavour-tagging algorithm evaluated on the adjusted MC.

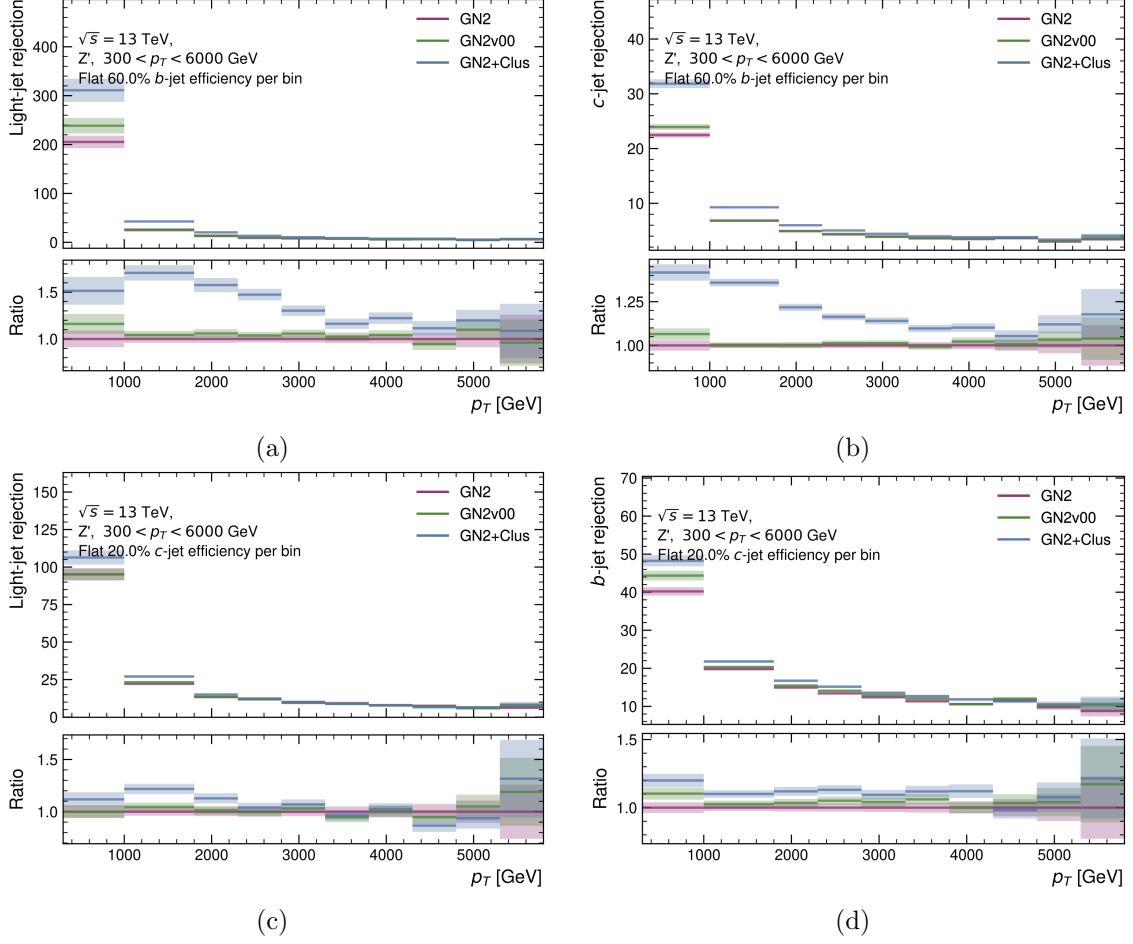


Figure 4.40: b -tagging (a and b) and c -tagging (c and d) background rejection rates of the GN2v00, GN2+Clus and of the retrained GN2 algorithms as a function of the jets p_T . The WP cut values are recomputed in every p_T bin to ensure a flat signal efficiency. GN2+Clus outperforms the other algorithms particularly in the 300-1800 GeV bins, in line with what was observed in the preliminary studies within the ATLAS collaboration

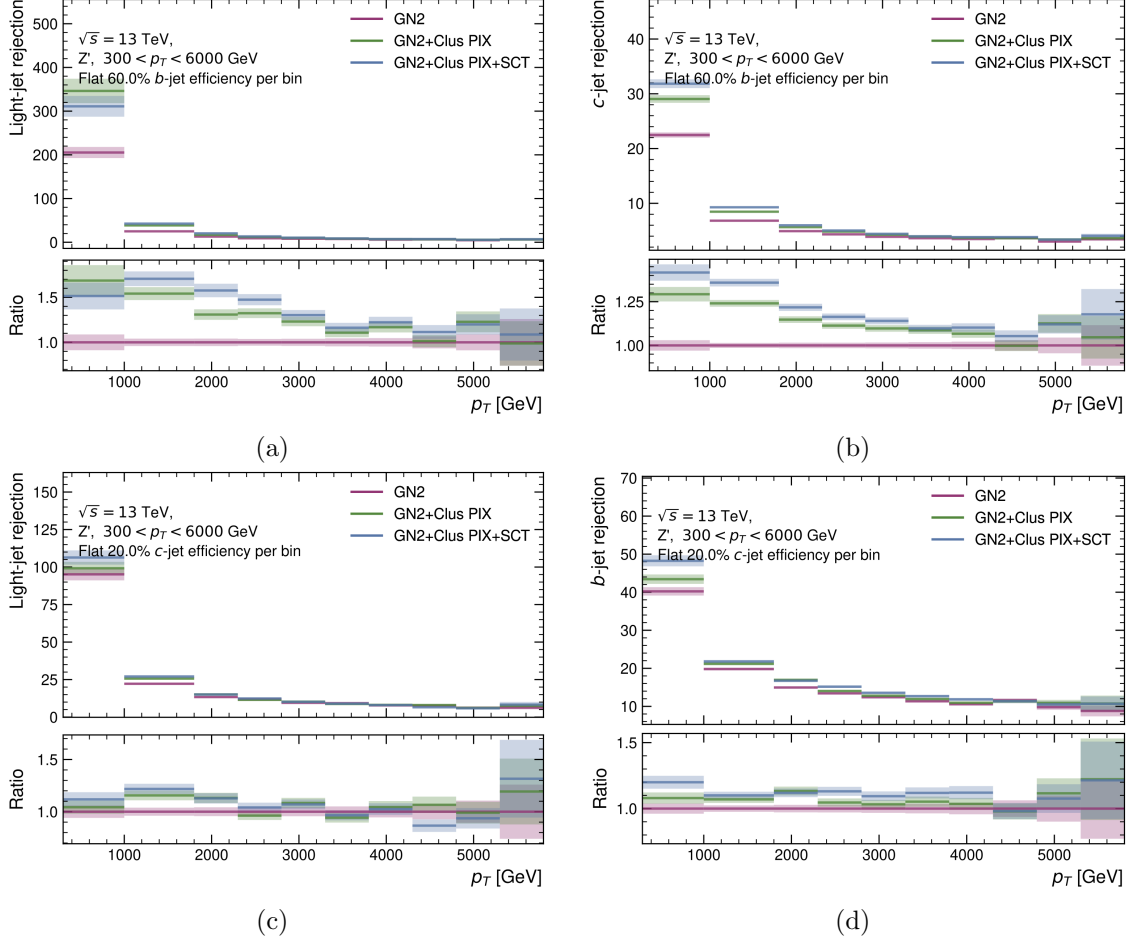


Figure 4.41: b -tagging (a and b) and c -tagging (c and d) background rejection rates of the GN2+Clus algorithm trained with and without utilizing the clusters of SCT hits, and of the retrained GN2 algorithm as a function of the jets p_T . The WP cut values are recomputed in every p_T bin to ensure a flat signal efficiency. Not using the cluster of SCT hits slightly decreases the b -tagging performance.

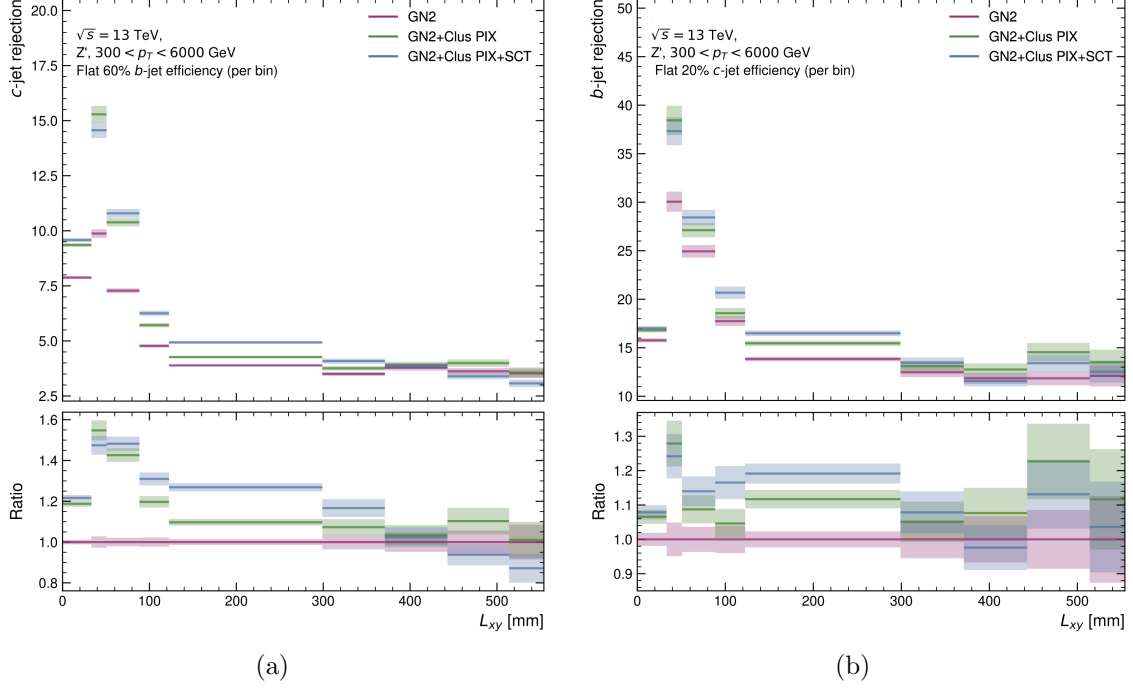


Figure 4.42: b -tagging (a) and c -tagging (b) background rejection rates of the GN2+Clus algorithm trained with and without utilizing the clusters of SCT hits, and of the retrained GN2 algorithm as a function of the transverse decay length of the heavy-flavour hadron associated to the heavy-flavour jet. The WP cut values are recomputed in every bin to ensure a flat signal efficiency. The edges of the bins correspond to the locations of the active layers in the ID (Fig. 2.4), *i.e.*, the IBL is located at the edge between the first and the second bin, the B-layer between the second and the third, and so on.

Chapter 5

Truth jet flavour-tagging with GNNs

Jet flavour-tagging algorithms are crucial tools for physics analyses studying processes with heavy-flavour jets in the final state. The simplest way to use them for assigning a flavour hypothesis to the jets can be defined as *pass-or-fail*, meaning that only the jets selected with a given WP are considered as flavour-tagged. Due to the high background rejection rates achieved by modern algorithms like DL1, GN1, GN2 and GN2+Clus (Chap. 3), this approach can require the production of extremely large simulated samples, which are used for modelling the background processes in many physics analyses. The problem is exacerbated when analysing collision events with multiple flavour-tagged jets, particularly when focusing on background processes enriched in light-flavour jets in the final state.

A good example is the background of Z + light-flavour jets events accepted in the signal region requiring 2 b -tagged jets of the analysis measuring the associated production of a Higgs boson and a vector gauge boson in the $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decay channels, detailed in Chap. 6. Such analysis uses the DL1r algorithm to identify the heavy-flavour jets. Fig. 3.22a shows that, for a 70% b -tagging WP, DL1r achieves a light-flavour jet rejection rate of $\sim 10^3$ when applied on jets with transverse momentum between 20 and 250 GeV. Thus, only a fraction of $\sim 10^{-3} \times \sim 10^{-3} = \sim 10^{-6}$ of the simulated Z + light-flavour jets events can be used to model the scenario in which both light-flavour jets are mistagged as b -jets¹. The resulting sample of Z + light-flavour jets events passing the flavour-tagging cut is very limited in statistics. This is a problem since this background sample is then used to model distributions of variables entering into the fit to data: large statistical uncertainties in the background templates can cause instabilities in the fit and propagate large uncertainties in

¹The analysis actually employs a *pseudo-continuous* flavour tagging approach, in which a more sophisticated definition of flavour-tagged jet is derived by defining bins in the D_b and D_c distributions of the DL1r tagger (Sec. 6.2.1). Hence, the percentages given here are not accurate, but rather useful for providing a sense of the scales involved in the problem.

the final result.

What was just described is an example of an overarching issue for experiments at the LHC. Applying selections on the flavour-tagging scores strongly restricts the phase space in which the analysis is conducted. The problem is that pantagruelian samples of collision events need to be simulated to estimate acceptance rates in very restricted regions of the phase space. An alternative approach to address this type of challenge is the event weighting technique, which replaces selection cuts with event weights. *Truth jet flavour-tagging* [93,94], described in this section, is an incarnation of this technique, providing an alternative approach to pass-or-fail flavour-tagging. It consists in weighting the events with their likelihood of having the required number of flavour-tagged jets.

Considering a counting experiment, the statistical uncertainty associated with the count N_s of events passing a selection is $\sqrt{N_s}$. If $N > N_s$ is the size of the sample before the selection, weighting the events with their probability w_i of passing the selection yields an estimate of N_s whose statistical uncertainty is computed with the variance of the weights $\sqrt{(\sum_{i \leq N} w_i^2)}$.

Therefore, weighting the events rather than applying the selection reduces the relative statistical uncertainty on the estimate of N_s if, as it typically happens, the variance of the weights is small compared to the poissonian uncertainty $\sqrt{N_s}$:

$$\frac{\sqrt{N_s}}{N_s} > \frac{\sqrt{(\sum_{i \leq N} w_i^2)}}{N_s} \text{ if } \sqrt{N_s} > \sqrt{(\sum_{i \leq N} w_i^2)}. \quad (5.1)$$

Thus, if the jet flavour-tagging efficiency function ϵ_f , defined in Eq. 3.4, is well known, it is possible to use it to compute event weights for the truth jet flavour-tagging technique, yielding samples distributed with the same overall shapes and normalisations that would be obtained with pass-or-fail flavour-tagging, but with a reduced statistical uncertainty.

This approach proves effective only when the ϵ_f function is appropriately parameterized. The following sections outline the difficulties in achieving such parametrization and presents a novel approach based on GNNs to address this challenge.

5.1 Jet flavour-tagging efficiency parametrization

The function ϵ_f is not known *a priori*. It can be thought of as the probability for a jet of flavour f to be classified as a heavy-flavour jet by using a jet flavour-tagging algorithm trained on a set of input variables denoted as $\mathbf{x}$. These variables, in turn, depend non-trivially on a set of underlying parameters θ whose members are only partially known. In the case of DL1r, $\mathbf{x}$ consists of the variables listed in Tab. 3.1. The most important parameters included in θ are the p_T and η of each jet, but there are others. For instance,

the amount of PU collisions and the proximity to other jets in the event both influence the probability ϵ_f of tagging a jet [95].

An intuition of the relation between input variables $\mathbf{x}$ and underlying parameters θ can be gained by reading again Sec. 4 under this perspective. There, a detailed description of the various ways in which the p_T of the jets influences the flavour-tagging efficiency ϵ_f is provided. Even though the different classes of jets used to train the jet flavour-tagging algorithms are resampled to share the same p_T distribution, this variable affects the other elements of $\mathbf{x}$ such that ϵ_f maintains a strong dependence on it. For instance, a high- p_T b -jet might contain a b -hadron decaying after traversing several layers of the ID, producing tracks reconstructed with diminished resolution, thus reducing the IP's significance, which is a member of $\mathbf{x}$. In this sense, the tagger's output D_x can be seen as function of $\mathbf{x}$ which depends on θ , and the efficiency function ϵ_f as the conditional probability $\epsilon_f(\mathbf{x}|\theta)$ of tagging a jet with features $\mathbf{x}$ given the parameters θ . The truth jet flavour-tagging technique is based on the assumption that $\epsilon_f(\mathbf{x}|\theta)$ can be parametrized using a set of variables θ , called the *relevant* variables, which collectively account for all the factors influencing the probabilities of $\mathbf{x}$. With such parametrization, the conditional probability $p(\mathbf{x}|\theta)$ of $\mathbf{x}$ at a given value of θ is the same between different samples or different regions of phase space, since all the relevant information that affects the probability of $\mathbf{x}$ is captured by θ .

5.1.1 Efficiency parametrization with (p_T, η) maps

An implementation of truth jet flavour-tagging developed in the ATLAS Collaboration and applied in [95, 96], two of the analyses which are improved and combined together in the analysis detailed in Chap. 6, used binned (p_T, η) maps such as the one displayed in Fig. 5.1 to parameterize ϵ_f . The map, like many of the results described in this chapter, is obtained using track-jets associated to large- R jets with $p_T > 250$ GeV reconstructed in simulated $t\bar{t}$ collision events². It displays $\epsilon_b(\mathbf{x}|\theta = (p_T, \eta))$ for the 77% WP of the DL1r algorithm.

Despite their conceptual simplicity, maps have inherent limitations that can hinder their accuracy in parametrizing ϵ_f . First, scaling maps to higher dimensions is an extremely challenging, if not impossible, task. As the number n of dimensions increases, the statistics needed to populate bins in the map grows exponentially with n^3 . Consequently, any dependencies on additional parameters must be disregarded. Secondly, maps of jets provide information about the characteristics of individual jets in the events but do not capture the relations between them or the features of the events. For example, light-flavour jets near b -jets are more likely to be b -tagged because tracks produced by charged particles from a b -hadron decay can be incorrectly attributed to the nearby light-flavour jet. Lastly,

²When applying flavour-tagging only on track-jets associated to a large- R jet, it is safe to assume that ϵ_f depends only on the objects within the large- R jet. Hence, the information needed to describe the event from the truth jet flavour-tagging perspective relates to the track jets associated to the large- R jet and to global event variables such as the average PU.

³An analogous effect was discussed in Sec. 3.2.1 when comparing the RNNIP and IPxD algorithms.

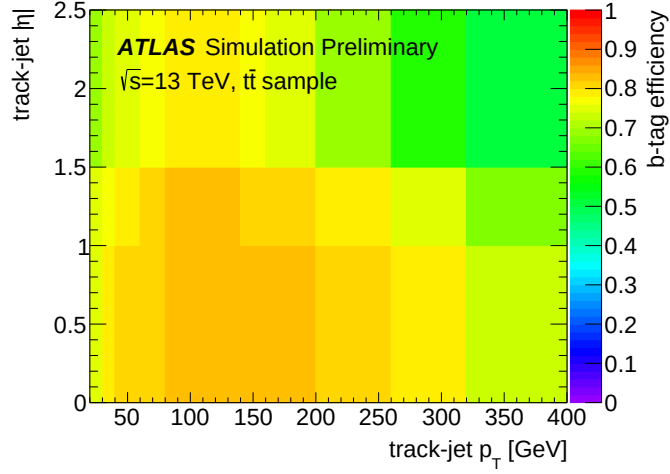


Figure 5.1: $\epsilon_b(\mathbf{x}|\theta = (p_T, \eta))$ efficiency map derived from b -jets in simulated $t\bar{t}$ events using the DL1r jet flavour-tagging algorithm at the 77% WP [94].

optimizing the binning is a non-trivial task due to the necessity to accurately capture all significant variations in the underlying distributions while ensuring a sufficient number of events in each bin.

These limitations led to unsatisfactory performance of the map-based truth jet flavour-tagging method in modelling the distributions of the important V +jets background sample when it was applied in the search for the $(W/Z)H, H \rightarrow c\bar{c}$ process described in [95]. There, differences of up to 20% were observed between the results of pass-or-fail tagging and truth jet flavour-tagging. As a remedy, an event-based correction was devised to take into account the ΔR between the tagged jet and its nearest neighbouring jet. This correction introduced an additional source of systematic uncertainty, ranking seventh among those causing the highest impact on the final result. Nevertheless, using truth jet flavour-tagging rather than pass-or-fail jet flavour-tagging improved the expected limit on the signal strength $\mu_{VH(c\bar{c})}$ of the $(W/Z)H, H \rightarrow c\bar{c}$ process by about 10% due to the improved statistical precision of the MC predictions.

5.1.2 Efficiency parametrization with GNNs

GNNs operating on representations of the events as fully connected graphs of jets can be trained to parameterize ϵ_f without being subject to the limitations that affect maps. $\epsilon_f(\mathbf{x}|\theta)$ depends on various jet and event-level variables, it requires parametrization for events that involve varying numbers of jets, and it was shown to be sensitive to jet-to-jet interactions. In general, NNs are particularly suitable to address high-dimensional problems. In addition, the architecture of GNNs allows to cope with unordered input data with variable size

Track-jet variables
Track-jet p_T
Track-jet η
Track-jet ϕ
Track-jet flavour-label
Mass of the p_T -leading b or c -hadron in the track-jet
p_T of the p_T -leading b or c -hadron in the track-jet
η of the p_T -leading b or c -hadron in the track-jet
ϕ of the p_T -leading b or c -hadron in the track-jet
Event variables
Average number of interactions per event, $\langle \mu \rangle$
Edge variables
Angular separation between two track-jets, ΔR

Table 5.1: Input features to the GNN used to parametrize ϵ_f .

optimally; and it encodes a representation of the data based on pair-wise relationships.

A method for parametrizing jet flavour-tagging efficiencies using GNNs was originally proposed in [93]. It was further developed within the ATLAS Collaboration in [94], a work in which I actively collaborated, for application to $t\bar{t}$ events where a large-R jet can be reconstructed. This method was then expanded to be applicable in both merged and resolved topologies (*i.e.*, topologies involving large-R jets containing variable-radius track jets and topologies involving small-R jets), for both b and c -tagging in the context of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ analysis, as detailed in Chap. 6. In this chapter, the results published in [94] are discussed. The implementation in the context of measuring the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ process is detailed in Sec. 6.11, as the discussion benefits from a prior introduction to the analysis.

Architecture and Training Each event is represented as a fully connected graph, with track-jets associated to the large-R jet serving as the nodes. In input, jets are represented as vectors that combine the track-jet and event variables outlined in Tab. 5.1. Differently from GN1 and GN2 (Sec. 3.4), edges are also given an initial representation, namely the angular distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between the connected jets.

Subsequently, a GNN is employed to generate new vector representations of the jets, incorporating information about the relationships between the jets and between each jet and the global event variables (Fig. 5.2). Each node representation is concatenated to one out of

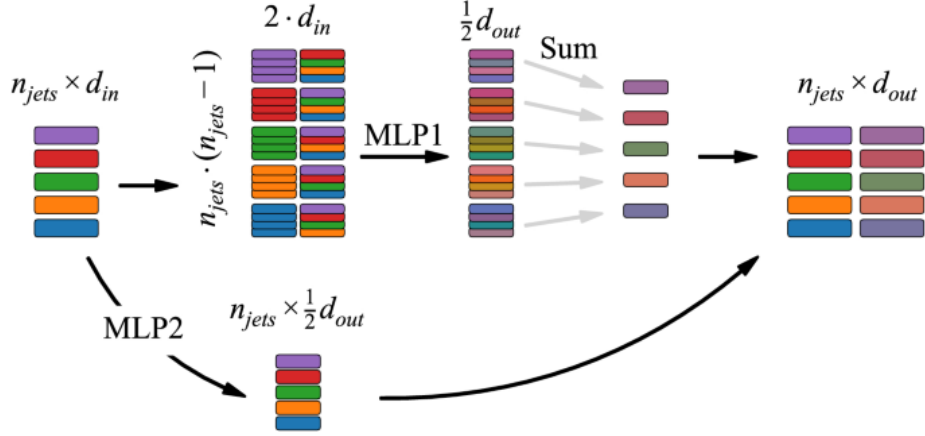


Figure 5.2: Architecture of a layer of the GNN used to parameterize ϵ_f [93].

four 3-dimensional vectors with learnable continuous values, derived from the true flavour of the jet through an embedding layer⁴. The resulting vector representation is indicated as h_i^0 , with $i = 1, \dots, N_{\text{jets}}$, and where the apex indexes the GNN layer. Edge representations e_{ij} are updated for every jet pair as follows. In the case of the first layer of the GNN, h_i^0 is concatenated with h_j^0 and with the ΔR separating the two jets. The resulting vector is processed through a NN, MLP1, which produces a higher dimensional representation of the edge:

$$e_{ij}^0 = \text{MLP1}(h_i^0 \oplus h_j^0 \oplus \Delta R). \quad (5.2)$$

In the case of the GNN layers following the first, the representation of the nodes i and j are concatenated to their original representations h_i^0 and h_j^0 , in addition to their ΔR :

$$e_{ij}^\lambda = \text{MLP1}(h_i^\lambda \oplus h_j^\lambda \oplus \Delta R \oplus h_i^0 \oplus h_j^0). \quad (5.3)$$

MLP1 consists of two linear layers, the first followed by a rectified linear unit, and the second by a hyperbolic tangent unit. Half of the updated node representation is obtained by summing along j the edge scores e_{ij} , and processing the resulting vector with a NN with the same structure as MLP1. The other half is obtained by passing the node representation directly through MLP2, which again is structured as MLP1. The two halves are concatenated into the updated node representation, which is normalized before being fed to the following layer.

After five GNN layers, each node is passed through a NN, the Efficiency Net, stacking linear layers and rectified linear units (Fig. 5.3). The Efficiency Net provides five scalar outputs

⁴Mapping discrete labels into continuous vectors, a procedure called *embedding* in the transformer’s nomenclature, provides the model with the flexibility necessary to implement a notion of “distance” between the classes labelled differently. For instance, in a natural language processing model, it can be useful to allow the network to label the word *have* with a value close to the word *possess* and far to the word *eudaemonia*, to encode the similarity of meaning between the first two, and the fact that they differ a lot from the third.

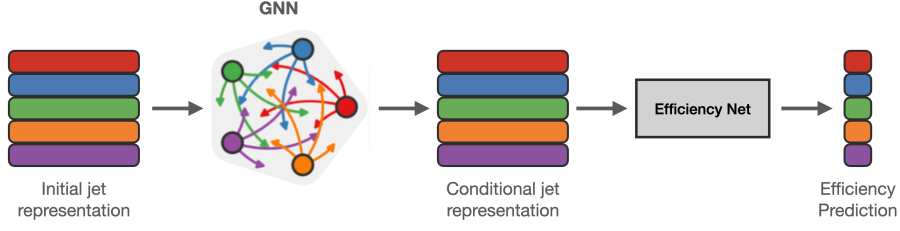


Figure 5.3: The network architecture used for truth jet flavour-tagging. The initial jet representation is created with the jet and the event features listed in Tab. 5.1. This representation is then updated by a fully connected graph network resulting in a jet representation which is also conditioned by its neighbours. A second fully-connected network is then applied on each of these updated jet representations independently to obtain the estimates of the jet flavour-tagging efficiencies [93].

for each jet, representing its probability of falling within each of the five bins of the DL1r D_b score which are maintained by the flavour-tagging group to implement the so-called *Pseudo Continuous B-Tagging* (PCBT). The b -tagging WPs are defined by selecting jets above a given threshold on the discriminant variable D_b built from the continuous outputs of the flavour-tagging algorithms. In PCBT, such discriminant variable is binned via the thresholds defining the 60%, 70%, 77% and 85% b -tagging efficiency WPs. Jets in upper⁵ bins are tagged with increasing likelihood and are used for tighter WPs. Jets belonging to the lowest bin (*i.e.* the bin whose upper edge is the 85% WP threshold) are defined as not tagged. The efficiency of a PCBT bin can be converted into jet flavour-tagging efficiency ϵ_f by summing over the bins above it. The training targets the DL1r PCBT bin in which the jet falls with a multi-class cross entropy loss function. The training converges to an estimate of the true b -tagging efficiency in each of the 5 PCBT bins by virtue of the density ratio estimation method introduced in [97]. By providing a continuous output directly, the GNN eliminates the need for bin optimization studies, which are unavoidable when performing the parametrization using maps.

The model was trained using 15 million events, with 1 million events allocated for validation, and it was evaluated on an additional 15 million events. The training sample was pre-processed to shift and scale the distributions of the variables listed in Tab. 5.1 to be centered in zero with unitary standard deviation. The training process employed the Adam optimizer with an initial learning rate of 10^{-3} and a batch size of 5000 events. The training spanned 30 epochs, as the validation loss, assessed after each epoch, typically reached a saturation point by then. Following the training phase, the best-performing model, as determined by the validation loss, was selected for further performance evaluations.

⁵In PCBT, the “up” and “down” directions are defined as increasing and decreasing values of the D_b scores defining the thresholds.

5.1.3 Event weighting

The maps and the GNN described in the previous sections serve the purpose of providing the probability $\epsilon_f(\mathbf{x}|\theta)$ of tagging a single jet of flavour f given the jet and event-level conditions θ . As anticipated in Sec. 5, though, the truth jet flavour-tagging technique is particularly beneficial for analyses in which multiple jets are required to be flavour-tagged. To apply the method to such cases, it is necessary to combine the ϵ_f^i of all the N_{jets} jets considered in the final state of the analysis to correctly derive the probability of selecting the events based on the jet flavour-tagging requirements.

For instance, the events used in the next section to validate the performance of the GNN were selected replicating the baseline selection criteria of analyses targeting a Higgs boson, or other resonances, decaying hadronically into two b -jets. In the high p_T regime, the decay products are typically reconstructed within a single large- R jet. To be considered, events must contain at least one large- R jet with a p_T greater than 250 GeV, and the two *leading* track-jets, *i.e.* the two track-jets with the highest p_T , must be tagged using the DL1r algorithm at the 77% WP. In cases in which more than one is clustered, only the leading large- R jet in the event is taken into account. For constructing the maps and training the GNN, all the track-jets that are ghost-associated with the leading large- R jet are considered. However, when it comes to validating the approaches and presenting the results, only the two leading ghost-associated track-jets are considered. As far as the flavour-tagging cuts are concerned, then, the probability ϵ_{event} of selecting an event is simply the product of the probabilities of selecting the two leading track-jets associated to the large- R jet: $\epsilon_{\text{event}} = \epsilon_f^1 \times \epsilon_f^2$.

A more general example would be the case in which exactly n_{tag} jets out of the N_{jets} in an event are required to be tagged. This requires a combinatorial approach, since all the $C = \binom{m}{n}$ possible combinations of n_{tag} and $m = N_{\text{jets}} - n_{\text{tag}}$ tagged and non-tagged jets contribute to the event weight with a term

$$\epsilon_c = \prod_i^{n_{\text{tag}}} \epsilon_f^i \prod_j^m (1 - \epsilon_f^j), \quad (5.4)$$

where c indexes the specific combination, and i and j run over the associated pools of tagged and untagged jets. The full event weight is then obtained by summing over all these terms:

$$\epsilon_{\text{event}} = \sum_c^C \epsilon_c. \quad (5.5)$$

Formulas for calculating the truth jet flavour-tagging event weights can be derived under various conditions, such as when n and m are not fixed or when multiple WPs are simultaneously applied, using similar reasoning.

In physics analyses, a precise definition of the b -tagged jets among the selected ones is often required. This is needed, for example, when reconstructing specific event information

in addition to meeting flavour-tagging criteria, such as calculating the mass m_{bb} of the tagged di-jet system in a three-jet event. In such cases, a specific combination is randomly chosen from all possible combinations, with its probability P_c being directly related to its truth-flavour tagging weight ϵ_c via:

$$P_c = \frac{\epsilon_c}{\epsilon_{\text{event}}}. \quad (5.6)$$

5.2 Performance

To be applicable in physics analyses, the distributions obtained through truth jet flavour-tagging must be consistent with those obtained through pass-or-fail flavour-tagging within statistical uncertainty. Additionally, when the condition on the variance of the weights defined in Eq. 5.1 is met, these distributions should display reduced statistical uncertainty. To validate the method, a comparison is performed of the distributions of physical observables from two track-jets, reconstructed from simulated $t\bar{t}$ events selected using the criteria outlined in Sec. 5.1.3. Efficiency maps are constructed using the events that make up the GNN training sample. Subsequently, distributions are generated from another set of events by employing both pass-or-fail flavour-tagging of the jets and the truth jet-flavour tagging method. The resulting distributions are then superimposed, and the ratios to the pass-or-fail flavour tagging distribution are examined in each scenario.

Fig. 5.4 shows the distributions of the large-R jet mass in events with 2 jets b -tagged with the 77% WP of the DL1r algorithm. The distributions are categorized based on the truth-flavour labels assigned to the b -tagged track-jet pairs, and only the flavour combinations that hold statistical significance are showcased.

As observed in all the distributions, the shapes and normalizations are as well, if not better, replicated when extracting truth jet-flavour tagging weights from the GNN-based approach compared to using efficiency maps. Using truth jet-flavour tagging with the efficiency maps tends to overestimate the events selected via pass-or-fail jet flavour-tagging by approximately 10-20%. This overestimation is significantly reduced when employing truth jet-flavour tagging with efficiencies estimated with the GNNs.

In Fig. 5.5, similar distributions as a function of ΔR between the two leading track-jets are showcased. Once more, the GNN approach surpasses the map-based method in terms of replicating the pass-or-fail flavour-tagged distribution. This contrast is especially noticeable in Fig. 5.5a and becomes more pronounced as ΔR decreases, as seen in Fig. 5.5a and 5.5c.

The capability of the GNN-based truth jet flavour-tagging method to correctly model the distribution of the ΔR between the two b -tagged track jets is particularly interesting when seen from the perspective of the works described in [95, 98], analyzing the $(W/Z)H, H \rightarrow b\bar{b}$ and $(W/Z)H, H \rightarrow c\bar{c}$ processes respectively. There, as anticipated in Sec. 5.1.1, the map-based truth jet flavour-tagging method was applied. To account for non-closures found between the ΔR distributions obtained when pass-or-fail flavour-tagging the jets and when

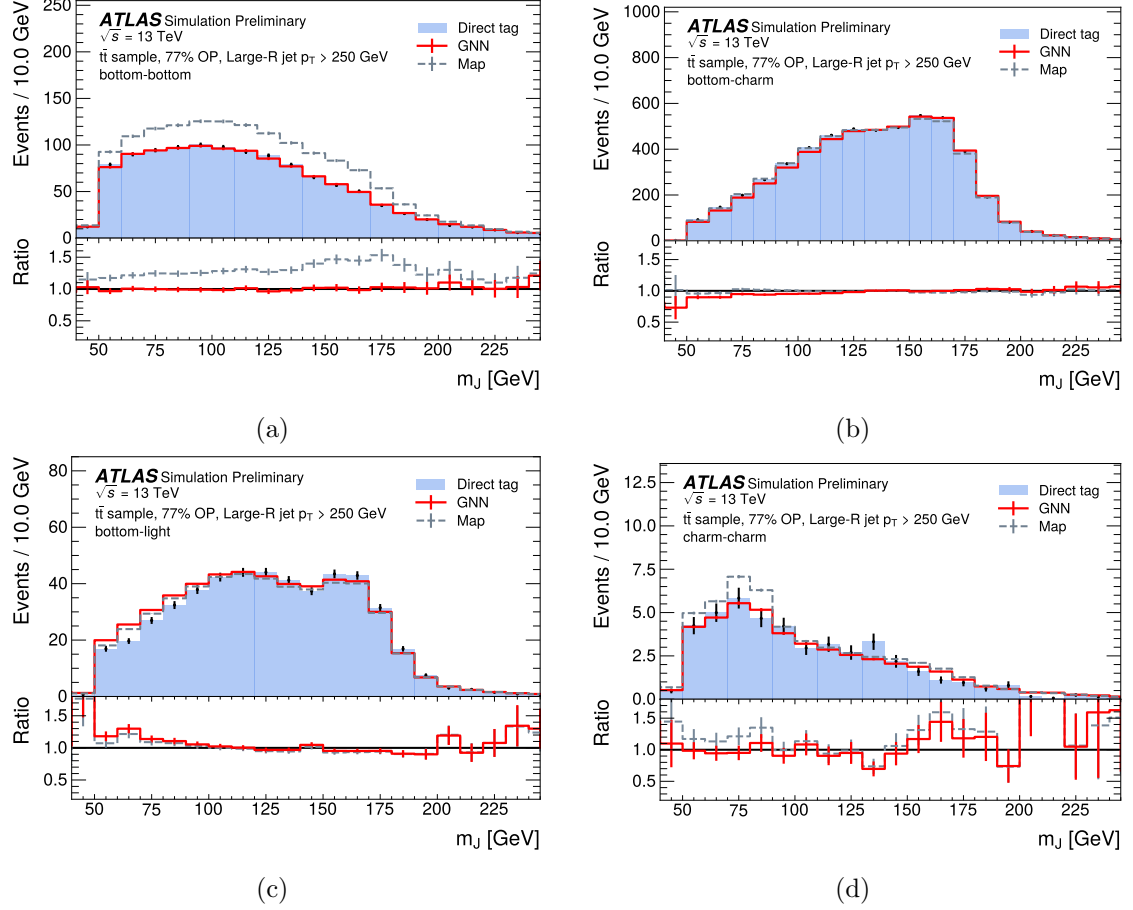


Figure 5.4: Comparison of the large-R jet mass distribution obtained using direct and truth jet flavour-tagging with DL1r at the 77% WP on the first two p_T -leading track-jets associated to the large-R jet in $t\bar{t}$ events. The pass-or-fail flavour tagging is represented by the blue-filled histogram. The black error bands represent the statistical error in pass-or-fail flavour tagging. The truth jet flavour-tagging obtained parametrizing the efficiencies with the map method is in grey, while the GNN method is in red. The bottom pad shows the ratio with respect to the pass-or-fail flavour tagging distribution. Each plot refers to a different truth-flavour combination of the track-jets associated to the large-R jet and used for flavour tagging. From [94].

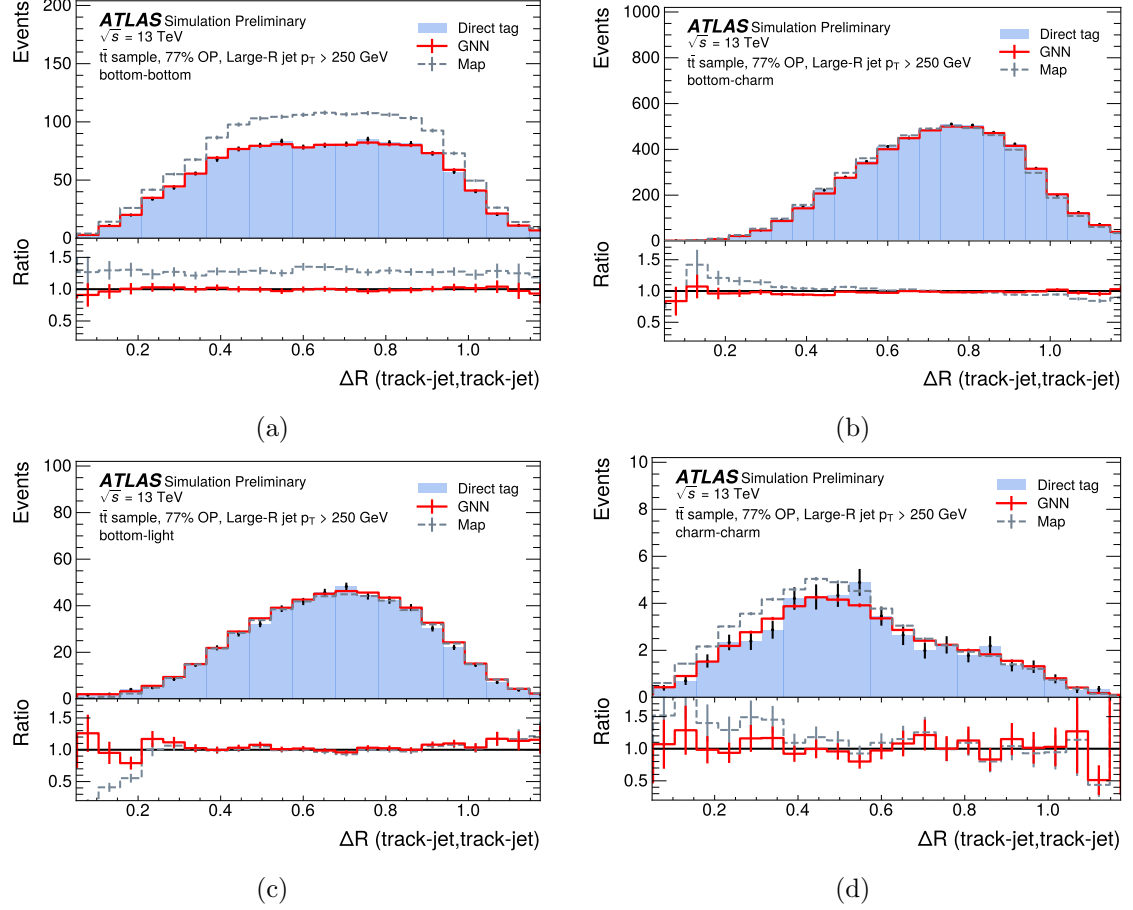


Figure 5.5: Comparison of the ΔR distributions obtained using pass-or-fail and truth jet flavour-tagging with DL1r at the 77% WP on the first two p_T -leading track-jets associated to the large-R jet in $t\bar{t}$ events. The pass-or-fail flavour tagging is represented by the blue-filled histogram. The black error bands represent the statistical error in pass-or-fail flavour tagging. The truth jet flavour-tagging obtained parametrizing the efficiencies with the map method is in grey, while the GNN method is in red. The bottom pad shows the ratio with respect to the pass-or-fail flavour tagging distribution. Each plot refers to a different truth-flavour combination of the track-jets associated to the large-R jet and used for flavour tagging. From [94].

applying the truth jet flavour-tagging method, a dedicated ΔR reweighting was derived and used to correct the non closure. Given the uncertain nature of this correction and the limited statistics available for its derivation, a substantial systematic uncertainty, equal to half of the correction, had to be attributed to the relevant MC templates; in a context in which the overall uncertainty related to the statistics of the MC templates contributed already up to approximately 20% to the total background uncertainty [93].

The good modeling of the ΔR distribution achieved through the GNN-based truth-flavour tagging method eliminates the need to introduce a similar correction in the analysis described in Chap. 6.

Chapter 6

Measurements of $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays in the VH production channel

This chapter presents a novel measurement of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ processes performed for a wide range of transverse momenta of the Higgs boson.

After motivating its importance, the elements of the analysis are described in-depth: Sec. 6.3 describes the employed proton-proton collision dataset and MC simulated samples, Sec. 6.4 the selection of the relevant physics objects, Sec. 6.5 the online and offline event selection and their categorization in multiple Signal and Control Regions (SRs and CRs), Sec. 6.6 the techniques used in the analysis to enhance the resolution of the Higgs-boson candidate mass, Sec. 6.7 the BDT trained to enhance the separation of signal and background events, Sec. 6.8 the systematic uncertainties affecting the measurement, Sec. 6.9 the modelling of the $V + \text{jets}$ background events, Sec. 6.10 the profile binned maximum likelihood fit used to extract the final results, Sec. 6.11 the implementation of the GNN-based truth flavour-tagging technique introduced in Chap. 5, which marks its inaugural application, and Sec. 6.12 the expected results.

Particular focus is placed on the elements of the analysis benefiting from my contribution, namely, the implementation of the GNN-based truth flavour-tagging technique, introduced in Chap. 5 and implemented within this analysis as detailed in Sec. 6.11; and the modelling of the dominant $V + \text{jets}$ background in the $H \rightarrow b\bar{b}$ decay mode, described in Sec. 6.9.

The most noticeable expected results, discussed in Sec. 6.12, include:

- The most precise determination of the signal strength $\mu_{VH, H \rightarrow b\bar{b}}$ of the $VH, H \rightarrow b\bar{b}$ process to date. Namely, $\mu_{VH, H \rightarrow b\bar{b}}^{\text{exp}} = 1.00 \pm 0.13$, corresponding to an expected significance of the $VH, H \rightarrow b\bar{b}$ signal over the background-only hypothesis of 8σ .

- The most stringent 95% CL upper limit on the signal strength of the $VH, H \rightarrow c\bar{c}$ process set by the ATLAS Collaboration, namely 11^{+4}_{-3} .
- The first observation of the $WH, H \rightarrow b\bar{b}$ process, with an expected significance of 5.2σ .
- The first observation of the $VZ, Z \rightarrow c\bar{c}$ process¹ by the ATLAS Collaboration, with an expected significance of 5.5σ .

Given that the analysis is not currently public, the results within this thesis and the details about the analysis elements are to be considered preliminary.

Unblinded results, obtained by the team after the defense of this thesis and published in [2], are added for completeness in App. A.

6.1 Motivation

During Run-2, the Yukawa couplings y_f of the Higgs boson with the third-generation fermions were established by the ATLAS and CMS collaborations [60, 99–102]. Among these is the y_b Yukawa coupling. Albeit $H \rightarrow b\bar{b}$ is the most common decay mode of the Higgs boson, predicted by the SM to have a branching ratio of 58% for $m_H = 125$ GeV (Fig. 6.2b), measuring it at the LHC is challenging because of the large background of events with multiple jets produced in QCD interactions between the colliding partons (Fig 6.3), henceforth termed *multijet* events.

The “golden” production mode that allows overcoming this experimental difficulty is the rather rare Higgs-strahlung (Fig. 6.1e and 6.2a), also termed VH production mode, in which a Higgs boson is radiated by a vector W or Z boson. Compared to the most prominent gluon-gluon Fusion (ggF, Fig. 6.1a) and Vector Boson Fusion (VBF, Fig. 6.1b), Higgs-strahlung has the additional advantage that the vector boson produced in association with the Higgs boson can decay into leptons. The presence of these leptons allows to distinguish VH events from the multijet background, as opposed to the fully hadronic final state expected in VBF and ggF, $H \rightarrow b\bar{b}$ events. Additionally, leptons offer clear signatures that can be used as triggers during data-taking.

The VH production mode was utilized for first detecting the $H \rightarrow b\bar{b}$ process [60, 102]. After the discovery, the focus shifted towards precision measurements, to characterize the cross-section of $VH, H \rightarrow b\bar{b}$ events in different phase-space regions. The ATLAS Collaboration performed three additional measurements targeting the $VH, H \rightarrow b\bar{b}$ process using the full Run-2 dataset. The first of these analyses [96] used the same strategy exploited for the first $H \rightarrow b\bar{b}$ observation to reconstruct the Higgs boson candidates, *i.e.* using the two separate

¹The $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ processes are measured in addition to $VH, H \rightarrow b\bar{b}/c\bar{c}$ to validate the analysis strategy (Sec. 6.12.2).

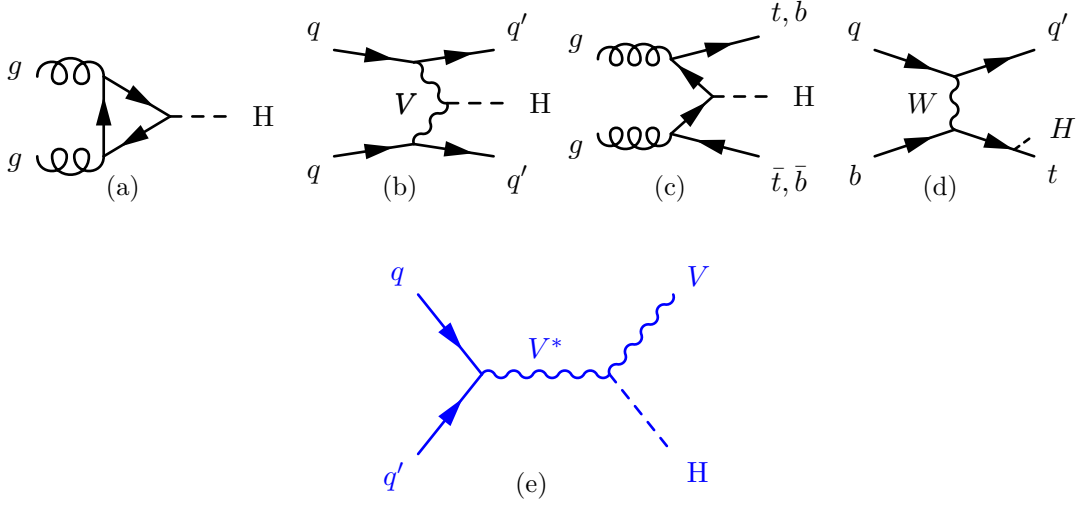
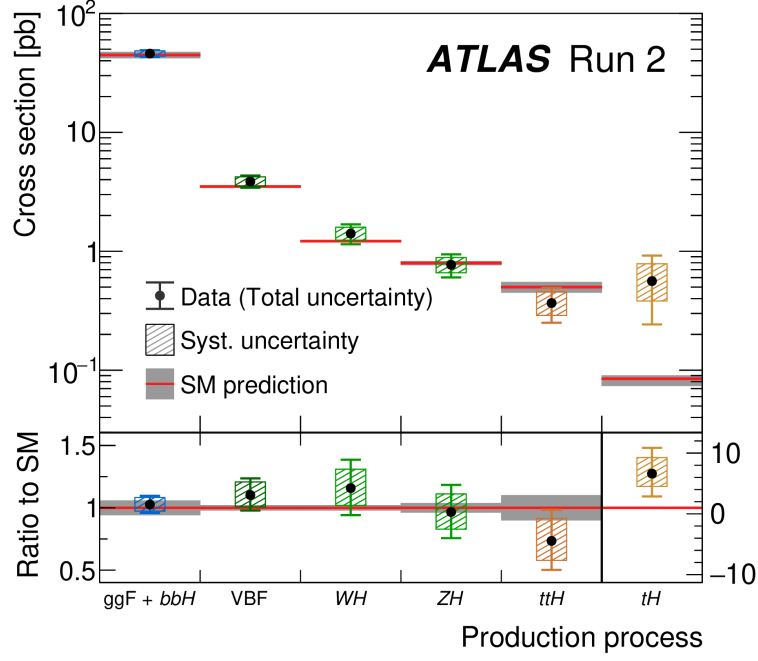


Figure 6.1: Feynman diagrams for the main Higgs production mechanisms at the LHC. (a) Gluon-gluon fusion (ggF), (b) Vector Boson Fusion (VBF), (c) ttH/bbH associated production, (d) tH associated production and (d) Higgs-strahlung. Higgs-strahlung is emphasized as it is the production mode targeted in this chapter.

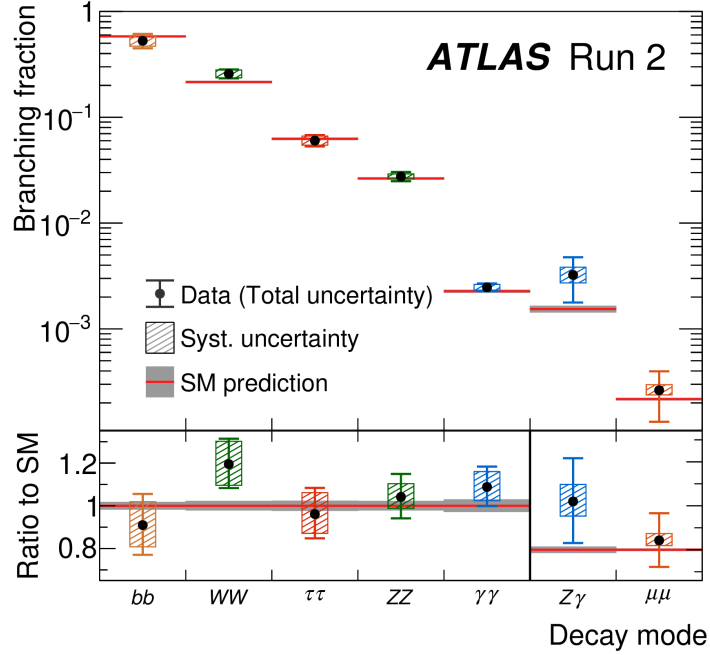
jets stemming from the hadronization of the b -quarks pair as a proxy for the Higgs particle. Since the analysis was performed in the p_T^H regime where the two jets can be reconstructed separately, it is addressed as *resolved*. The majority of Higgs bosons in $VH, H \rightarrow b\bar{b}$ events are generated within this regime in the ATLAS experiment [96]. The second analysis [105] was optimized to target $VH, H \rightarrow b\bar{b}$ events in which the Higgs boson is produced with high transverse momentum. Given that strongly boosted Higgs bosons decay in b -quark pairs whose hadronization jets cannot be separated, but must be reconstructed within single, larger radius jets, this analysis is commonly termed *boosted*, and the jet topology that it addresses is termed *merged*. Finally, the resolved and boosted analyses were combined [106] by treating the resolved and boosted cases as two orthogonal kinematic regions.

Since it gives the highest statistical contribution, the resolved region is essential to measure the total rate of the $VH, H \rightarrow b\bar{b}$ process precisely and to perform differential cross-section measurements. Even if it accounts for only 1‰ of the total VH production cross-section, the boosted region, instead, is particularly interesting for its sensitivity to heavy new physics, including two-Higgs doublets models [107, 108], spontaneous flavour violating Higgs sectors [109], additional heavy-quarks and others [110]: at high energies, additional interactions with new heavy particles could result in detectable deviations from the SM predictions. In this energy regime, Higgs-strahlung becomes increasingly prominent in the overall Higgs-boson production, eventually emerging as the dominant mode for $p_T^H > 1.2 \text{ TeV}$ (Fig. 6.4).

While establishing and characterizing the coupling of the Higgs boson to third-generation fermions, efforts have been made to explore Higgs boson decay processes involving fermions



(a)



(b)

Figure 6.2: Observed and predicted Higgs boson production cross-sections and branching fractions. (a) Cross-sections for different Higgs boson production processes are measured assuming SM values for the decay branching fractions. Albeit WH and ZH are not the most frequent production modes, leptons emerging from the decay of the W or the Z boson give a handle for measuring $H \rightarrow b\bar{b}$ at the LHC. (b) Branching fractions for different Higgs boson decay modes are measured assuming SM values for the production cross-sections. The lower panels show the ratios of the measured values to their SM predictions. The vertical bar on each point denotes the 68% confidence interval. The most frequent decay channel of the Higgs boson is $H \rightarrow b\bar{b}$. From [103].

Standard Model Production Cross Section Measurements

Status: October 2023

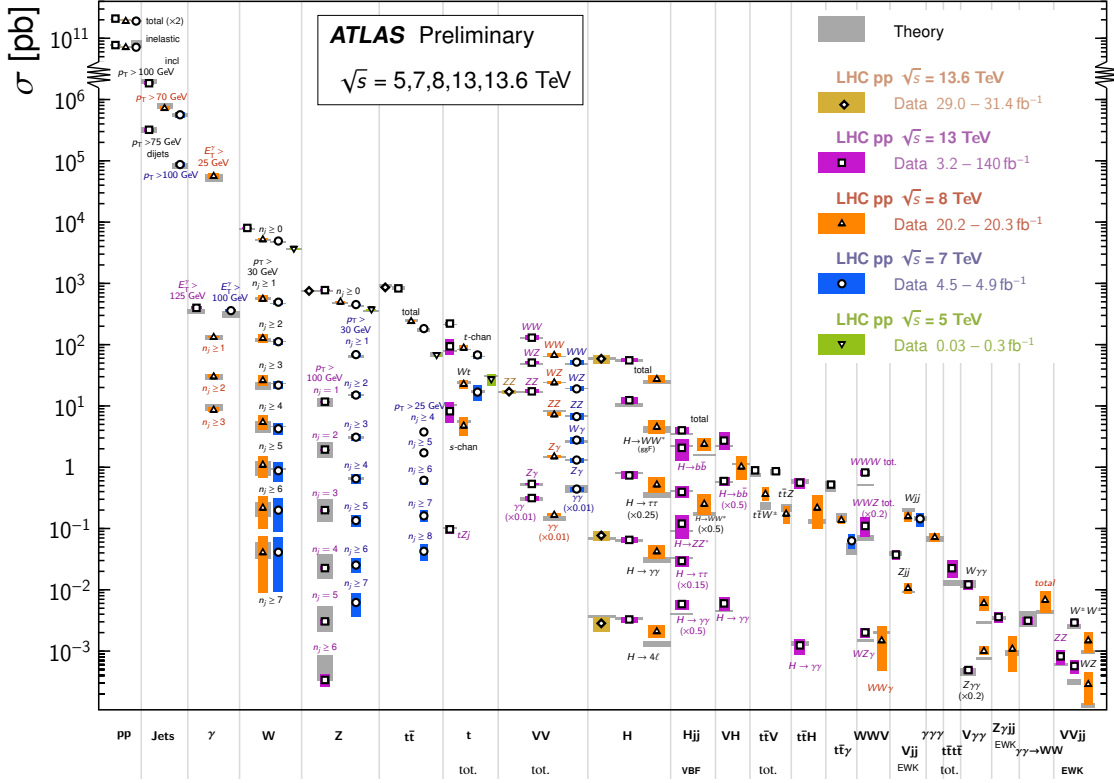


Figure 6.3: Summary of several SM total and fiducial production cross-section measurements. The measurements are corrected for branching fractions, compared to the corresponding theoretical expectations. In some cases, the fiducial selection differs between measurements in the same final state for different centre-of-mass energies $\sqrt{s}$, resulting in lower cross-section values at higher $\sqrt{s}$. The cross-sections of multijet events, indicated in the second column, exceed those of events containing Higgs bosons of various orders of magnitude. From [104].

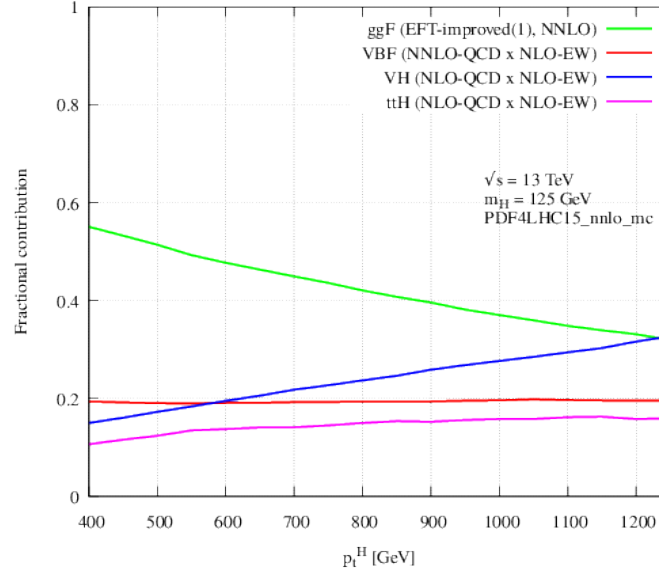


Figure 6.4: Relative share of the cross-section for producing a Higgs boson as a function of the lowest cut on the Higgs boson transverse momentum. From [111].

from the first and second generations. These are probes of the corresponding Yukawa couplings, holding particular interests due to their smallness, making them susceptible to significant modifications in the new physics scenarios. Couplings of the Higgs boson to first and second-generation particles still have not been observed as of today. However, the CMS Collaboration has reported the first 3σ evidence of Higgs boson interactions with second-generation fermions from studying $H \rightarrow \mu^- \mu^+$ [112], a decay channel in which ATLAS observed a 2σ excess over the background only hypothesis [113].

Among the decay modes of the Higgs boson yet unobserved, the SM predicts $H \rightarrow c\bar{c}$ to be the most frequent, with a branching ratio of 2.89% [114]. In [95], the most recent direct search for $H \rightarrow c\bar{c}$ decays in the VH production mode performed in ATLAS is detailed. The analysis shares many features with the resolved $VH, H \rightarrow b\bar{b}$ measurement owing to the similarity of the final states' topologies, leveraging leptonic decays of the vector bosons and the presence of flavour-tagged jets. Using the full Run-2 dataset, an upper limit was set on the rate of the $VH, H \rightarrow c\bar{c}$ process of 26 times the value predicted by the SM, with a 95% CL. More recently, the CMS collaboration pushed such limit down to observed(expected) 14.4(7.6) times the SM, a result made possible by applying sophisticated $H \rightarrow c\bar{c}$ tagging techniques based on deep-learning and targeting the boosted region [115].

Owing to their similarities, the $VH, H \rightarrow c\bar{c}$ and the resolved $VH, H \rightarrow b\bar{b}$ analyses were combined in [95]. This allowed to simultaneously measure y_b and constrain the Higgs-charm Yukawa coupling y_c , and constrain the ratio y_c/y_b , hence removing all the assumptions on

the total Higgs boson width in the interpretation. The measurements of $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ events are also challenged by the same main background events, namely V +jets and $t\bar{t}$: combining the analyses can increase the precision with which the contamination from these backgrounds is determined.

6.1.1 Improving the $(W/Z)H, H \rightarrow b\bar{b}$ and $(W/Z)H, H \rightarrow c\bar{c}$ measurements in ATLAS

The two combinations hitherto described, namely that involving the resolved and boosted $VH, H \rightarrow b\bar{b}$ analyses, and that involving the $VH, H \rightarrow c\bar{c}$ and resolved $VH, H \rightarrow b\bar{b}$ analyses, were performed while preserving the distinct strategies of each analysis under consideration. In contrast, the analysis presented in this thesis specifically targets resolved and boosted $VH, H \rightarrow b\bar{b}$ events, along with $VH, H \rightarrow c\bar{c}$ events, establishing a *coherent* strategy optimized for both $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decay modes. The goal is to correlate modelling and experimental uncertainties as comprehensively as possible, maximizing the statistical power of the fit. This surpasses previous analyses and combinations not only due to a harmonized design but also through the use of improved physics objects and event reconstruction techniques, and innovative machine learning methods.

Most noticeably, the DL1r flavour-tagging algorithm is applied to identify both resolved and merged b -jets and c -jets, in contrast to the previous $VH, H \rightarrow b\bar{b}$ analyses that used its predecessor MV2, and $VH, H \rightarrow c\bar{c}$, that used DL1 and MV2 to identify c -jets and b -jets, respectively. Another major improvement consists in consistently fitting the output distribution of a BDT designed to enhance the separation of signal from background events in the SR, whereas the earlier $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ analyses fitted the resonant mass of the flavour-tagged jets. The BDT improves the separation of signal from backgrounds by factoring in other discriminant variables in addition to the mass of this di-jet system. Additionally, the MC samples used to simulate the important V +jets background events were produced with Next to Leading Order (NLO) precision and substantially enhanced in statistics.

On the front of advanced machine-learning techniques, the GNN-based truth flavour-tagging method described in Chap.5 is employed to increase the statistical precision on the background events containing mistagged jets, whereas the previous $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ analyses adopted a simplified parametrization of the flavour-tagging efficiencies using (p_T, η) maps and the resolved $VH, H \rightarrow b\bar{b}$ analysis applied a pass-or-fail flavour tagging. Using the inferior map-based parametrization forced including additional systematic uncertainties in the previous $VH, H \rightarrow c\bar{c}$ analysis, ranking as high as seventh among those causing the highest impact on the final results. A new reweighting technique, the Calibrated Likelihood Ratio Estimator (CARL), is also introduced [116]. CARL is a NN that maps distributions obtained from the nominal MC generator to those produced by alternative generators. These are used to estimate theoretical systematic uncertainties regarding the shapes of simulated distributions. Pragmatically, alternative samples are made with signifi-

cantly fewer events due to computational resource considerations. Reweighting the nominal distributions to the alternatives allows for exploiting their larger statistics when estimating this source of systematic effects. CARL replaces a previous reweighting technique based on BDTs, formerly applied in the resolved $VH, H \rightarrow b\bar{b}$ analysis, offering a new, more flexible tool.

Analysis deliverables Scope of the analysis is to extract, by performing a binned profile likelihood fit in numerous SRs and CRs:

- The $VH, H \rightarrow b\bar{b}$ signal strength and its corresponding significance over the background only hypothesis;
- The upper limit on the signal strength of the $VH, H \rightarrow c\bar{c}$ process, set with a 95 % CL;
- The exclusive signal strengths of the $ZH, H \rightarrow b\bar{b}$ and $WH, H \rightarrow b\bar{b}$ processes;
- The exclusive signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes, split in p_T^V and jet multiplicity bins for the resolved regime, and only in p_T^V bins for the boosted regime;

Additionally, a measurement of the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ processes is performed. Since $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ processes are much more frequent than $VH, H \rightarrow b\bar{b}/c\bar{c}$ processes but produce very similar final states, measuring them helps validate the analysis strategy by providing SM candles for the $VH, H \rightarrow b\bar{b}/c\bar{c}$ processes.

6.2 Identikit of the signal and the main backgrounds

6.2.1 Reconstructing the Higgs boson

The analysis primarily relies on reconstructing the $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays over a wide range of p_T^H values as accurately as possible. This requires different techniques depending on the Lorentz boost of the Higgs boson, which is proportional to the Higgs boson mass m_H and its transverse momentum p_T^H . To simplify the picture, for a Higgs boson decaying in the transverse plane with momentum $p = p_T^H$, the angular distance ΔR between the two b or c -quarks produced in the decay can be approximated by the formula:

$$\Delta R(q_1, q_2) \approx \frac{m_H}{p_T^H} \frac{1}{\sqrt{z(1-z)}} \approx \frac{2m_H}{p_T^H}, \quad (6.1)$$

where z represents the fraction of momentum of the Higgs boson carried by q_1 . Given that b and c -quarks carry the majority of the energy of the jets they generate, Eq. 6.1 also approximates the angular distance between these jets.

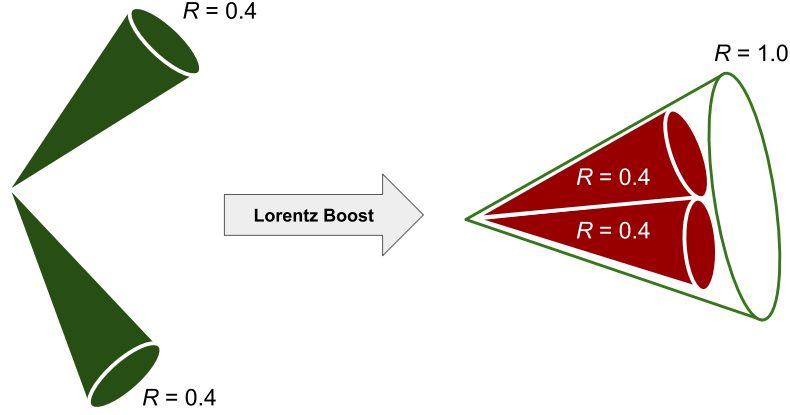


Figure 6.5: Different techniques to reconstruct the $H \rightarrow b\bar{b}$ decays at relatively low and high Higgs boson transverse momentum values.

In the regime with $p_T^H < \sim 300$ GeV, encompassing most signal events, the jets exhibit significant separation, allowing for individual reconstruction. The decay in this scenario is reconstructed using two small- R jets with $R = 0.4$, and the Higgs boson candidate is formed by adding their respective four-momenta.

However, as the transverse momentum surpasses $p_T^H = 400$, bringing the two jets closer than twice their radius $R = 0.4$, the efficiency of this reconstruction technique diminishes. In such instances, a more effective approach is reconstructing both jets within a large- R jet with $R=1.0$ (Fig. 6.5).

In the $300 \text{ GeV} < p_T^H < 400 \text{ GeV}$ range, it is possible to reconstruct the decay of the Higgs boson with both techniques approximately in 50% of the cases. This analysis introduces a $p_T^H = 400 \text{ GeV}$ boundary that defines two orthogonal regions to reconstruct the $H \rightarrow b\bar{b}$ decay using the alternative techniques. The most reliable experimental proxy for the true transverse momentum of the Higgs boson is the transverse momentum of the vector boson against which it recoils, denoted as p_T^V . Given the negligible transverse momentum of the colliding partons at the LHC, p_T^V and p_T^H must cancel out due to the conservation of momentum: in vectors, $\mathbf{p}_T^V = -\mathbf{p}_T^H$. In the cases where V decays leptonically, though, p_T^V is a better proxy than the sum of the transverse momenta of the jets stemming from the Higgs decay due to the better resolution with which the leptons' momenta can be measured in the detector compared to the jets'. For what concerns $H \rightarrow b\bar{b}$ decays, hence, the Higgs candidate is reconstructed via resolved jets in events where $p_T^V < 400 \text{ GeV}$ is measured, and with merged jets otherwise. $H \rightarrow c\bar{c}$ decays, instead, are reconstructed only with the resolved technique. Introducing a boosted $H \rightarrow c\bar{c}$ region was considered, but no significant improvement was found using the current analysis design, due to the low rate of signal

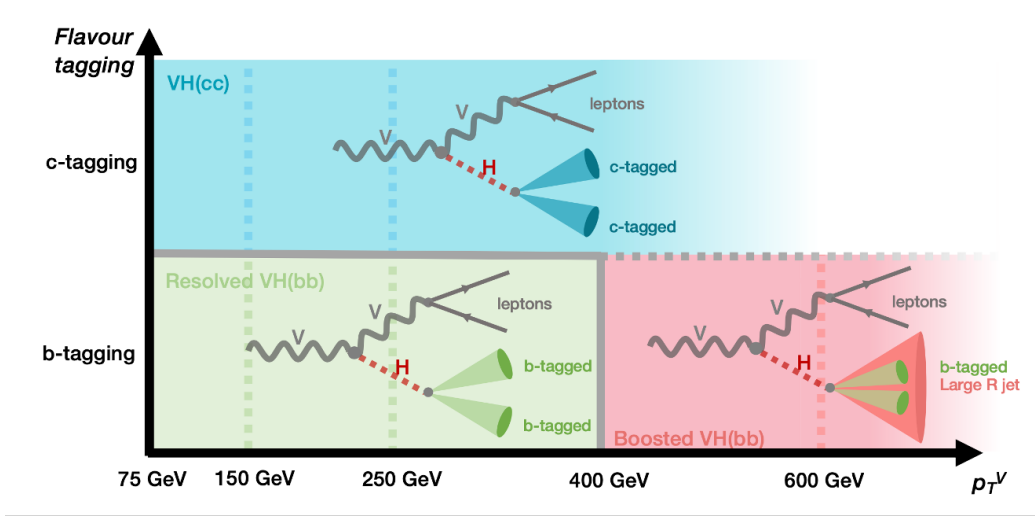


Figure 6.6: Illustration of the Higgs boson reconstruction in the different regions considered in the analysis. There are rare cases where some events enter both the $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ regions, causing a small overlap indicated by the dotted line separating the two regions.

events. New promising GNN-based $H \rightarrow c\bar{c}$ tagging algorithms [117] being calibrated in ATLAS will enable future analysis designs benefiting from the boosted $H \rightarrow c\bar{c}$ region.

Fig. 6.6 schematises the techniques used to reconstruct the Higgs boson candidate in the various regions of the analysis. The requirements on the flavour-tagging scores of the jets used in the reconstruction are described in the next section. There are rare cases where these requirements allow some events to enter both the $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ regions, causing a small overlap indicated in Fig. 6.6 by the dotted line separating the two regions.

Flavour-tagging scheme Among the novel design choices introduced in this analysis to harmonise the measurements of $VH, H \rightarrow c\bar{c}$ and resolved $VH, H \rightarrow b\bar{b}$ events, to benefit as much as possible from their combination, is the scheme adopted to identify the b and c -jets: the so-called Pseudo Continuous Flavour Tagging (PCFT) scheme. In PCFT, the D_b and D_c outputs of the DL1r algorithm are used to define the two axes of a 2D plane (Fig. 6.7a). The PCFT plane is divided into five regions, for b and c -jets identified with increasing stringency. Thus, a perfect tagger, capable of separating completely all the classes of jets, would assign all the b -jets to the rightmost bin, all the c -jets to the top-left one, and all the light-flavour jets to the bottom left one. As it was detailed extensively in Chap. 3 and 4, jet flavour-tagging algorithms are imperfect classifiers, meaning that the distributions of their output scores D_x when they are evaluated on the different classes of jets overlap. As a result, part of the jets fall outside the bins mentioned above, and all the bins contain a

PCFT bin name	ϵ_{jet}			
	b -jets	c -jets	light-jets	τ -jets
Untagged	14.5%	47.4%	92.4%	60.3%
Loose c -tag	11.5%	20.5%	6.5%	18.5%
Tight c -tag	4.8%	24.2%	0.9%	19.5%
70% b -tag	11.2%	5.2%	0.1%	1.7%
60% b -tag	58%	2.6%	0.05%	0.5%

Table 6.1: The jet flavour-tagging efficiencies of the various PCFT bins for b -jets, c -jets, light-flavour jets and τ -jet, when measured on a $t\bar{t}$ sample where at least one W -boson decays leptonically, simulated with the POWHEG+PYTHIA 8 generator.

mixture of all the classes. When evaluated on a simulated $t\bar{t}$ sample, 60 % of the b -jets fall in the rightmost *60% b -tag* bin, the purest for b -tagging. An additional 10% of them fall in the less pure *70% b -tag* bin² 20% of the c -jets fall in the top-left *tight c -tag* bin, and an additional 20% in the *loose c -tag* bin (Tab. 6.1). All the remaining jets, which fall in the *untagged* bin, are considered as light-flavour jets.

Resolved $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decay candidates are selected based on the PCFT scores of the jets in the events, as illustrated in Fig. 6.7b. Decays of Higgs boson candidates in which both the resolved jets are tagged in the 70% or 60% b -tag bins are accepted for the resolved $VH, H \rightarrow b\bar{b}$ regime. Due to the lower efficiency of the c -tagging bins, $H \rightarrow c\bar{c}$ candidates are accepted in all cases where at least one of the two candidate jets falls in the tight c -tag bin. This choice is justified by the rarity of the $H \rightarrow c\bar{c}$ process, which requires to favour signal efficiency over purity.

Boosted $H \rightarrow b\bar{b}$ events are selected by applying the PCBT method (Sec. 5.1.2) on the jets reconstructed within the merged large- R jet. Both jets must be identified with the 85% b -tagging WP, the most efficient maintained by the ATLAS flavour-tagging group. Efficiency is favoured over purity due to the scarcity of events and the smaller contribution of the backgrounds in the boosted region.

The PCFT plane defines orthogonal regions, meaning no overlap exists between the $VH, H \rightarrow c\bar{c}$ and resolved $VH, H \rightarrow b\bar{b}$ regions³. There are events, instead, where both an $H \rightarrow c\bar{c}$ and a merged $H \rightarrow b\bar{b}$ decay candidate can be reconstructed. The impact of such events, which account for 0.4% and 4.3% of the $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ events reconstructed on data, was found to be negligible on the final results in dedicated studies.

²The name 70% b -tag bin indicates its lower edge corresponds to the threshold for the 70% b -tagging WP.

³Orthogonality is also ensured in events in which more than two jets are reconstructed by vetoing any b -tagged jet in $VH, H \rightarrow c\bar{c}$ candidate events.

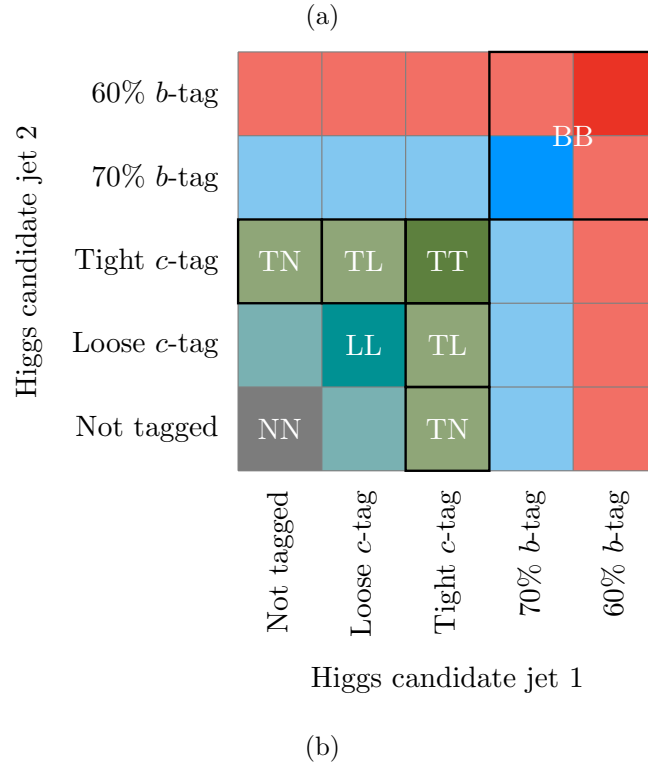
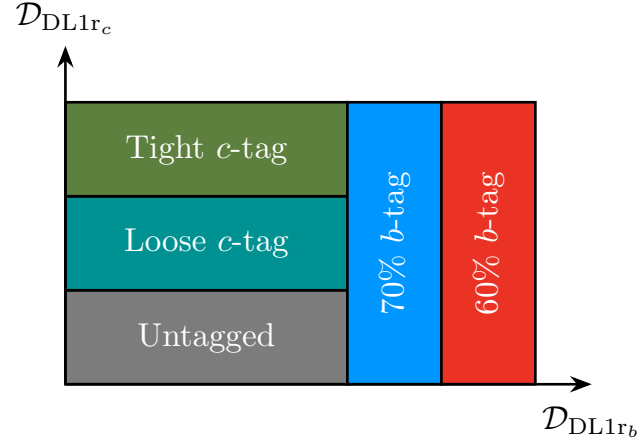


Figure 6.7: (a) PCFT plane used in the analysis to perform b and c -tagging simultaneously. (b) Combinations of PCFT scores used to reconstruct $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays. T stands for tight c -tag, L for loose c -tag, B for b -tagged with the 70% b -tagging WP (and hence is used for jets falling either in the 70% b -tagging or in the 60% b -tagging bins), and N stands for untagged. LL and NN combinations are not used.

6.2.2 Reconstructing the vector boson

As anticipated in previous sections, in order to filter out most of multijet events generated in proton-proton collisions, which would otherwise overshadow the observation of $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ events within this background, this analysis specifically focuses on the leptonic decays of vector bosons. The signatures which are targeted are $Z \rightarrow \nu\bar{\nu}$, $W \rightarrow l^-\bar{\nu}$, and $Z \rightarrow l^-l^+$ where l stands for electrons or muons. τ leptons are not targeted explicitly, but, in the rare cases in which their decay products pass the selections optimised on the V decays mentioned above, they are treated accordingly.

Each of the three V decays yields a distinct detector signature, leading to the division of the analysis into three sub-channels. The *0-lepton* (0L) channel identifies events with missing transverse energy E_T^{miss} from the $Z \rightarrow \nu\bar{\nu}$ decay. The *1-lepton* (1L) channel selects events with precisely one electron or muon and E_T^{miss} ⁴. The *2-lepton* (2L) channel targets events with either two electrons or two muons.

In the 0L channel, only information about the transverse momentum component of the Z boson candidate is available due to the unknown component in the beamline direction for both neutrinos. In the 1L channel, the W boson is fully reconstructed using the quadrimomentum of the measured lepton, E_T^{miss} , and a W mass constraint. Finally, in the 2L channel, the Z boson candidate is constructed from the two quadrimomenta of the reconstructed leptons.

Reconstructed events in the different leptonic channels and in the resolved and boosted regions are showcased in Fig. 6.8, 6.9 and 6.10.

⁴An exception is represented by $W \rightarrow \tau\bar{\nu}_\tau$ processes with hadronically decaying taus. When decaying into hadrons, τ can be erroneously reconstructed as jets. Hence, $WH, W \rightarrow \tau\bar{\nu}_\tau$ events often pass the selections of the 0L channel. For instance, almost 20% of the 0L $VH, H \rightarrow b\bar{b}$ events are WH , with the W decaying to hadronic τ . These τ are identified with an RNN named τ_{had} -ID [118], and the events containing them are moved to the 1L channel.

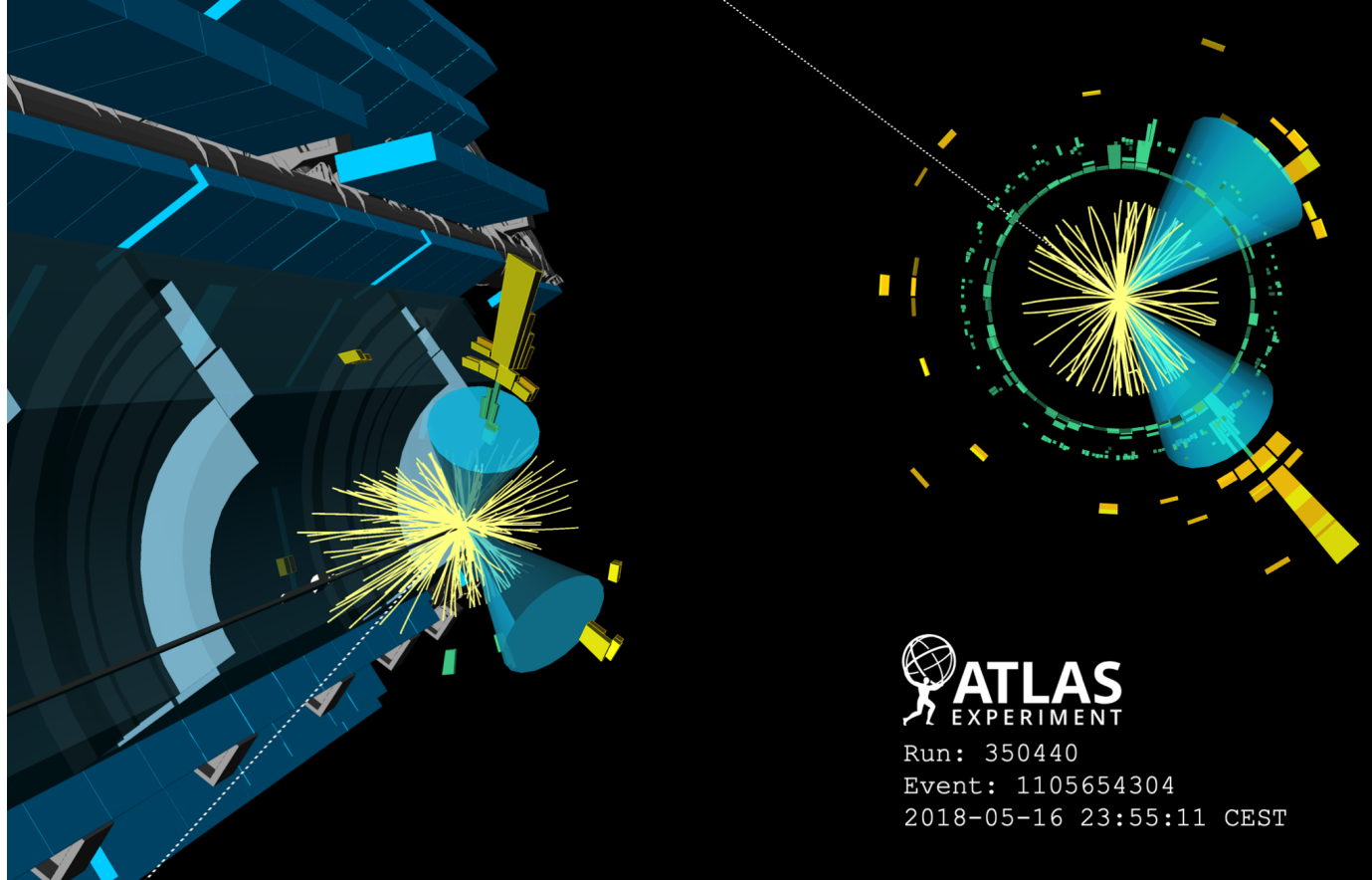


Figure 6.8: Candidate event for the process $ZH \rightarrow \nu\bar{\nu}c\bar{c}$. Missing transverse energy is identified by a white dashed line and it has a magnitude of 155 GeV. The Higgs boson candidate is reconstructed from two c -tagged $R = 0.4$ jets (blue cones), which have associated energy deposits in both the electromagnetic (green) and hadronic (yellow) calorimeters. The leading and sub-leading jets have transverse momenta of 176 GeV and 22 GeV respectively. The Higgs boson candidate has a reconstructed invariant mass of 125 GeV.

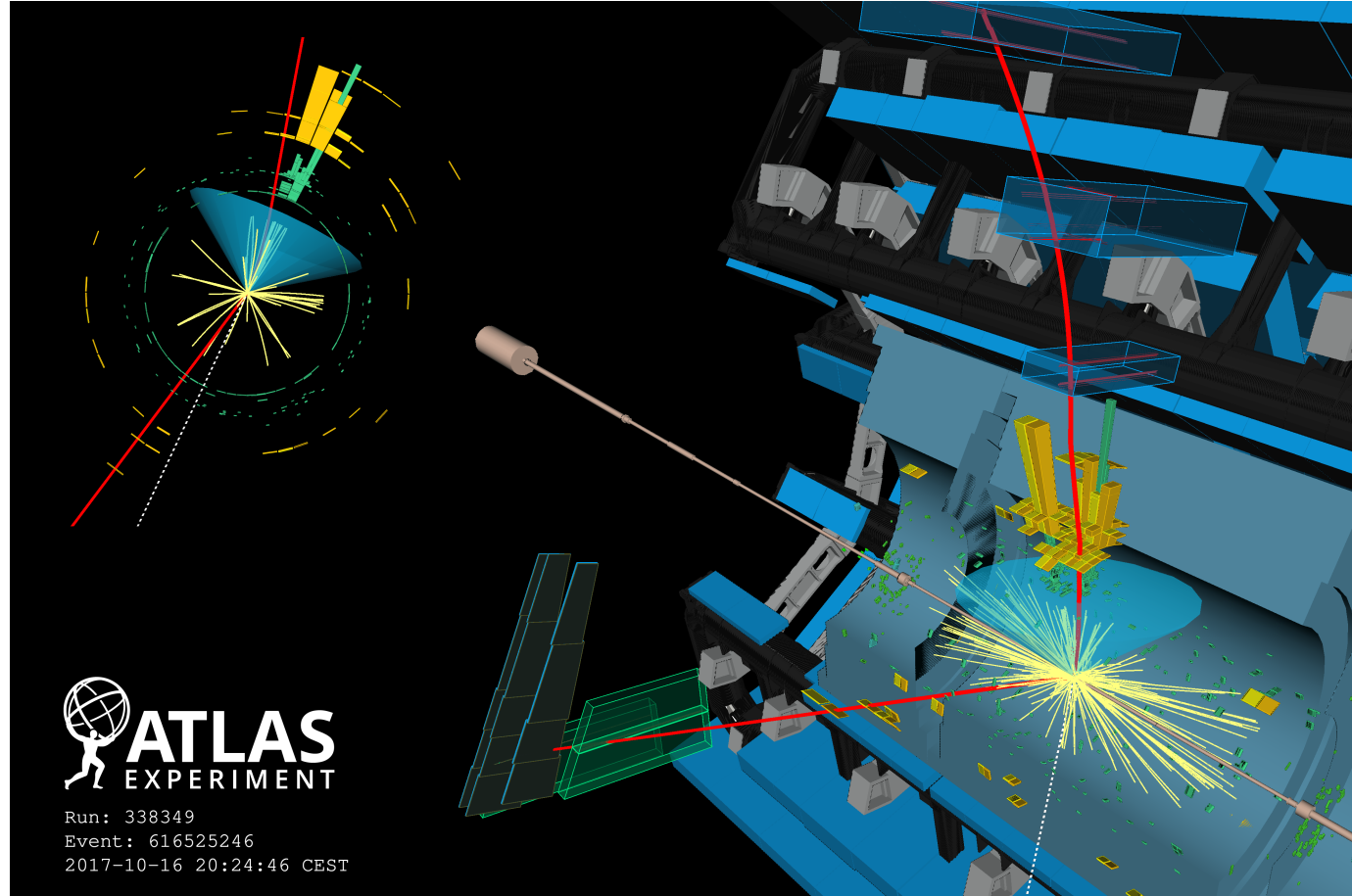


Figure 6.9: Candidate event for the process $WH \rightarrow \mu \bar{\nu} b \bar{b}$. An isolated muon is shown as the lower red track producing hits in the endcap-muon chambers; it has a transverse momentum of 196 GeV and a pseudorapidity of 1.16. The missing transverse momentum is identified by a white dashed line and it has a magnitude of 832 GeV; the transverse momentum (mass) of the muon- E_T^{miss} system is 1.03 TeV (112 GeV). The Higgs-boson candidate is reconstructed as a single large- R ($R = 1.0$) jet (blue cone) containing two b -tagged sub-jets built from tracking information. The jet has a transverse momentum of 973 GeV and a mass of 116 GeV. A clear two-prong structure is visible as energy deposits in both the electromagnetic (green) and hadronic (yellow) calorimeters. A second muon (upper red track), shown with hits in the barrel muon chambers, is found within $\Delta R = 1.0$ of the Higgs-boson candidate.

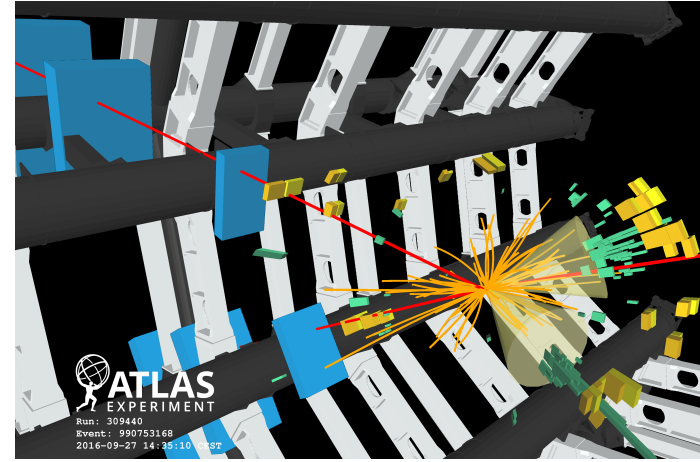
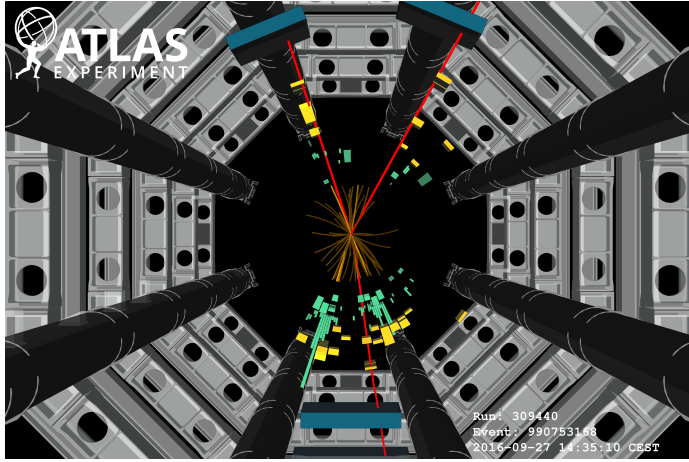


Figure 6.10: A 2 b -tag, 2-jet, 2-muon event within the most signal-like portion of the BDT output distribution (Sec. 6.7) is shown. Muons are shown as red tracks. Two of them form an invariant mass of 89 GeV, compatible with a Z boson, while a third one is found within ΔR of 0.4 of a jet and used in the muon-in-jet correction. The two central high- p_T b -tagged jets are represented by the green and yellow bars corresponding to the energy depositions in the electromagnetic and hadronic calorimeters, respectively, and they have an invariant mass of 123 GeV. The value of p_T^V is 204 GeV.

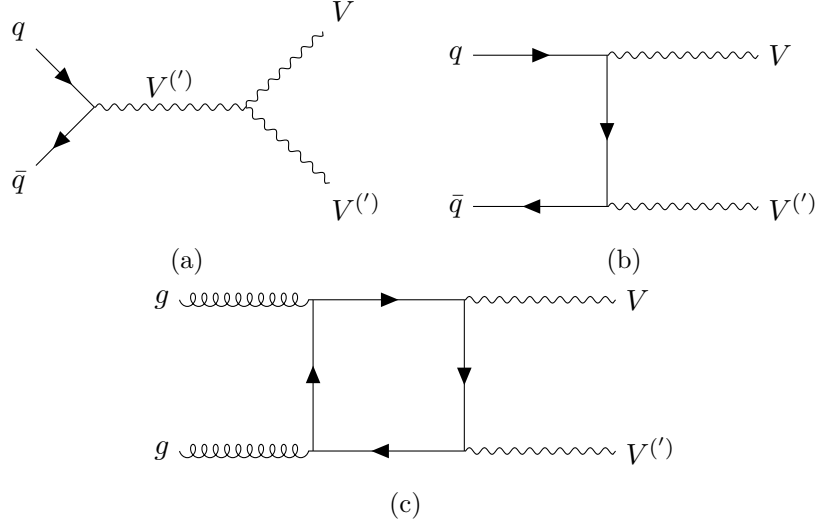


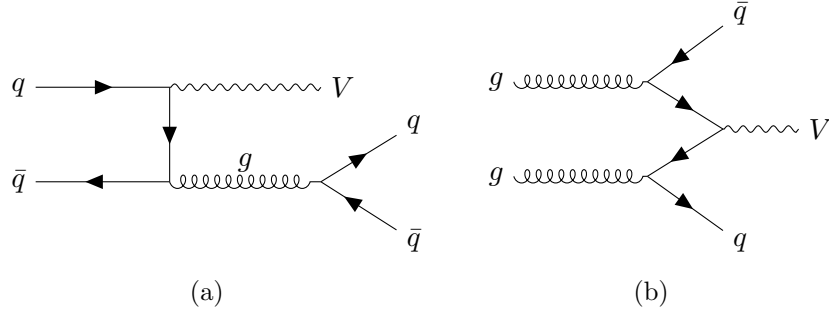
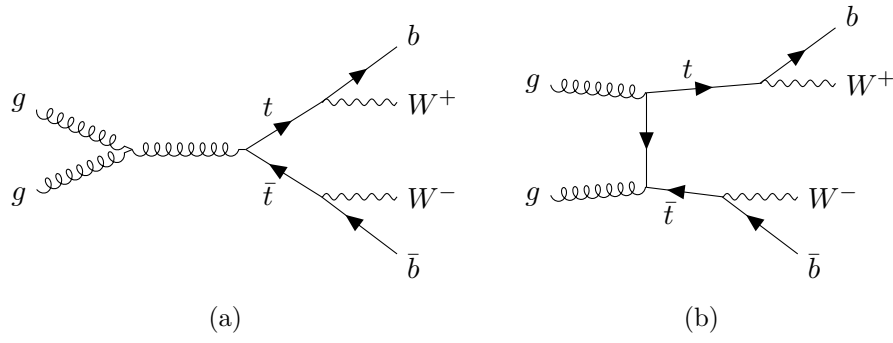
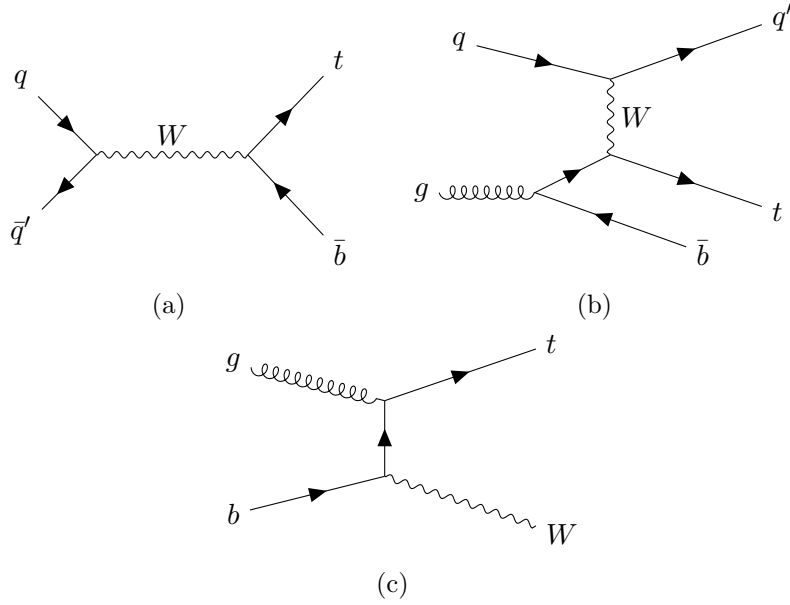
Figure 6.11: Feynman diagrams for diboson production.

6.2.3 The main backgrounds

Several background processes can yield the same signatures as the signal events. The backgrounds are categorized into reducible and irreducible, with distinct compositions in the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ regions, and across the three lepton channels. Irreducible backgrounds, namely diboson and V +jets production (Fig. 6.11 and 6.12), share the same final state particle count as the VH signal. The VZ part of the diboson process closely resembles the signal, resulting in a mass peak around the Z boson mass, analogous to the mass peak around the Higgs boson mass for the VH signal. The cross-section of VZ production is higher than that of VH production (Fig. 6.3), leading to a more prominent VZ peak. The VZ process is measured as a cross-check for the analysis strategy.

Reducible backgrounds, including $t\bar{t}$ production (Fig. 6.13), single top production (Fig. 6.14), and QCD multijet production (Fig. 6.15), pass the event selection due to instrumental limitations. These limitations arise from limited detector coverage, reconstruction inefficiencies, or objects falling outside kinematic acceptance. Although such inaccuracies are infrequent, these processes typically have cross-sections much larger than the signal, resulting in many events passing the selection.

Fig. 6.16, 6.17 and 6.18 show the contributions of the different processes to the SRs of the analysis in the $VH, H \rightarrow b\bar{b}$ resolved and boosted and $VH, H \rightarrow c\bar{c}$ regimes respectively. SRs are designed to contain most of the signal events and are described in Sec. 6.5.3.

Figure 6.12: Feynman diagrams for V +jets production.Figure 6.13: Feynman diagrams for $t\bar{t}$ production.Figure 6.14: Feynman diagrams for single top production in the s -channel (a), t -channel (b), and Wt -channel (c).

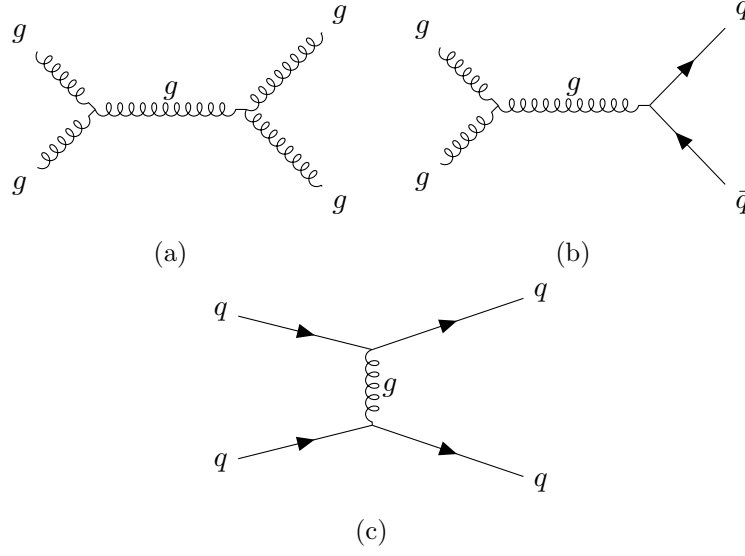


Figure 6.15: Feynman diagrams for QCD multijet production.

6.3 Data and simulated samples

6.3.1 Data

This measurement uses the Run-2 dataset, collected during the second operational run of the LHC from 2015 to 2018. In Run-2, as detailed in Chap. 2, the ATLAS detector recorded proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The data was recorded only for proton collisions happening during stable beam conditions. Furthermore, all relevant sub-detectors needed to be fully operational and the data had to be flagged as exempt from detector problems in the GRLs compiled by the ATLAS Data Quality group (Sec. 2.4). The analysed sample has an integrated luminosity of $140.1 \pm 1.2 \text{ fb}^{-1}$. Depending on the analysis sub-channel, a combination of triggers selecting events with high $E_{\text{T}}^{\text{miss}}$ or high- p_{T} leptons, described in Sec. 6.5.1, is used.

6.3.2 Simulated samples

The signal process and most background descriptions rely on MC simulations. The first exception is the multijet background, which is very challenging to simulate accurately, and is hence estimated from data using a dedicated data-driven technique. The second exception is the $t\bar{t}$ background in the 2L channel of the $VH, H \rightarrow b\bar{b}$ regions, where $t\bar{t}$ events can be identified very well in data using the top($e\mu$) CRs (Sec. 6.5.3).

The MC generators used in the analysis mostly reach NLO QCD accuracy in calculating

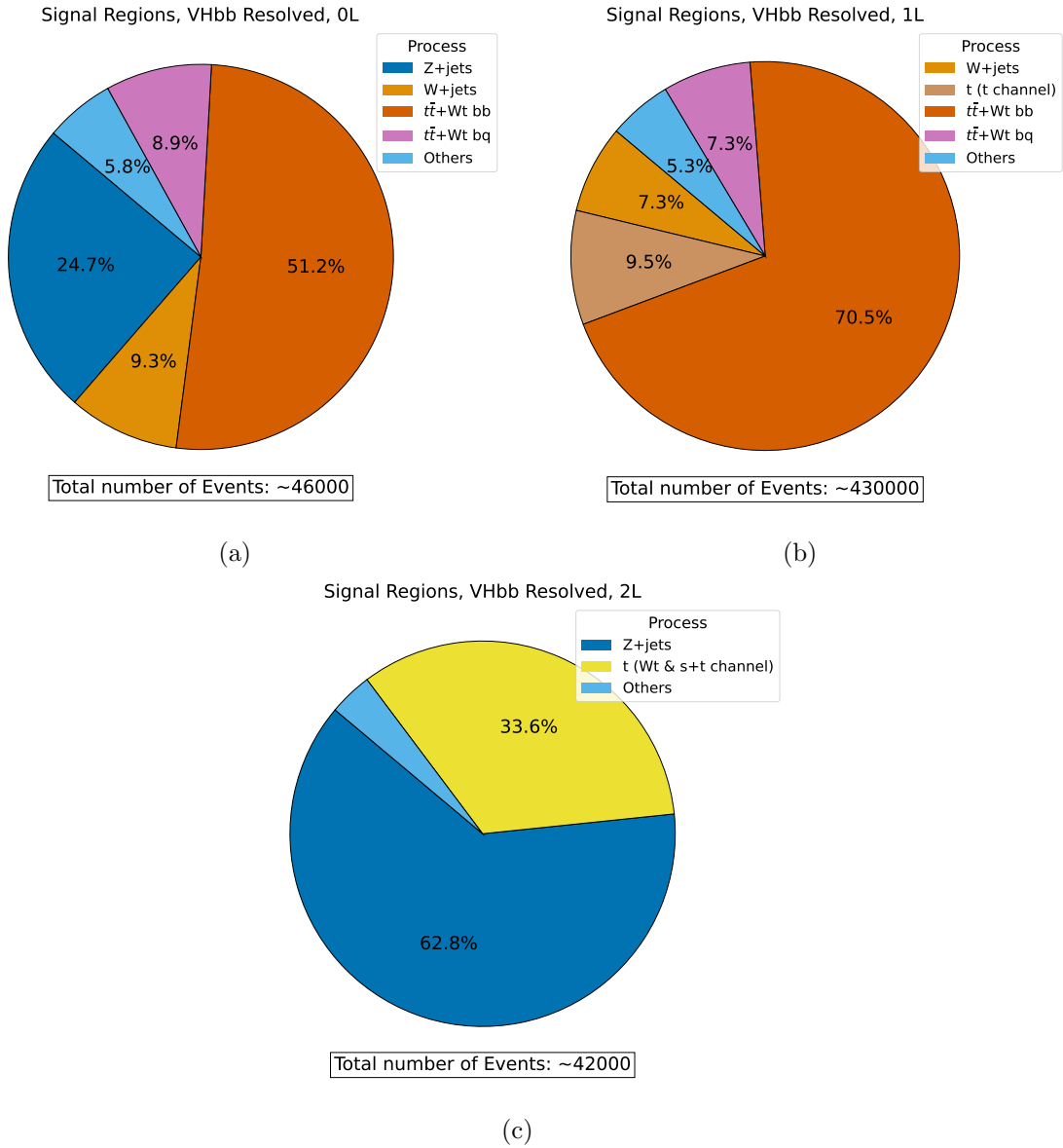


Figure 6.16: Estimated contribution, before the fit, of the signal and background processes in the SRs of the $VH, H \rightarrow b\bar{b}$ resolved analysis. All the SRs are merged together. Processes contributing with less than 5% of the total events are indicated as “Others”. The total count of events for each sample is normalized to the Run-2 luminosity times the most accurate theoretical prediction of its cross-sections.

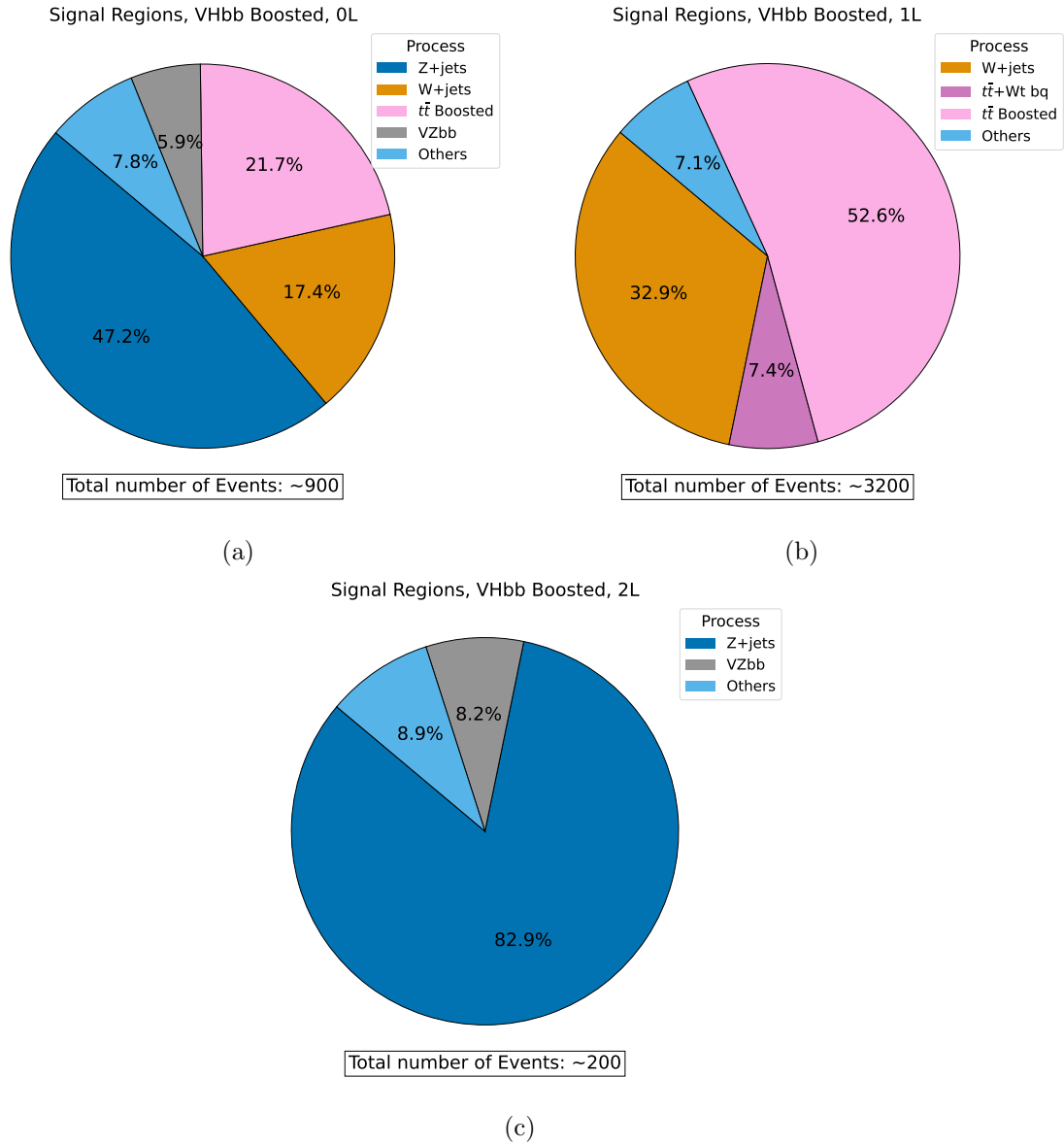


Figure 6.17: Estimated contribution, before the fit, of the signal and background processes in the SRs of the $VH, H \rightarrow b\bar{b}$ boosted analysis. All the SRs are merged together. Processes contributing with less than 5% of the total events are indicated as “Others”. The total count of events for each sample is normalized to the Run-2 luminosity times the most accurate theoretical prediction of its cross-sections.

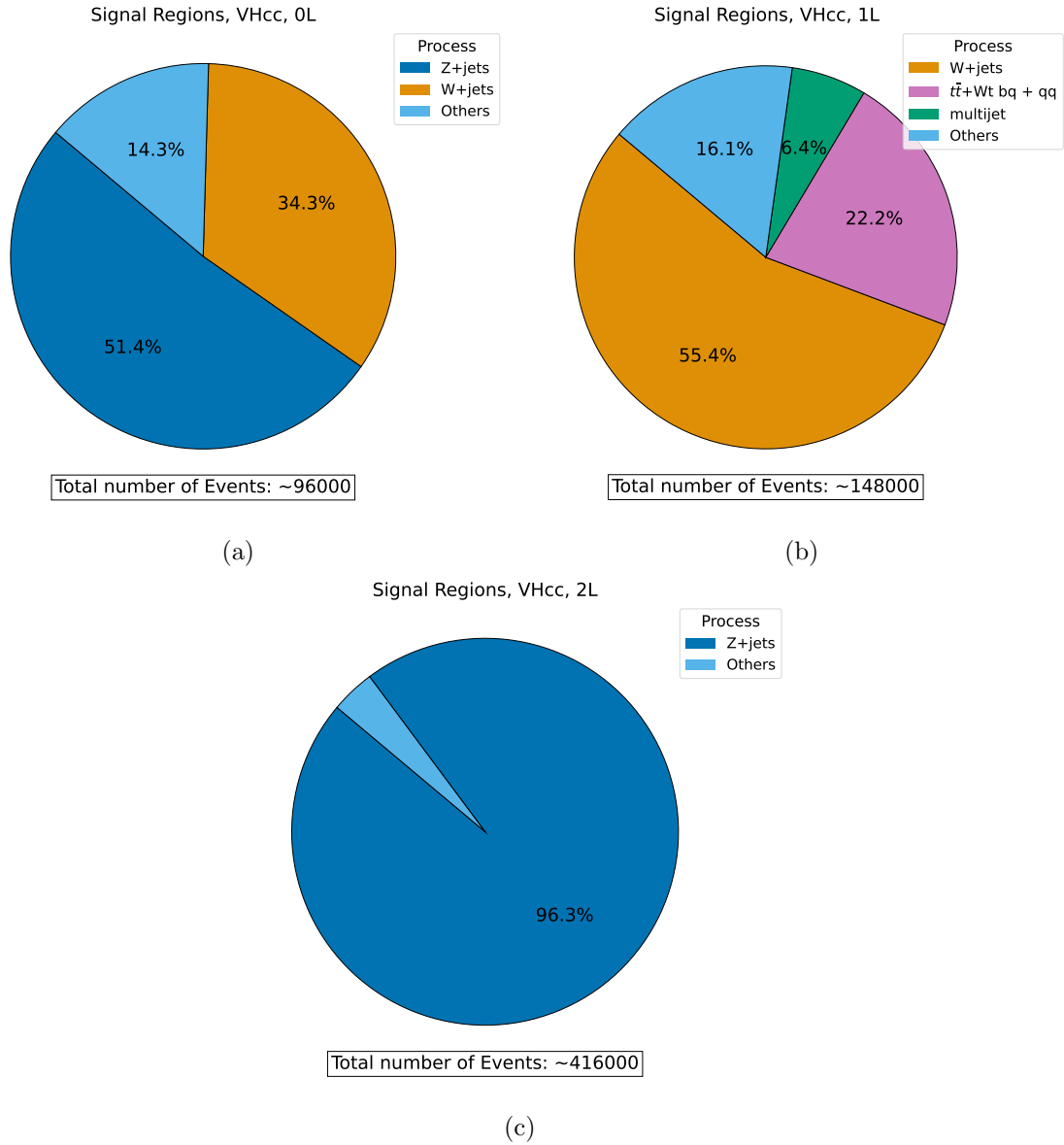


Figure 6.18: Estimated contribution, before the fit, of the signal and background processes in the SRs of the $VH, H \rightarrow c\bar{c}$ analysis. All the SRs are merged together. Processes contributing with less than 5% of the total events are indicated as "Others". The total count of events for each sample is normalized to the Run-2 luminosity times the most accurate theoretical prediction of its cross-sections.

matrix element amplitudes. Additionally, they include a parton shower step for shower evolution and hadronisation. In cases requiring accurate modelling of extra jets, like V +jets, multi-leg techniques ensure (N)LO accuracy in describing up to n hard, well-separated emissions, as detailed in Sec. 6.9.

Samples of simulated events were generated for VH production with a Higgs boson mass, m_H , of 125 GeV, for both $H \rightarrow b\bar{b}$ and $H \rightarrow c\bar{c}$ decays, with branching fractions of 2.89% and 58.2%, respectively, and for the main background processes. The setup of the *nominal* samples is summarized in Tab. 6.2. *Alternative* samples were produced with different generators to assess systematic uncertainties in the modelling of the processes, as detailed in Sec. 6.8.2.

Process	ME generator	ME PDF	PS and hadronisation	Tune	Cross-section order
$qq \rightarrow VH$ ($H \rightarrow c\bar{c}/b\bar{b}$)	POWHEG BOX v2 [119, 120] + GoSAM [129] + MINLO [130, 131]	NNPDF3.0NLO [121]	PYTHIA 8.212 [122]	AZNLO [123]	NNLO(QCD) +NLO(EW) [124–128]
$gg \rightarrow ZH$ ($H \rightarrow c\bar{c}/b\bar{b}$)	POWHEG BOX v2	NNPDF3.0NLO	PYTHIA 8.212	AZNLO	NLO+NLL [132, 133]
$t\bar{t}$	POWHEG BOX v2 [134]	NNPDF3.0NLO	PYTHIA 8.230	A14 [135]	NNLO +NNLL [136–142] i
t/s -channel single top	POWHEG BOX v2 [143]	NNPDF3.0NLO	PYTHIA 8.230	A14	NLO [144, 145]
Wt -channel single top	POWHEG BOX v2 [146]	NNPDF3.0NLO	PYTHIA 8.230	A14	Approx. NNLO [147, 148]
V + jets	SHERPA 2.2.11 [149]	NNPDF3.0NNLO [121]	SHERPA 2.2.11	Default	NNLO [150]
$qq \rightarrow VV$	SHERPA 2.2.11	NNPDF3.0NNLO	SHERPA 2.2.11	Default	NLO
$gg \rightarrow VV$	SHERPA 2.2.2	NNPDF3.0NNLO	SHERPA 2.2.2	Default	NLO

Table 6.2: Signal and background processes and their corresponding nominal MC generators used in this analysis. The acronyms ME and PS indicate matrix element and parton shower, respectively. The cross-section order refers to the order of the cross-section calculation used for process normalisation in QCD, unless otherwise stated, with NNLO and NNLL standing for next-to-next-to-leading order and next-to-next-to-leading log, respectively. *Tune* refers to the sets of parameters tuned in the MC generator to best model the effects of multi-partonic interactions, and of initial and final state radiation in data. Details on the chosen sets are provided in their references.

For many processes, cross-section computations are available at higher orders in perturbation theory than what is offered by MC generators. These calculations are used to normalize the simulated event yields derived from MC generators [135]. Given my contribution to its modelling, more details about the samples simulating the V +jets background are provided in Sec. 6.9.

6.4 Physics objects selection

In addition to the baseline requirements outlined in Sec. 2.3, additional selections of physics objects guided by the topology of the $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ signals are developed and here reported.

- **Electrons and muons** must satisfy specific identification and p_T -dependent isolation criteria that aim at discarding those compatible with non-prompt decays and fakes (Tab. 6.3 and 6.4).
- **τ leptons** decaying hadronically (Tab. 6.5) are clustered into jets and identified using a RNN developed within the ATLAS tau-identification group [118]. The RNN has increased performance compared to the algorithms used in the previous measurements of $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ processes. τ -leptons decaying leptonically are not specifically targeted, and enter the analysis only through the reconstruction of the leptons stemming from their decays.
- **Jets** used for the resolved topologies are PFlow small- R jets. As anticipated in Sec. 2.3, PFlow jets combine information from tracks and from topological clusters of energy deposits in the calorimeter. This allows for improvement in the resolution of their energy compared to the jets used in the previous measurements of $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ processes, which used only the information from the calorimeter. PFlow large- R jets are used to reconstruct the Higgs-boson candidate in the boosted regions. To reconstruct the individual quarks from the $H \rightarrow b\bar{b}$ decay within the large- R jet, variable-radius track jets are used. Tab. 6.6 summarizes the selection applied to small- R jets. Large- R jets are required to have $p_T > 250$ GeV and $|\eta| < 2.0$, and the variable- R track jets associated to them must satisfy $p_T > 10$ GeV and $|\eta| < 2.5$.

An overlap removal procedure is implemented to prevent the double-counting of reconstructed leptons, including hadronically decaying τ -leptons, and small- R jets.

6.5 Event selection and categorisation

To be accepted, events must satisfy a set of online and offline selections, customized in the different lepton channels and in the various regimes to target Higgs and vector boson

Electron Selection	p_T	η	ID	tracks d_0^{sig}	$ \Delta z_0 \sin \theta $	Isolation
<i>VH-Loose</i>	>7 GeV	$ \eta < 2.47$	LH Loose	< 5	< 0.5 mm	Loose
<i>ZH-Signal</i>	>27 GeV		Same as VH-Loose			
<i>WH-Signal</i>	Same as ZH-Signal		LH Tight	Same as ZH-Signal		Tight

Table 6.3: Criteria for selecting electrons. Electrons selected with the *VH-Loose* criterion are used to categorise the events into the 0-, 1- or 2L channels, whereas *VH-Signal* selection is optimised to reduce fakes in 1- and 2L channels separately based on the different composition of their backgrounds. LH Loose and LH Tight refer to increasingly stringent criteria to identify electrons using a likelihood-based method [50]. *WH-Signal* selection, designed for the 1L channel, applies the tighter criterion due to the greater contribution from the multijet background (Fig. 6.18b). For the same reason, these electrons are also required to satisfy tighter isolation criteria. Cuts on the IP of tracks decrease the impact of pile-up collisions.

Muon Selection	p_T	η	ID	d_0^{sig}	$ \Delta z_0 \sin \theta $	Isolation
<i>VH-Loose</i>	>7 GeV	$ \eta < 2.7$	Loose quality	< 3	< 0.5 mm	Loose
<i>ZH-Signal</i>	>27 GeV	$ \eta < 2.5$	Same as VH-Loose			
<i>WH-Signal</i>	>25 GeV when $p_T^V > 150$ GeV >27 GeV when $p_T^V < 150$ GeV	$ \eta < 2.5$	Medium quality	< 3	< 0.5 mm	Tight

Table 6.4: Criteria for selecting muons. Muons selected with the *VH-Loose* criterion are used to categorise the events into the 0-, 1- or 2L channels, whereas *VH-Signal* selection is optimised to reduce fakes in 1- and 2L channels separately based on the different composition of their backgrounds. Loose quality and Medium quality refer to increasingly stringent criteria to identify muons based on requirements on the ID and muon spectrometer tracks used to reconstruct them. *WH-Signal* selection, designed for the 1L channel, applies the tighter criterion due to the greater contribution from the multijet background (Fig. 6.18b). For the same reason, these muons are also required to satisfy tighter isolation criteria. Cuts on the IP of tracks decrease the impact of pile-up collisions.

Lepton	p_T	η	n_{trk}	ID
τ	>20 GeV	$ \eta < 2.5$	1 or 3 tracks	Loose

Table 6.5: Criteria for selecting hadronically decaying τ -leptons. These decays are characterised by the presence of one or three charged particles, with a neutrino, neutral pions and kaons. This explains the requirement on n_{trk} . The loose WP of the RNN used to identify τ -leptons corresponds to 85% and 75% efficiency on decays with 1 and 3 charged particles, respectively.

Jet Category	Tight Jet Cleaning	p_T	η	JVT/fJVT
<i>Signal Jet</i>	true	> 20 GeV	< 2.5	JVT >0.5 for $ \eta < 2.4$, $p_T < 60$ GeV jets
<i>Forward Jet</i>	true	> 30 GeV	$2.5 < \eta < 4.5$	fJVT <0.5 for $p_T < 120$ GeV jets

Table 6.6: Criteria for selecting small- R jets. *Signal jets* ($|\eta| < 2.5$) are reconstructed in the region of the detector where tracking information is available. Since tracking is crucial for flavour-tagging, only these are used for reconstructing the Higgs boson. *Forward jets* ($2.5 < |\eta| < 4.5$), instead, are used only for event categorisation. Jet cleaning removes signals in regions corresponding to hot calorimeter cells, and helps reduce non-collision noise. To suppress jets arising from pileup, Tight and Loose Jet Vertex Tagger selections (Cap. 4) are applied to signal or forward jets, respectively.

candidates while rejecting as many multijet events as possible.

6.5.1 Triggers

During data-taking, events considered in this measurement were selected using lepton triggers to target the leptonic decays of the vector bosons. Thus, 0L channel event candidates were selected with E_T^{miss} triggers. For the 1L channel, motivated by studies on the respective efficiencies in the two p_T^V regimes, different triggers are used. Muon events with $p_T^V > 150$ GeV were selected using the same E_T^{miss} triggers, as, at the trigger level, E_T^{miss} is computed using only the calorimeters and not the muon spectrometer. Muon events in the $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ region were selected by single-muon triggers. Electron events were selected using single-electron triggers. The strategy adopted in the 2L channel is similar to 1L channel's. In the electron sub-channel, events were selected if at least one single-electron trigger fired. In the muon sub-channel, E_T^{miss} triggers were used for events with $p_T^V > 250$ GeV, while single-muon triggers were used in the $75 \text{ GeV} < p_T^V < 250 \text{ GeV}$ region. The larger range of application of single muon trigger in this case is motivated by their higher efficiency in the 2L channel compared to 1L.

Leptons firing triggers are required to match those reconstructed offline.

6.5.2 Offline selections

First, events are categorized in the different lepton channels based on the numbers of *VH-Loose* leptons they contain (Tab. 6.3 and 6.4).

In the 0L channel, p_T^Z is reconstructed as the event E_T^{miss} ; in the 1L channel, p_T^W is the vector sum of the lepton's p_T with the event missing transverse momentum; and, in the 2L channel, p_T^Z is reconstructed as the vector sum of the transverse momenta of the two leptons. Events with $150(75) \text{ GeV} < p_T^V$ are considered for the 0L (1 and 2L) channels. The

different lower p_T^V limits are motivated by detailed studies on their impact on the expected significance of the final results.

In the 0L channel, the scalar sum S_T of the p_T of the jets must be larger than 120(150) GeV for events with 2 (or more than 2) jets. This cut removes a region mismodelled in the simulations, due to difficulties in modelling the trigger efficiencies turn-on curves as they exhibit a non-trivial dependence on the jet activity [98].

In the 2L channel, the two leptons must have $p_T > 25$ GeV and have the same flavour, with the exception of the events entering the $\text{top}(e\mu)$ CR, described in Sec. 6.5.3. Furthermore, at least one must be identified with the ZH -Signal criterion (Tab. 6.3 and 6.4). Di-muon events must have opposite charges, as opposed to di-electron events, which are affected by higher rates of charge misidentification. Since in the 2L channel it is possible to reconstruct completely the leptons emerging from the decay of the Z -boson, only events with a mass m_{ll} of the di-lepton system reconstructed within an $81(66) \text{ GeV} < m_{ll} < 101(116 \text{ GeV})$ around the Z mass are accepted in the resolved (boosted) analysis. The different widths of these mass windows are justified by the different resolutions for reconstructing m_{ll} in the different regimes.

In the resolved regime, 0L events containing a τ lepton that decays hadronically are moved to the 1L channel and required to have $m_T^W > 10$ GeV. This cut reduces contamination from fake hadronic tau candidates. In the 1L channel, events containing a hadronically decaying τ -lepton and a VH -loose lepton are discarded, rather than being assigned to the 2L channel. This is because they can be mimicked by $t\bar{t}$ events with one W -boson decaying to a neutrino and a hadronically decaying τ -lepton.

Cuts concerning the reconstruction of the Higgs-boson candidate Higgs-boson candidates are reconstructed with different procedures in the resolved and merged topologies.

First, events without any b -tagged small- R jet, but having at least one tight c -tagged small- R jet or exactly one loose c -tagged small- R jet are considered for the resolved $VH, H \rightarrow c\bar{c}$ regions.

Events with $p_T^V < 400$ GeV, exactly two b -tagged small- R jets, and, in 0L and 1L, no tight c -tagged small- R jet are considered for the resolved $VH, H \rightarrow b\bar{b}$ regions.

Events in the 0 and 1L channel containing both b -tagged and tight c -tagged small- R jets are considered for a $\text{top}(bc)$ CR which is useful for constraining $t\bar{t}$ contributions in the fit⁵.

Events with $p_T^V \geq 400$ GeV and at least two b -tagged variable-radius track-jets are considered for the $VH, H \rightarrow b\bar{b}$ boosted regions.

These selections ensure that the events contain the ingredients for identifying a Higgs-boson

⁵This CR is not implemented in 2L due to the possibility of introducing in that channel the $\text{top}(e\mu)$ CR, which is much purer in $t\bar{t}$ events (Sec. 6.5.3).

candidate H_{cand} .

In the resolved $VH, H \rightarrow b\bar{b}$ regions, H_{cand} is produced by summing the quadrimomenta of the two b -tagged jets.

In $VH, H \rightarrow c\bar{c}$, jets are prioritized based on their c -tagging score according to the Tight c -tag > Loose c -tag > No-tag rule, and, for jets falling in the same c -tagging bin, according to decreasing transverse momentum. H_{cand} is produced by summing the two jets ranked the highest.

In both the $VH, H \rightarrow c\bar{c}$ and resolved $VH, H \rightarrow b\bar{b}$ regions, the two jets forming H_{cand} are sorted according to their p_T . The leading jet (j_1) is required to have $p_T > 45$ GeV. Due to difficulties of modelling gluon splitting in V +jets events with low-mass Higgs-boson candidates, the invariant mass of H_{cand} must be larger than 50 GeV. Additionally, it is difficult to model the angular separation ΔR between j_1 and the sub-leading jet j_2 in events with ΔR larger than π . Hence, these events are discarded. This cut is justified because only a tiny fraction of signal events form the Higgs-boson candidate with jets so separated.

In the boosted $VH, H \rightarrow b\bar{b}$ analysis, H_{cand} is the large- R jet with the highest p_T . It must be matched with at least two variable-radius track jets. These are sorted according to their p_T , and the first three are considered for flavour-tagging. If at least two are b -tagged, and the mass of the large- R jet m_J is larger than 50 GeV, the event is selected.

In the 0 and 1L channels, some energy is undetected due to the presence of neutrinos. Background multijet events can be selected in these channels when mismeasurements of the energies of their jets result in fake E_T^{miss} . Dedicated cuts are therefore applied to reduce multijet contamination. In the 0L channel, the following conditions must be met:

- $\Delta\phi(E_T^{\text{miss}}, H_{cand}) > 120^\circ$,
- $\min[\Delta\phi(E_T^{\text{miss}}, \text{small-}R \text{ jets})] > 20(30)^\circ$ in events with 2(3) small- R jets,
- $\Delta\phi(j_1, j_2) < 140^\circ$ in events in the resolved regions.

The first condition ensures that E_T^{miss} points away from the Higgs-boson candidate, as expected in the signal topologies. The second condition prevents E_T^{miss} from being aligned with any jet, reducing the probability that it arises from mismeasurements of the jet energy. The third condition discards events with leading jets too much separated in azimuthal angles, which happens more often in QCD events than in events with $H \rightarrow c\bar{c}$ or $H \rightarrow b\bar{b}$ decays. Together, these cuts reduce multijet events to less than 1% of the total, making such background negligible in the 0L channel. In 1L, E_T^{miss} is required to be larger than 30 (50) GeV in the resolved (boosted) regions; and the mass m_W^T of the W boson is required to be larger than 20 GeV for events with $75 \text{ GeV} < p_T^W < 150 \text{ GeV}$, as it is reconstructed relying on E_T^{miss} .

Tab. 6.7 and 6.8 summarize the online and offline selections in the various lepton channels for the resolved and boosted regimes.

Resolved Regime	$VH, H \rightarrow b\bar{b}$	$VH, H \rightarrow c\bar{c}$
	Selections applied to all lepton channels	
Jets		≥ 2 signal jets
Candidate jets tagging	2 B-tags	≥ 1 T-tag (or exactly 1 L-Tag)
Leading Higgs candidate jet p_T		> 45 GeV
Sub-leading Higgs candidate jet p_T		> 20 GeV
m_{bb} or m_{cc}		> 50 GeV (before correction)
$\Delta R(\text{jet1, jet2})$		Upper cut $\Delta R < \pi$
	0 lepton	
Trigger		E_T^{miss} triggers
Jets	≤ 4 jets	≤ 3 jets
Additional jets tagging	no T-tag	no B-tag (except for top(bc) CR)
Leptons		0 VH -loose lepton
E_T^{miss}		> 150 GeV
S_T		> 120 (2 jets), > 150 GeV (3p jets)
m_T^W		> 10 GeV (for events with at least one hadronic τ)
$ \min\Delta\phi(E_T^{\text{miss}}, \text{jet}) $		$> 20^\circ$ (2 jets), $> 30^\circ$ (3 jets)
$ \Delta\phi(E_T^{\text{miss}}, H) $		$> 120^\circ$
$ \Delta\phi(\text{jet1, jet2}) $		$< 140^\circ$
	1 lepton	
Trigger		e channel: single electron trigger μ channel: single muon trigger ($p_T^V < 150$ GeV) and E_T^{miss} triggers (above)
Jets		≤ 3 jets
Additional jets tagging	no T-tag	no B-tag (except for top(bc) CR)
hadronic τ -veto		no hadronic τ
Leptons		1 WH -signal lepton
		> 1 VH -loose lepton veto
E_T^{miss}		> 30 GeV (e channel)
S_T		> 120 (2 jets), > 150 GeV (3 jets) (μ channel with E_T^{miss} trigger)
m_T^W		> 20 GeV ($75 < p_T^V < 150$ GeV only)
	2 lepton	
Trigger		e channel: single electron trigger μ channel: single muon trigger ($p_T^V < 250$ GeV) and E_T^{miss} triggers ($p_T^V \geq 250$ GeV)
Additional jets tagging	-	no B-tag
Leptons		2 VH -loose leptons (≥ 1 ZH -signal lepton)
		Same flavour (Except for top($e\mu$) CR), opposite-charge for $\mu\mu$
m_{ll}		$81 < m_{ll} < 101$ GeV

Table 6.7: Summary of the event selection during and after data-taking in the resolved regime.

Selection	0 lepton channel	1 lepton channel		2 lepton channel	
		e sub-channel	μ sub-channel	e sub-channel	μ sub-channel
Trigger	$E_{\text{T}}^{\text{miss}}$	Single electron	$E_{\text{T}}^{\text{miss}}$	Single electron	$E_{\text{T}}^{\text{miss}}$
Leptons	0 VH -Loose lepton	1 WH -Signal lepton no second VH -Loose lepton no hadronic τ		≥ 1 ZH -Signal lepton 2 VH -Loose leptons	
$E_{\text{T}}^{\text{miss}}$	$> 250 \text{ GeV}$	$> 50 \text{ GeV}$	-	-	
p_{T}^V	$p_{\text{T}}^V > 400 \text{ GeV}$				
large- R jet	at least one large- R jet, $p_{\text{T}} > 250 \text{ GeV}$, $ \eta < 2$				
Track-Jets	at least two track-jets, $p_{\text{T}} > 10 \text{ GeV}$, $ \eta < 2.5$, matched to the leading large- R jet				
b -jets	exactly two of the leading three track-jets matched to the leading large- R must be b -tagged				
m_{J}	$> 50 \text{ GeV}$				
$\min \Delta\phi(E_{\text{T}}^{\text{miss}}, \text{jets})$	$> 30^\circ$	-			
$\Delta\phi(E_{\text{T}}^{\text{miss}}, H_{\text{cand}})$	$> 120^\circ$	-			
$m_{\ell\ell}$	-			$66 \text{ GeV} < m_{\ell\ell} < 116 \text{ GeV}$	
lepton flavour	-			two lepton same flavour	
lepton charge	-			opposite sign muons	

Table 6.8: Summary of the event selection during and after data-taking in the boosted regime.

6.5.3 Event categorisation

The selected events are then split into exclusive categories, based on the events' p_T^V , number of jets, flavour-tagging content, angular separation between the jets forming the Higgs-boson candidate, and flavours of the leptons. These categories are schematized in Fig. 6.19, and described in the following paragraphs.

p_T^V Categorisation: According to the p_T^V of the reconstructed V -boson candidate they contain, events are distributed in the following categories: $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$, $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$, $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$, $p_T^V > 250 \text{ GeV}$, $400 \text{ GeV} < p_T^V < 600 \text{ GeV}$, $p_T^V > 600 \text{ GeV}$. This categorization is motivated by the fact that the signal-to-background ratio increases with p_T^V , as well as that a variety of BSM models predict deviations from the SM for specific values of the Higgs-boson transverse momentum. Furthermore, it enables measuring differential cross-sections in distinct regions.

Categorisation in Number of Jets: According to the number of selected jets they contain, events are further distributed in 2, 3, 4, 3 or more (3p) and 4 or more (4p) jets categories. This categorization is motivated by studies to maximise the overall signal significance in the fit. Furthermore, it enables measuring differential cross-sections for events with distinct jet multiplicities in the final state.

Categorisation According to the Jet Flavour-Tagging Score: Events used for the $VH, H \rightarrow c\bar{c}$ regions are categorized according to the PCFT scores of the jets forming the Higgs-boson candidate. Events are distributed into TT, LT, NT and LN⁶ regions. This categorization is motivated by the fact that different PCFT bins define regions with very different populations of backgrounds. Separating the events provides the analysis with the flexibility to target the individual sources of backgrounds with dedicated CRs, while accepting as much signal events as possible.

Control Regions : CRs are designed to be enriched in events from background processes with strong kinematic similarities with signal processes. Thanks to their constraining power on individual sources of background events, CRs allow for the normalisation of templates representing the main backgrounds to be left free to float during the fitting process. This enables extracting directly from data the contribution of background events to the total yields in the SRs (SRs). Additionally, CRs effectively constrain the shapes of these contributions in the SRs thanks to dedicated shape uncertainties, through the methods described in Sec.6.8.2. In the resolved regimes, the CRs are:

- **High- ΔR CRs:** Designed to constrain the V +jets and $t\bar{t}$ background normalisation and shapes. For each SR, a p_T^V -dependent continuous upper cut on the ΔR between the two jets forming the Higgs-boson candidate is defined to select 95% (85%) of signal events with 2 (more than 2) jets in each p_T^V region. The events in the region above the cut, instead, are used for the high- ΔR CRs. Fig. 6.21 exemplifies the composition of the backgrounds in the high- ΔR CRs. In the figure's caption, MC distributions are termed *prefit*, meaning that the templates are shown as predicted by the MC generators before fitting them to data.

⁶The LN combination is used only for the V +light CR described later in this section.

- **Top(bc) CR:** The $t\bar{t}$ events entering in the high- ΔR CRs mostly contain Higgs-boson candidates formed with two b -jets, which will be addressed as top(bb) henceforth. The top(bc) CRs are designed to control the top(bc) and top(bl) background events. They are constructed by selecting events with at least one tight c -tagged and at least one b -tagged jet. Fig. 6.20 exemplifies the composition of the backgrounds in the top(bc) CRs.
- **Top($e\mu$) CRs:** In the 2L channel, the dominant background events with top quarks arise from di-lepton $t\bar{t}$ decays. The top($e\mu$) CRs are designed to be very pure in these events by selecting differently flavoured leptons. This selection eliminates most of the other background and signal events since they are expected to contain lepton pairs of the same flavour originating from the Z -boson decay. In the resolved $VH, H \rightarrow b\bar{b}$ region, when performing the fit, the top($e\mu$) CRs are employed to derive the $t\bar{t}$ background templates for the SRs, rather than as separate fitting regions with their own floating templates. This data-driven approach is not applied in $VH, H \rightarrow c\bar{c}$ due to the more limited statistics. Furthermore, $t\bar{t}$ is a tiny background in the $VH, H \rightarrow c\bar{c}$ case. There, events in the top($e\mu$) CRs are required to contain at least one tight c -tagged jet. In the fit, these CRs are used as separate fitting regions. Fig. 6.22 exemplifies the composition of the backgrounds in the top($e\mu$) CRs.
- **Low- ΔR CRs:** In the 1L channel, it is possible to design a CR enriched in W +jets events where the two jets forming the Higgs-boson candidate are either both b or both c -jets. The average angular separation between the two jets is smaller for these events than for the signal. Leveraging this feature, a lower p_T^V -dependent continuous cut on the ΔR between the two jets is defined. Rather than being defined to retain a given proportion of signal events as in the SR, as was the case for the high- ΔR cut, the low- ΔR cut ensures that 90% of the diboson events are maintained in the SR. As the selected diboson events tend to produce more collimated jet pairs, this requirement ensures that the SR retains enough diboson events to perform a fit of the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$, which provides a good SM candle to validate the analysis design, while effectively isolating a W +jets enriched CR. The composition of the background in the low- ΔR CR is exemplified in Fig. 6.23, which displays the score $mvaCRLow$ of a BDT used only in this region to separate W +heavy flavour ($W + hf$) jet events from the others.
- **V +light CRs** Consisting of events where one of the jets forming the Higgs-boson candidate is loose c -tagged, while the other is not tagged, these CRs are enriched in V +light-flavour jets. Fig. 6.24 exemplifies the composition of the backgrounds in the V +light CRs.

In the boosted regime, only a top-CR is defined by requiring an additional b -tagged track jet outside the H_{cand} jet. The region is enriched in $t\bar{t}$ events because jets produced by the hadronic decay of the W boson emerging from the decay of one of the two top quarks can be misinterpreted as a Higgs-boson candidate. This means that the b -quark stemming from

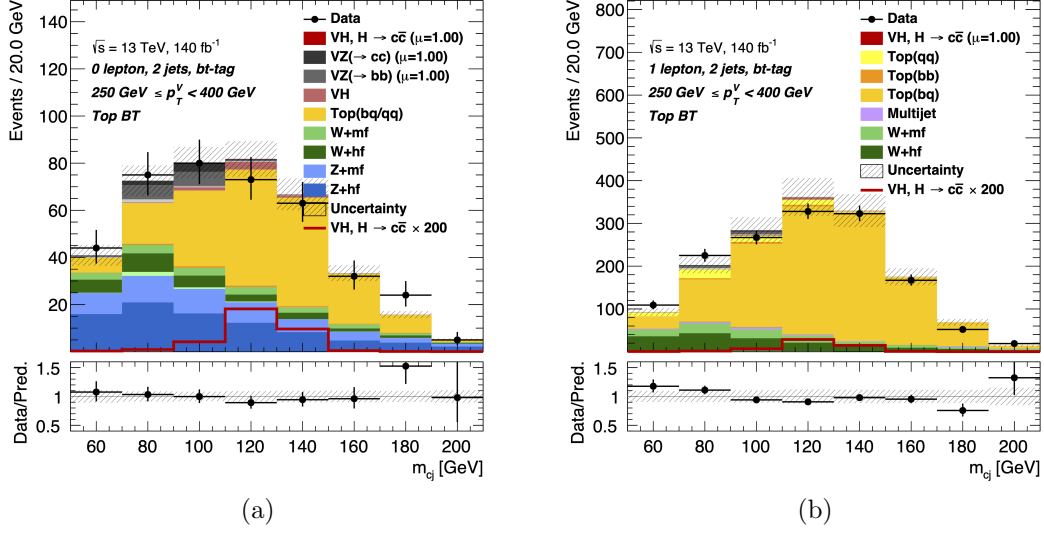


Figure 6.20: Data and prefit MC distributions of events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ $\text{top}(bc)$ CR. The regions are enriched in $\text{top}(bq)$ and $\text{top}(qq)$ events. The distributions that are displayed are those used as templates in the fit.

the decay of the other top quark, which is often produced with a wide angular separation from the former, is expected to be reconstructed outside the Higgs-boson candidate large- R jet. Fig. 6.25 exemplifies the composition of the backgrounds in the top CR of the boosted regime.

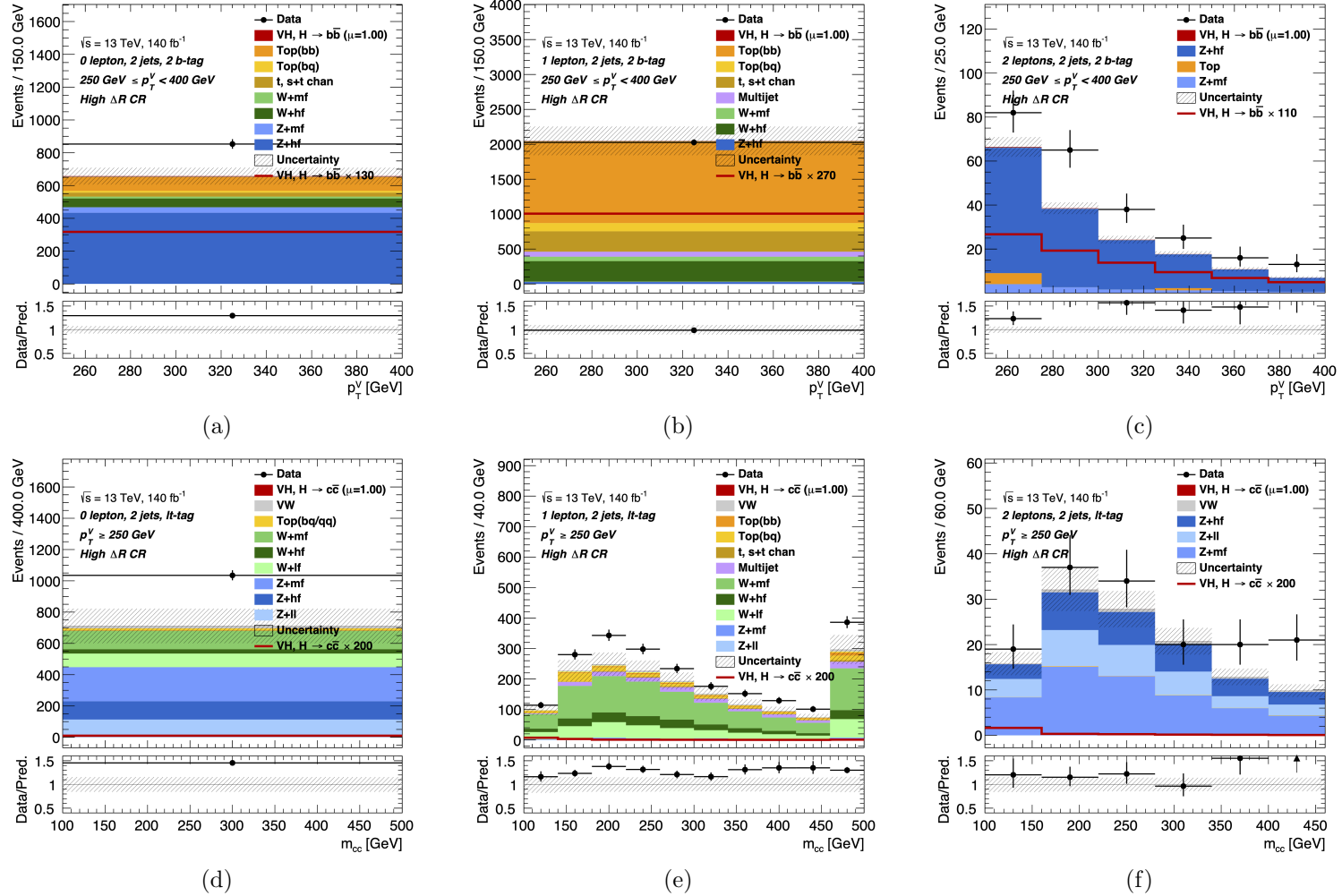


Figure 6.21: (a, b and c) Data and prefit MC distributions of events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ high- ΔR CR in the resolved $VH, H \rightarrow b\bar{b}$ regime. (d, e, f) Data and prefit MC distributions of events entering the 2-jet $p_T^V > 250 \text{ GeV}$ LT c -tag high- ΔR CR in the $VH, H \rightarrow c\bar{c}$ regime. The regions are enriched in V +jets and top(bb) events. The distributions that are displayed are those used as templates in the fit.

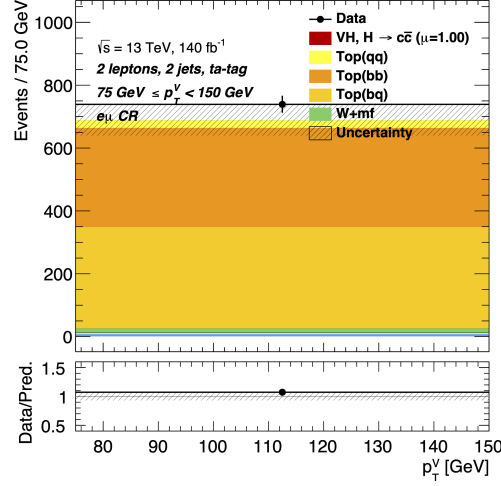


Figure 6.22: Data and prefit MC distributions of events entering the 2-jet $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ $\text{top}(e\mu)$ CR in the $VH, H \rightarrow c\bar{c}$ regime. The region is enriched in $\text{top}(bb)$ and $\text{top}(bq)$ events. The distribution that is displayed is used as a template in the fit.

6.6 Corrections to the Higgs-boson candidate mass

In contrast to the previous $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ analyses, where the H_{cand} mass distribution was directly used as a template in the SRs during the fit, the distribution of the output scores of BDTs separating signal from background events are employed in this case (Sec. 6.7). However, the sensitivity of the analysis still depends strongly on accurately determining the invariant mass of the Higgs-boson candidate H_{cand} , since such a variable remains crucial in training the BDTs. To further improve the mass resolution and, thereby, the sensitivity of the analysis, some analysis-specific corrections are applied. Their impacts in the resolved and boosted $VH, H \rightarrow b\bar{b}$ and in the $VH, H \rightarrow c\bar{c}$ regions are showcased in Fig. 6.26.

Muon-in-Jet Corrections Muons deposit little fractions of their energies in the calorimeters. When muons reconstructed in the muon spectrometer are matched to jets forming H_{cand} , their quadrimomenta are added back to the jets'. This contributes to a more accurate determination of the energy of the b -jets and c -jets used to reconstruct H_{cand} , as the heavy-flavour hadrons they contain frequently undergo semi-leptonic decays, producing muons in approximately 10% of cases.

Final State Radiation (FSR) Recovery Quarks, produced in the final states of the events targeted in this analysis, emit QCD radiation and hadronize. It happens that jets not used to reconstruct H_{cand} stem from the hadronization of colour charges radiated by the

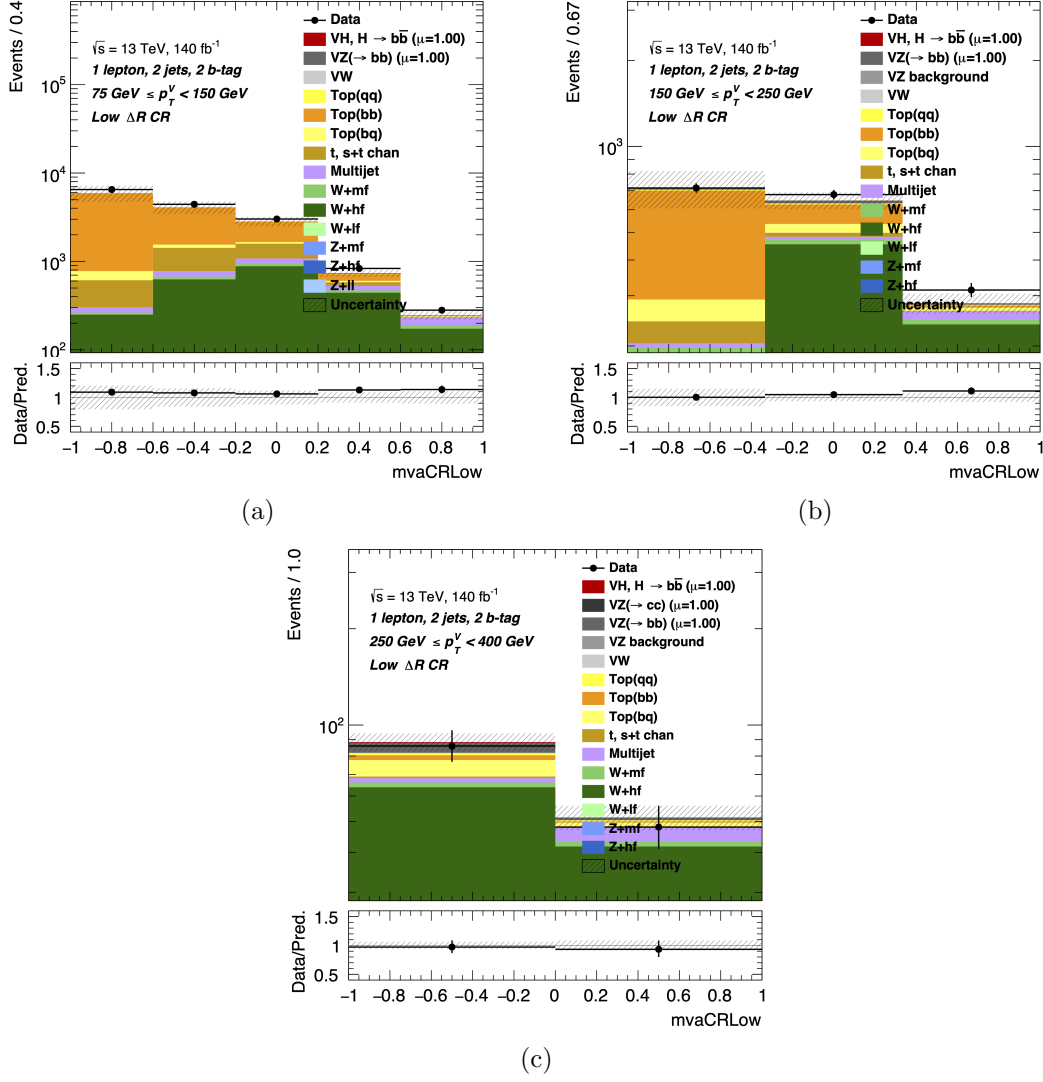


Figure 6.23: Data and prefit MC distributions of events entering the 2-jet $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ low- ΔR CR. In this region, a BDT is trained to separate $W + hf$ events from the other processes. As a result, the bins with a higher score of the BDT's output distribution are very pure in $W + hf$ events. The distributions that are displayed are those used as templates in the fit.

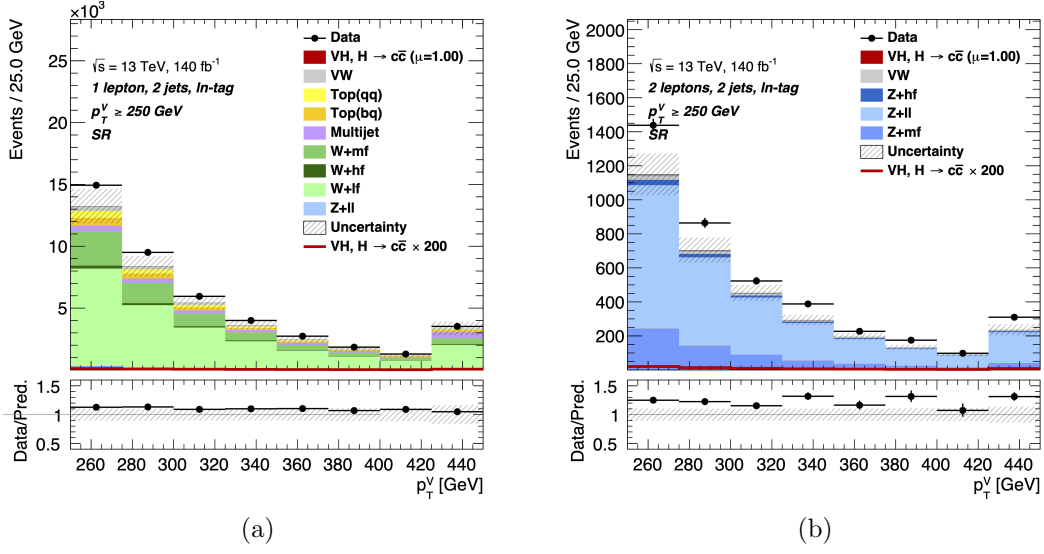


Figure 6.24: Data and prefit MC distributions of events entering the 2-jet $p_T^V > 250$ GeV V +light-flavour jets CR. The regions are enriched in V +light-flavour jets events. The distributions that are displayed are those used as templates in the fit.

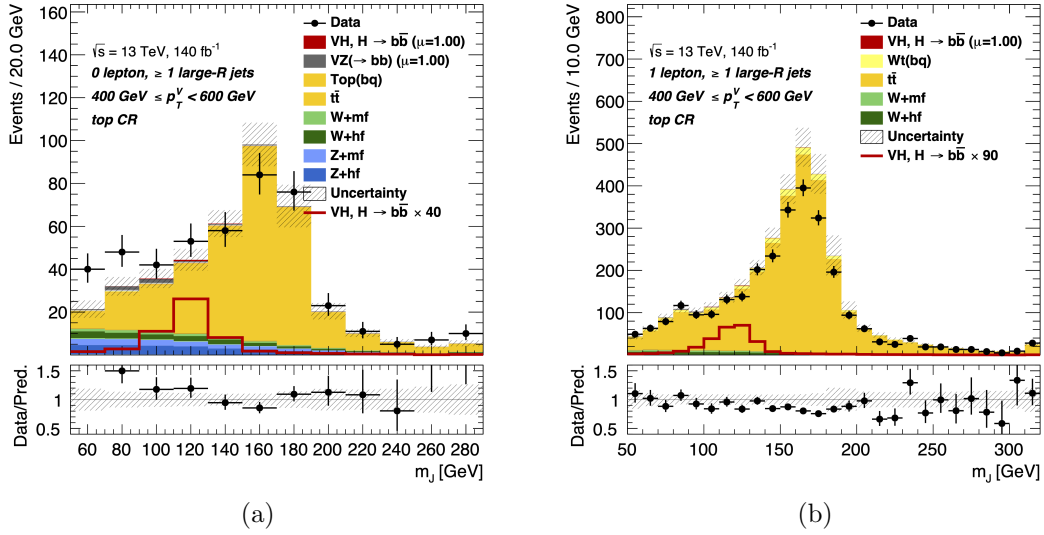


Figure 6.25: Data and prefit MC distributions of events entering the 400 GeV $< p_T^V < 600$ GeV Top CRs of the boosted $VH, H \rightarrow b\bar{b}$ analysis. The regions are enriched in $t\bar{t}$ events. The distributions that are displayed are those used as templates in the fit.

final state quarks. A procedure is implemented in the analysis to re-assign their energies to the jets forming H_{cand} . A continuous cut is established on the sum of the ΔR between the additional jets and the two jets forming H_{cand} as a function of p_{T}^V . Jets falling below this cut are designated as radiation jets, and their quadrimomenta are added back to the nearest jet forming H_{cand} . The cut is optimized using signal and background samples to maximize the H_{cand} -mass based significance $S = \sum_i [2(s_i + b_i) \ln(1 + \frac{s_i}{b_i})]$, where s indicates signal events, b indicates background events, and where i runs over the bins of the distributions of the H_{cand} mass in the 3-jet and 4-jet events of the resolved SR.

Truth jet- p_{T} correction In simulated events, the determination of the energy of the reconstructed heavy-flavour jets can be refined by adding back the undetected energy of the neutrinos and the energy of soft-particles not associated to the jet during the clustering process, which are often produced when heavy-flavour hadrons undergo semi-leptonic decays. This correction is possible because of the truth record about the origin of every simulated particle kept in the simulated events.

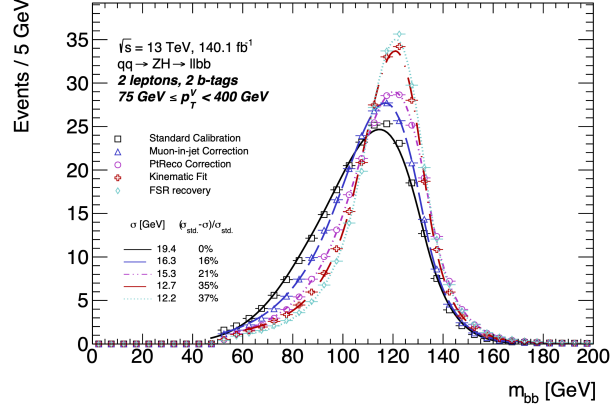
Kinematic fit of the H_{cand} p_{T} In events entering the 2L channel, it is possible to fully reconstruct the vector boson candidate from the leptons emerging from its decay. A kinematic fit is hence performed to improve the resolution of the measurement of the hadronic system in the transverse plane by balancing it against the well-measured leptonic system and applying a constraint on the reconstructed dilepton mass. The kinematic fit enables utilizing the percent-level resolution on the measurement of the lepton energies and momenta to improve the corresponding measurements in jets, whose resolutions are of the order of $\sim 5 - 10\%$ [151].

6.7 Multivariate BDT to enhance signal vs background separation

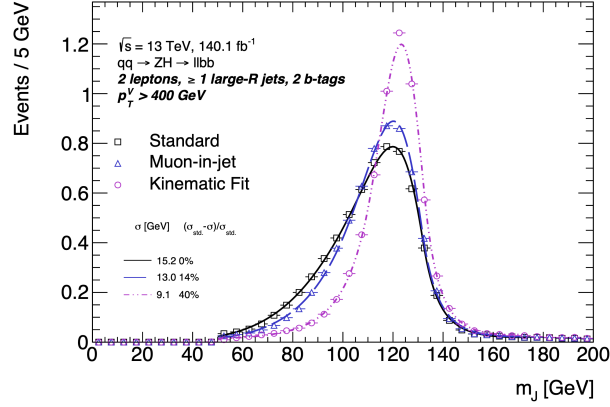
Similar to the high-level taggers described in Chap 3 for separating jets with different flavours, BDTs can be used to combine a set of input variables discriminating signal from background events into a single distribution with enhanced separation. It was already underlined that their introduction in the $VH, H \rightarrow c\bar{c}$ and boosted $VH, H \rightarrow b\bar{b}$ regions, in addition to the resolved $VH, H \rightarrow b\bar{b}$ region, represents one of the main improvements of the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis.

To optimize training statistics, this analysis' BDTs, although employed only in the SRs, are trained using events from both the SRs and high- ΔR CRs, post-correction of the H_{cand} mass as described in Sec. 6.6⁷. The BDTs were trained with $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$

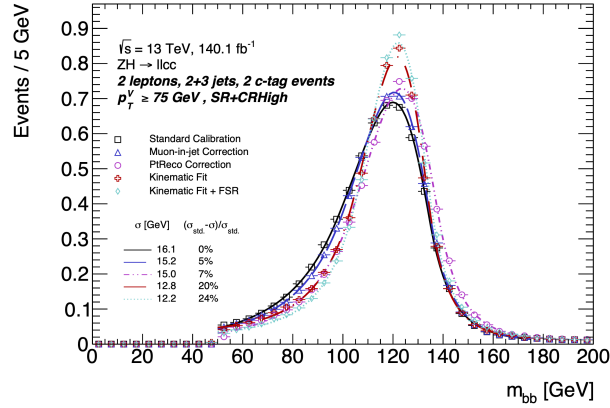
⁷Training solely on SRs was considered, but it led to overtraining issues without performance improvement.



(a)



(b)



(c)

Figure 6.26: Reconstructed Higgs-boson candidate mass of simulated $ZH, H \rightarrow b\bar{b}$ events in the 2L channel of the resolved (a) and boosted (b) $VH, H \rightarrow b\bar{b}$ regions and of simulated $ZH, H \rightarrow c\bar{c}$ events in the 2L channel of the $VH, H \rightarrow c\bar{c}$ regions (c). Only events passing the event selection are included. The Higgs-boson candidate mass distribution after the standard calibration is shown in black, after the muon-in-jet correction in blue, after the kinematic fit in violet, and after the FSR in light blue. The best improvement in the mass resolution measurement, of the order of 15%, is brought about by the kinematic fit.

samples as signals respectively in the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ regions. In both cases, V +jets, $t\bar{t}$ and diboson samples were used as backgrounds⁸. Additional BDTs were trained using $VZ, Z \rightarrow b\bar{b}$ and $VZ, Z \rightarrow c\bar{c}$ diboson samples as signal, to validate the BDT method and to provide the templates used in the diboson fit (Sec. 6.12.2).

Due to significant kinematics and background composition variations across different analysis regions, distinct BDTs are trained for application in distinct regions.

For the boosted topology, three BDTs are trained for the different lepton channels.

For the resolved topologies, more distinctions are introduced. As far as tagging regions are concerned, separate BDTs are trained for the 2 c -tag region of the $VH, H \rightarrow c\bar{c}$ regime, comprising TT and TL events, and for the 1 c -tag region, which contains TN events; while no separation based on b -tagging is applied for $VH, H \rightarrow b\bar{b}$ events. In the 0L channels of both the $VH, H \rightarrow b\bar{b}$ resolved and $VH, H \rightarrow c\bar{c}$ regimes, separate BDTs are trained for 2, 3, and 4-jet events with $p_T^V > 150$ GeV. In the 1L channel, separate BDTs are trained in the 2 and 3-jet categories, further divided in $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ and $p_T^V > 150 \text{ GeV}$. In the 2L channel, separate BDTs are trained in the 2 and 3p-jet categories, divided into $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ and $p_T^V > 150 \text{ GeV}$.

The BDTs are trained using distinct input variables across different lepton channels, with no further differentiation introduced for the additional categories within each lepton channel. They are summarized in Tab. 6.9. Following are their definitions.

Resolved input variables

- p_T^V : transverse energy of the reconstructed vector boson candidate.
- $p_T^{j_1}$ and $p_T^{j_2}$: transverse momenta of the jets forming H_{cand} .
- $m_{j_1 j_2}$: H_{cand} invariant mass after corrections.
- $\Delta R(j_1, j_2)$: angular distance between the two jets forming H_{cand} .
- $m_{j_1 j_2 j_3}$: invariant mass of the two jets forming H_{cand} and the remaining jet with the highest p_T .
- $\Delta\phi(V, H_{\text{cand}})$: azimuthal distance between the reconstructed vector boson and Higgs-boson candidates.
- $\text{bin}_{\text{DL1r}}(j_1), \text{bin}_{\text{DL1r}}(j_2)$: PCFT bins of the two jets forming H_{cand} .
- $\sum_{i \neq 1, 2} p_T^{j_i}$: sum of the transverse momenta of jets not used to reconstruct H_{cand} .

⁸As a reminder, before the event selection described in Sec. 6.5, events are reweighted to the Run-2 luminosity times the most accurate theoretical prediction of the process cross-section.

- **0L channel specific variables:**

- $\Delta\eta(j_1, j_2)$: pseudorapidity distance between the two jets forming H_{cand} .
- $\min\{\Delta R(j_i, j)\}$, $i = 1, 2$: closest angular distance between a jet forming H_{cand} and an additional jet.
- m_{eff} : scalar sum of the $E_{\text{T}}^{\text{miss}}$ and the p_{T} of all the jets in the event.

- **1L channel specific variables:**

- m_{T}^W : transverse mass of the reconstructed W boson candidate.
- $\Delta y(V, H_{\text{cand}})$: rapidity distance between the vector boson and Higgs-boson candidates.
- $\min\{\Delta R(j_i, j)\}$, $i = 1, 2$: closest angular distance between a jet forming H_{cand} and an additional jet.
- $\min[\Delta\phi(l, j_i)]$, $i = 1, 2$: closest ϕ distance between a jet forming H_{cand} and an additional jet.
- m_{top} : reconstructed mass of the leptonically decaying top quark.

- **2L channel specific variables:**

- $m_{\ell\ell}$: invariant mass of the 2 leptons system.
- $\cos\theta(\ell^-, \vec{Z})$: Z boson polarization sensitive angle [152].
- $E_{\text{T}}^{\text{miss}}/\sqrt{S_{\text{T}}}$: ratio of $E_{\text{T}}^{\text{miss}}$ over the square root of S_{T} , the scalar sum of the p_{T} of the leptons and jets in the event.
- $\Delta y(V, H_{\text{cand}})$: rapidity distance between the reconstructed vector boson and Higgs-boson candidates.

Boosted input variables

- p_{T}^V : transverse energy of the reconstructed vector boson candidate.
- $p_{\text{T}}^{j_1, \text{trk}}$, $p_{\text{T}}^{j_2, \text{trk}}$ and $p_{\text{T}}^{j_3, \text{trk}}$: transverse momenta of the track-jets inside H_{cand} . j_1 and j_2 refer to the b -tagged track-jets, and j_3 refers to the leading non- b -tagged track-jet.
- m_J : invariant mass of H_{cand} .
- $\Delta R(j_{1, \text{trk}}, j_{2, \text{trk}})$: angular distance between the two b -tagged track-jets.
- $N(\text{track-jets in } J)$: number of track-jets associated to H_{cand} .
- $N(\text{add. small } R\text{-jets})$: number of additional small- R jets that are not matched to the leading large- R jet. This happens when $\Delta R(\text{small-}R \text{ jet}, \text{large-}R \text{ jet}) > 1.0$.

- $\Delta\phi(V, H_{\text{cand}})$: azimuthal distance between the reconstructed vector boson and Higgs-boson candidates.
- Colour: a variable designed to exploit colour-sensitive differences in the substructures of large- R jets emerging from the products of gluon splittings and from decays of QCD singlets states such as the Higgs-boson [153].
- $\text{bin}_{\text{DL1r}}(j1, \text{trk}), \text{bin}_{\text{DL1r}}(j2, \text{trk})$: PCBT bin of the b -tagged track-jets.
- **1L channel specific variables**
 - $\Delta y(V, H_{\text{cand}})$: rapidity distance between the reconstructed vector boson and Higgs-boson candidates.
 - p_{T}^{ℓ} : transverse momentum of the lepton.
 - $(p_{\text{T}}^{\ell} - E_{\text{T}}^{\text{miss}})/p_{\text{T}}^W$: proxy for the p_{T} imbalance of the charged lepton and the neutrino of the W -boson.
- **2L channel specific variables**
 - $\Delta y(V, H_{\text{cand}})$: rapidity distance between the reconstructed vector boson and Higgs-boson candidates.
 - $\cos\theta(\ell^-, \vec{Z})$: Z boson polarization sensitive angle [152].

The distributions of the two most important variables in the BDTs training, namely the mass and the angular separation ΔR between the jets forming H_{cand} , are displayed in Fig. 6.27 for signal and background events of the 2-jet training regions with $p_{\text{T}}^V > 150 \text{ GeV}$ of the 0L channel. The corresponding distributions look very similar in the same categories of the 1 and 2L channels. The 2-jet category brings the best sensitivity in the final fit in all three lepton channels.

Training and output transformation The hyperparameters that shape the architectures of the BDTs and define their training campaigns are optimized for each model to maximise the *Asimov significance* S of their signal vs background separation [154] and to avoid overtraining. S is computed from the distribution of the BDT discriminant output according to the formula:

$$S = \sqrt{\sum_i 2 \left[(s_i + b_i) \ln \left(1 + \frac{s_i}{b_i} \right) - s_i \right]}, \quad (6.2)$$

where s_i and b_i indicate the number of simulated signal and background events in the i^{th} bin of the distribution⁹.

⁹The Asimov significance is computed on the BDT discriminant variable taking into account only statistical uncertainties. Thus, systematic effects are ignored.

Variable	$VH, H \rightarrow b\bar{b}/c\bar{c}$ Resolved			$VH, H \rightarrow b\bar{b}$ Boosted		
	0L	1L	2L	0L	1L	2L
$m_{j_1 j_2}$ or m_J	×	×	×	×	×	×
$m_{j_1 j_2 j_3}$	×	×	×			
$p_T^{j_1}$ or $p_T^{j_1, \text{trk}}$	×	×	×	×	×	×
$p_T^{j_2}$ or $p_T^{j_2, \text{trk}}$	×	×	×	×	×	×
$p_T^{j_3, \text{trk}}$				×	×	×
$\sum_{i \neq 1, 2} p_T^{j_i}$	×	×	×			
$\Delta R(\vec{j}_1, \vec{j}_2)$ or $\Delta R(j_{1, \text{trk}}, j_{2, \text{trk}})$	×	×	×	×	×	×
$ \Delta\eta(\vec{j}_1, \vec{j}_2) $	×					
$\text{bin}_{\text{DL1r}}(j_1)$	×	×	×	×	×	×
$\text{bin}_{\text{DL1r}}(j_2)$	×	×	×	×	×	×
p_T^V	$\equiv E_T^{\text{miss}}$	×	×	$\equiv E_T^{\text{miss}}$	×	×
E_T^{miss}	×	×		×	×	
$E_T^{\text{miss}}/\sqrt{S_T}$			×			
$ \Delta y(\vec{V}, H_{\text{cand}}^{\vec{}}) $		×	×		×	×
$ \Delta\phi(\vec{V}, H_{\text{cand}}^{\vec{}}) $	×	×	×	×	×	×
$\min[\Delta\phi(\vec{\ell}, \vec{j}_1 \text{ or } \vec{j}_2)]$		×				
m_{eff}	×					
m_T^W		×				
m_{top}		×				
$m_{\ell\ell}$			×			
$\cos\theta(\vec{\ell}^{\vec{}}, \vec{Z})$			×			×
$(p_T^{\ell_1} - E_T^{\text{miss}})/p_T^W$					×	
p_T^ℓ					×	
$N(\text{track-jets in } J)$				×	×	×
$N(\text{add. small R-jets})$				×	×	×
Colour				×	×	×
$\min\{\Delta R(j_i, j)\}$	×	×				

Table 6.9: Input variables used to train the BDTs employed in the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis to enhance signal vs background separation.

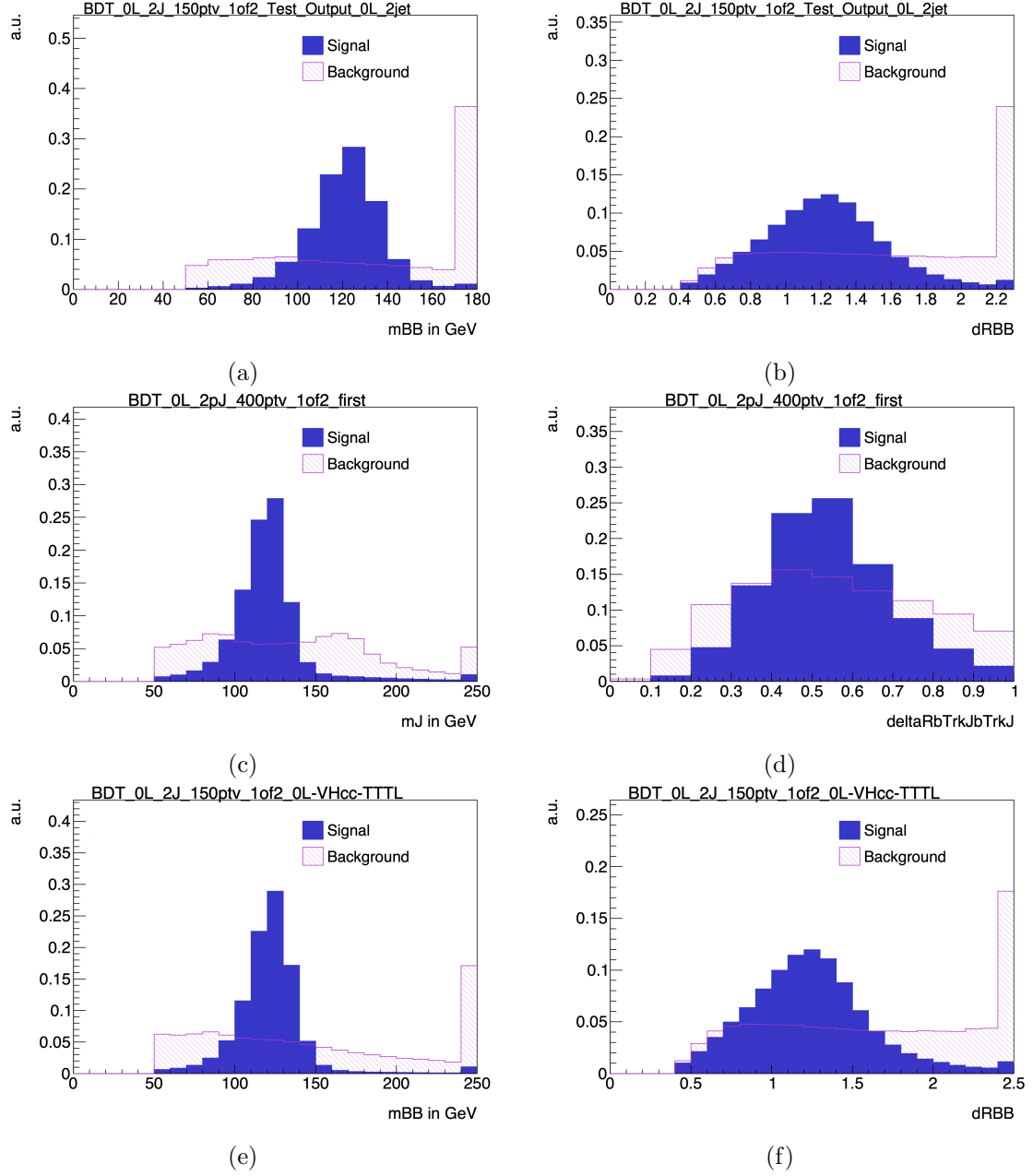


Figure 6.27: Normalized distributions of the mass and of the angular separation ΔR between the jets forming H_{cand} for 2-jet training regions with $p_T^V > 150 \text{ GeV}$ of the 0L channel of the resolved (a and b) and boosted (c and d) $VH, H \rightarrow b\bar{b}$ regimes and for the $VH, H \rightarrow c\bar{c}$ regime (e and f). Due to the long tails, the range of each variable is limited from above to include 99% of the signal events, a prescription which increases training stability. All events exceeding the range are assigned the maximum value, explaining the spikes in the upper bins of the distributions.

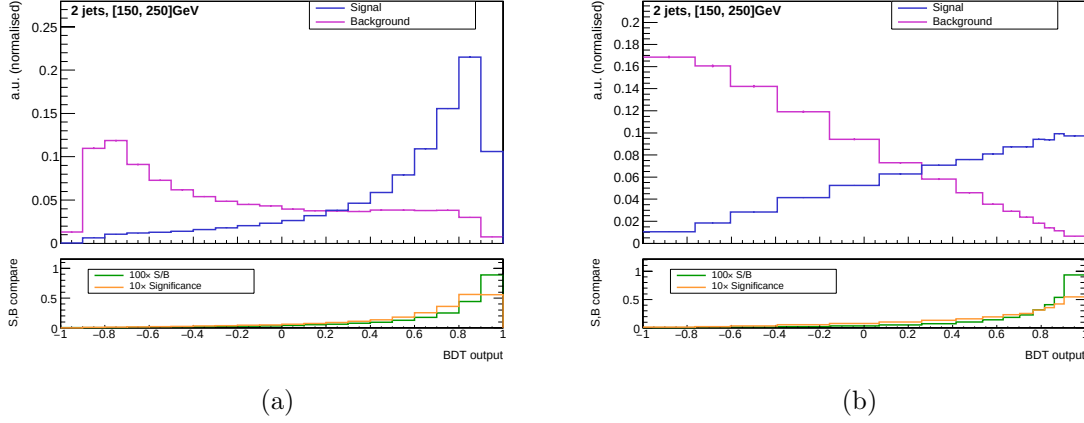


Figure 6.28: Comparison of the output distribution of the BDT in the $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ 2-jet, TT+TL region of the $VH, H \rightarrow c\bar{c}$ regime before (a) and after (b) transformation D. The binning obtained with the transformation is finer in the range with higher signal purity, and it makes the distributions smoother.

Before computing S , the distribution is rebinned using a transformation technique named *transformation D*. The default binning is equidistant and very fine: transformation D aims to optimise the bins' sensitivity by considering that low BDT output scores are predominantly associated with background events and lack signal events, while high BDT output scores exhibit a favourable signal-to-background ratio; and by factoring in the statistical uncertainty of each bin, which should be as low as possible. Starting from the highest bin, the Z quantity is computed:

$$Z = z_s \frac{s}{N_s} + z_b \frac{b}{N_b} \quad (6.3)$$

where N_s and N_b are the total number of signal and background events in the distribution, and z_s and z_b are tunable parameters. The bin is merged with its neighbours until Z exceeds 1¹⁰. This forms one bin in the transformed histogram, and the process restarts from the next unmerged bin. Tuning z_s and z_b modifies the number of signal-enriched and background-enriched bins obtained after transformation D. Fig. 6.28 illustrates the effect of transformation D in the $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ 2-jet, TT+TL region of the $VH, H \rightarrow c\bar{c}$ regime.

Trainings are performed using the k -fold method with $k = 2$: in each training region, events are equally divided into even and odd-numbered, and one BDT is trained per sub-set. Then, the BDT trained on the even-numbered events is evaluated on the odd-numbered, and vice-versa. To rule out overtraining, the distributions of BDT scores of both models are computed on both even and odd-numbered events and are monitored to ensure that they

¹⁰Protection measures are added to avoid bins with too low MC statistics or too few events once applied to real data.

are equivalent. The final templates of the BDT scores used in the fit are built by merging those comprising the even-numbered and odd-numbered events.

6.7.1 Data and MC BDT output distributions in the SRs

Fig. 6.29 shows the pre-fit BDT distributions in data and in MC samples belonging to the 2-jet 250 GeV < p_T^V < 400 GeV SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime. Fig. 6.30 displays those in the 2-jet 150 GeV < p_T^V < 250 GeV TL+TT SRs of the $VH, H \rightarrow c\bar{c}$ regimes; and Fig. 6.31 those in the 400 GeV < p_T^V < 600 GeV SRs of the boosted regime. The histograms are displayed with bins with equal width even though transformation D is applied, a choice that improves readability.

The plots provide valuable insights. They demonstrate the effectiveness of the BDT in separating signal from backgrounds by assigning higher scores to signal events and lower scores to background events. Additionally, they illustrate the intricate composition of background processes in the different SRs and lepton channels, whose contributions are controlled through the dedicated CRs as described in Sec. 6.5.3. The data/MC ratios in the lower panel of the plots indicate well-modelled physics processes in each region. The absence of significant shape effects in the ratios is noteworthy: while pre-fit data/MC ratios may have offsets, these can be easily absorbed in the fit due to the normalisation of main backgrounds varying freely; whereas complex dependencies of the ratio on the BDT output score must be characterized in terms of systematic effects.

6.8 Systematic uncertainties

Aside from the statistical uncertainty of the measurement, various systematic effects must be characterised and considered. They fall into two categories: experimental uncertainties and theoretical uncertainties related to the modelling of the signal and background processes. Experimental uncertainties are tied to the performance of the ATLAS detector, most noticeably in how it reconstructs, identifies, and calibrates physics objects. Modelling uncertainties are associated with assumptions made in simulating signal and background processes using MC methods, and with their cross-section calculations.

6.8.1 Experimental uncertainties

The leading experimental uncertainties in both the resolved and boosted $VH, H \rightarrow b\bar{b}$ regimes, as well as in $VH, H \rightarrow c\bar{c}$, are due to the imperfect calibration of the jet energies scales and resolutions.

Small- R jet energy scale uncertainties have contributions extracted from calibrations performed on well understood processes such as $Z \rightarrow e^+e^-$ decays, which allow to characterise

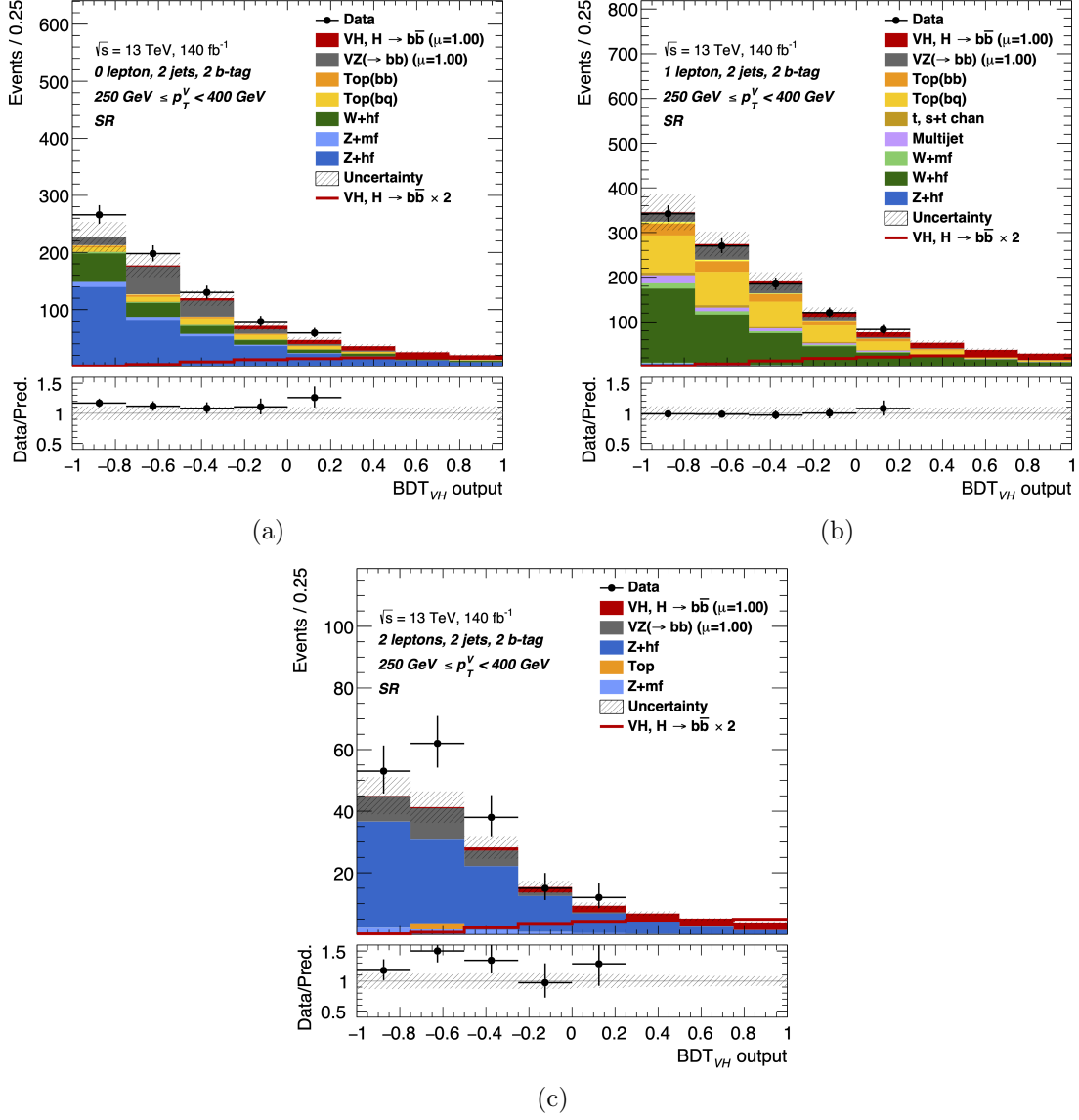


Figure 6.29: Data and prefit MC distributions of the output score of the BDT evaluated on events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ SR in the resolved $VH, H \rightarrow b\bar{b}$ regime.

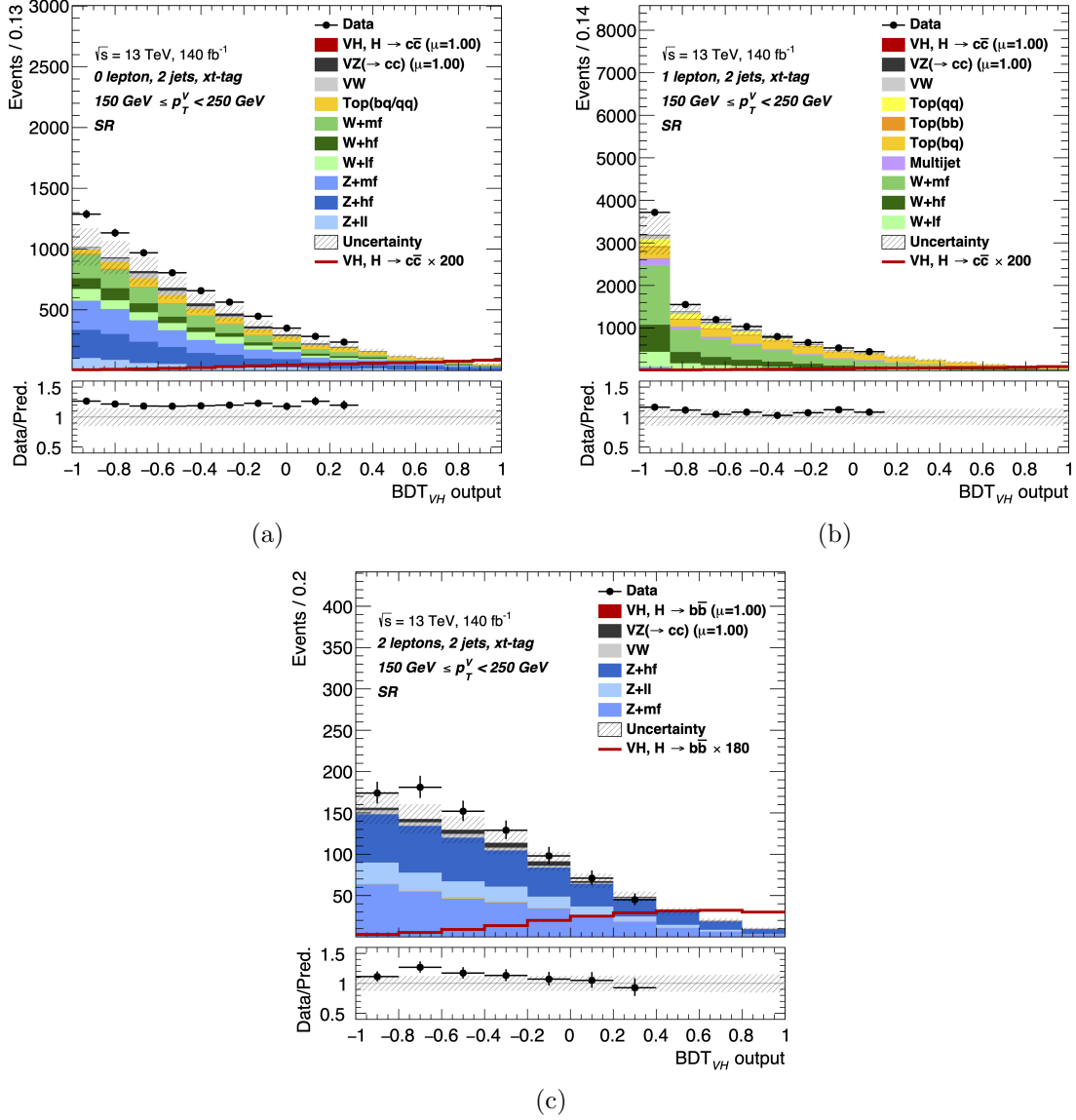


Figure 6.30: Data and prefit MC distributions of the output score of the BDT evaluated on events entering the 2-jet $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ SRs in the $VH, H \rightarrow c\bar{c}$ regime. *xt-tag*, a name used in the legend, indicates TT+TL *c*-tag regions.

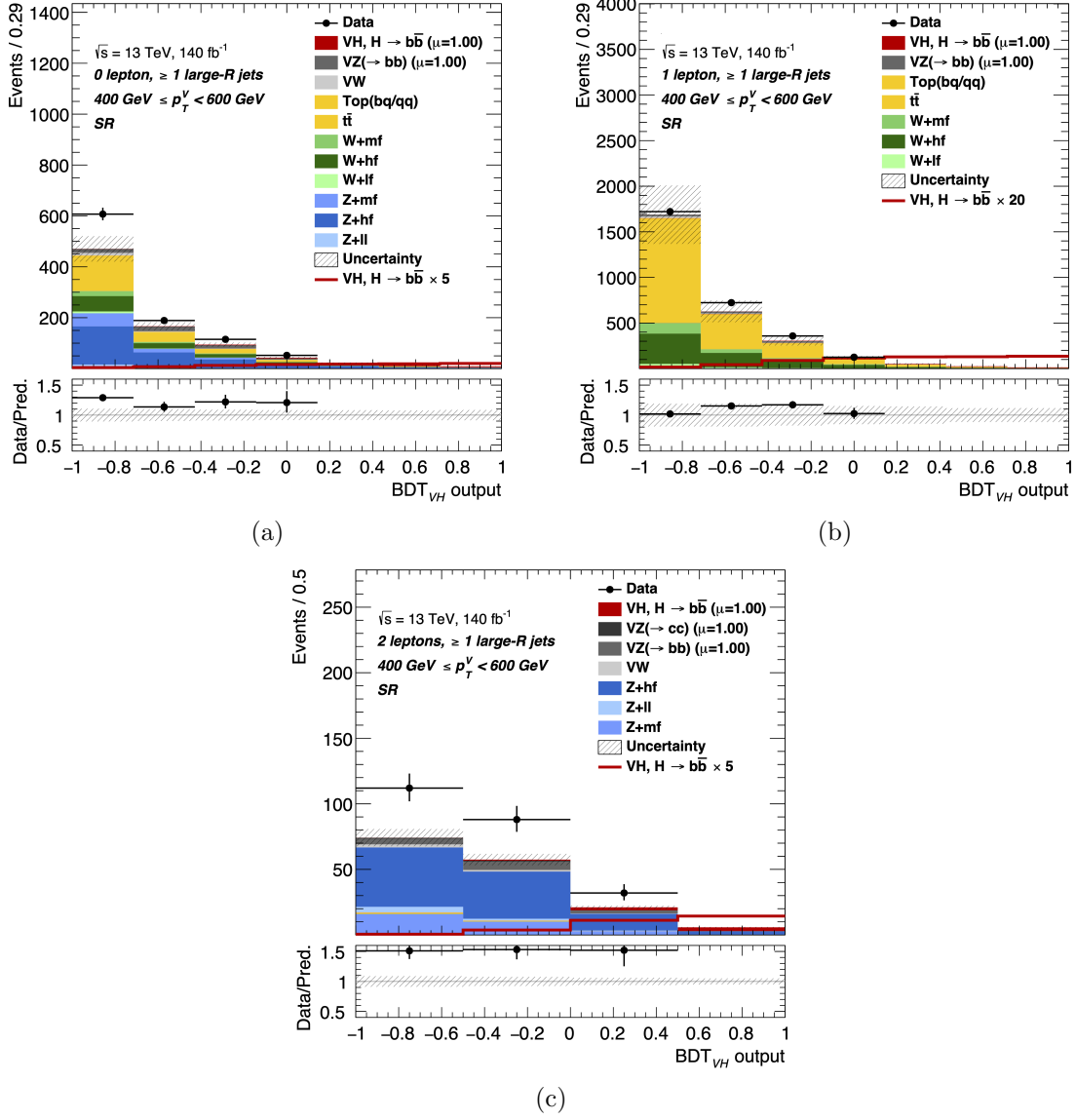


Figure 6.31: Data and prefit MC distributions of the output score of the BDT evaluated on events entering the SRs in the boosted $VH, H \rightarrow b\bar{b}$ regime.

variations in the detector response when dealing with jets with different p_T and η , with events with different pile-up activities, with different flavour-composition of the jets, or with boosted hadrons that do not deposit all their energies in the calorimeters.

Uncertainties on the energy resolution of the jets, extracted via calibrations leveraging the p_T asymmetries measured in di-jet events, arise from mismodelling in MC events used to estimate it and from noise in the detector.

Large- R jet energy scale and resolution uncertainties arise from similar effects. Additional contributions take into account the topology of the event in which the large- R jet is produced, as the modelling of the response for large- R jets stemming from prompt light-quarks, gluons, or from hadrons produced in decays of heavier particles such as the Higgs-boson can vary in different generators. Furthermore, uncertainties relative to the determination of the large- R jet mass scale and resolutions are taken into account [155].

Uncertainties in the data-MC flavour-tagging efficiency SFs have a comparable impact. These SFs are determined using calibration data in various kinematic regions, separately considering b , c , and light-flavour jets (see Chapter 4). Since no data-MC SF is evaluated for flavour-tagging τ jets, SF_c is applied also to τ jets, given their similar flavour-tagging efficiencies in the MC samples. An additional uncertainty is introduced to handle the extrapolation to jets beyond the kinematic range covered by the data calibration.

Smaller impact experimental systematic uncertainties include those in lepton energy and momentum scales, lepton reconstruction and identification efficiency, and trigger efficiency. Uncertainties in the energy scale and resolution of track-jets, small- R jets and leptons are considered in calculating E_T^{miss} , with additional uncertainties stemming from the scale, resolution, and reconstruction efficiency of tracks used to compute the soft term, along with the modelling of the underlying event [156].

6.8.2 Modelling uncertainties

The MC simulation packages employed to model signal and background processes in the analysis phase space and the cross-section calculations for normalizing the samples are based on several assumptions. These, as anticipated in Sec. 2.5, include choices such as the QCD renormalization and factorization scales, the PDFs describing the distribution of momentum among the quarks and gluons inside the proton, the NLO+PS algorithm connecting the colour charges produced at the matrix element level with those described by the parton shower algorithms, the parton shower and hadronization model, the particles that the final state particles can emit, and experimental input parameters like the strong coupling constant α_S or the mass of the b -quark m_b . These assumptions introduce modelling uncertainties that must be assessed. This is accomplished by varying the internal parameters of the nominal simulation controlling these choices, and by comparing the nominal simulation results with those obtained using an alternative generator setup. The modelling uncertainties affecting this analysis can be categorized into three groups.

Normalisation uncertainties The normalisations of the main backgrounds, V +jets and top events, rely on data in the dedicated CRs. This is achieved by allowing them to float freely in the likelihood fit, without any constrain from auxiliary measurements, as described in Sec.6.10. The floating normalisations are split by p_T^V and jet multiplicity as much as the data statistics allows. Minor backgrounds' normalisations, instead, are scaled to the SM prediction and assigned the uncertainty on the theory cross-section prediction coherently over all the analysis regions.

Relative acceptance uncertainties Nominal and alternative MC generators may distribute events across analysis regions in distinct ways. This leads to a variation in the event ratios among regions. Relative acceptance uncertainties parameterise the uncertainty regarding how the probability of accepting events of a process varies across different regions. In simpler terms, they represent the uncertainty on the ratio of the number of accepted events of a given process in different regions of the analysis. They are calculated by comparing the relative acceptance between two regions A and B with the nominal and the alternative generators, via the double ratio formula:

$$\text{Acceptance unc}_{var_i} = \frac{(N_B/N_A)_{var_i}}{(N_B/N_A)_{nom}} - 1 \quad (6.4)$$

Here, N_A and N_B represent the number of selected events in regions A and B, respectively. Region A, known as the *anchor*, has the highest purity or statistics. The uncertainty is applied to the normalisation of the process in region B, and the use of the double ratio prevents double-counting with normalisation uncertainties. If the statistical error surpasses the relative acceptance uncertainty, a conservative approach is taken, and the statistical error is used as the relative acceptance uncertainty. The total relative acceptance uncertainty is the pythagorean addition of all considered variations:

$$\text{Total acceptance} = \sqrt{\sum_i (\text{Acceptance unc}_{var_i})^2} \quad (6.5)$$

The general strategy to account for relative acceptance uncertainties in the fit involves deriving uncertainties across regions and for processes that share a common normalisation factor ¹¹. If three regions A, B and C share the same normalisation factor, two uncertainties are derived. Let A be the anchor region, B the intermediate, and C the region with less purity or statistics: the relative acceptance uncertainty is first computed from the anchor region to the other two (*i.e.*, $A \rightarrow B+C$), and then from the intermediate to the remaining region (*i.e.*, $B \rightarrow C$).

¹¹Relative acceptance uncertainties might be confusing at first. Explaining their purpose can guide their understanding. In the fitting process, the variation of a normalisation factor shared among regions is mainly constrained in the anchor region. The strength of this constraint depends on how signals and backgrounds are divided across the different regions. This division may change when simulating events with different generators. Relative acceptance uncertainties provide an extra degree of freedom to the MC templates in the regions anchored by A, which accounts for the uncertainty on how the events are distributed. This mitigates the extrapolation of the constraining power of the anchor region to those linked to it in the fit.

In the analysis, normalisation factors are shared between many types of regions and processes. The corresponding relative acceptance uncertainties account for these correlations across different p_T^V and jet multiplicity regions, lepton channels, samples with different flavour compositions of the events, Wt and $t\bar{t}$ samples floating together, SRs and CRs.

Shape uncertainties from CARL and other sources Shape systematic uncertainties quantify how the shapes of the distributions used in the fit change when comparing different MC samples.

Shape uncertainties for most samples are derived using the CARL algorithm, a reweighting method based on a NN. The NN is trained to distinguish between nominal and alternative samples using the same kinematic variables as the BDT described in Sec. 6.7. It assigns a probability to each event, representing its estimated likelihood of belonging to the alternative sample. Applying this probability as an event weight to the nominal sample reweights it to mimic the shape of the alternative sample. Since nominal samples are simulated with much larger statistics than alternative ones, this process helps smooth out statistical fluctuations in the distribution obtained with the alternative generators. The CARL-reweighted sample, normalized to the nominal's yield, is then used to compute the envelope in which the nominal MC template can vary during the fit, according to the interpolation procedure described in Sec. 6.10.2.

Fig. 6.32 compares the distribution of the BDT discriminant variable computed on 2L $Z+b$ -jets events with $p_T^V > 400$ GeV selected in the SR of the $VH, H \rightarrow b\bar{b}$ boosted regime; simulated with the nominal SHERPA 2.2.11 generator, with the alternative MADGRAPH FxFx generator (described in Sec. 6.9), and with the CARL-reweighted sample.

The top panel displays the distributions. It shows the distribution obtained with the CARL-reweighted sample closing nicely on that obtained with the alternative generator. The statistical uncertainties on the overall yields of the distributions are listed in the legend. The uncertainties of the nominal and of the CARL-reweighted distributions are equal and smaller than the alternative's.

The medium panel displays the ratios among the distributions. The good performance of the CARL method is reflected by the ratio of the CARL-reweighted over the alternative distribution being close to unity everywhere, aside from some deviations in the poorly populated tails; and by the ratios of the alternative over the nominal sample and of the CARL-reweighted over the nominal sample being very similar.

The bottom panel shows the residuals of the nominal vs alternative and of the CARL-reweighted vs alternative distributions, defined as $\frac{N-A}{\sqrt{\sigma(N)^2 + \sigma(A)^2}}$. The residual of the CARL-reweighted vs alternative distribution is close to zero, and within a 3σ variation everywhere.

The distance between the alternative and CARL-reweighted distribution is quantified via a χ^2 test and a Kolmogorov-Smirnov (K-S) test [157]. These tests are monitored in the hundreds of regions where the CARL method is used. Good performance is indicated by

χ^2 being close to unity and K-S being close to zero.

Besides CARL, shape uncertainties are also included based on the internal weight variations within the nominal samples. In this scenario, systematic variations have the same statistics as the nominal samples.

A detailed description of the modelling uncertainties for the V +jets background is provided in the following section, as I was personally involved in their derivation. Equally careful treatments were reserved for the signal and other background processes. The exact simulation setup differs from process to process. Therefore, the sources of uncertainties can vary. Nevertheless, the guiding principles to determine the normalisation, relative acceptance and shape uncertainties are the same. For the sake of brevity, this work does not delve into further details about their derivation. Tab. 6.10 summarizes all the normalisation and relative acceptance uncertainties applied in the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis.

6.9 Modelling V +jets events

In the previous sections, it was shown how V +jets events represent one of the main backgrounds in all the lepton channels. These events were simulated with the latest MC generators available on the market, introducing updates with respect to the previous $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ analyses. The most noteworthy involve using an alternative generator capable of simulating part of the events at NLO precision in the strong coupling constant α_s and introducing improved statistical enhancement techniques to populate low cross-section regions of the phase space.

As anticipated in the Introduction, I contributed to the studies described in this section in multiple ways.

First, I played a crucial role in kick-starting the code that is used to produce all the plots, modulo those with captions citing other published works. Such code was widely used for modelling also the other types of backgrounds and to compute relative acceptance uncertainties.

I used the code to perform various studies comparing the nominal SHERPA 2.2.11 sample with its predecessor SHERPA 2.2.1, focusing on Z + jets events in the $VH, H \rightarrow b\bar{b}$ regions. These studies proved crucial for assessing the statistics needed to model this type of background accurately. They resulted in interactions with the ATLAS Physics Modelling Group to request additional sample productions to gather sufficient statistics. Eventually, the SHERPA 2.2.11 sample largely surpassed the statistics of SHERPA 2.2.1 in this analysis phase space.

Together with the pre-fit data-MC closure studies that I performed on 2L events in the $VH, H \rightarrow b\bar{b}$ regions, these studies also helped quantify the issue encountered by the

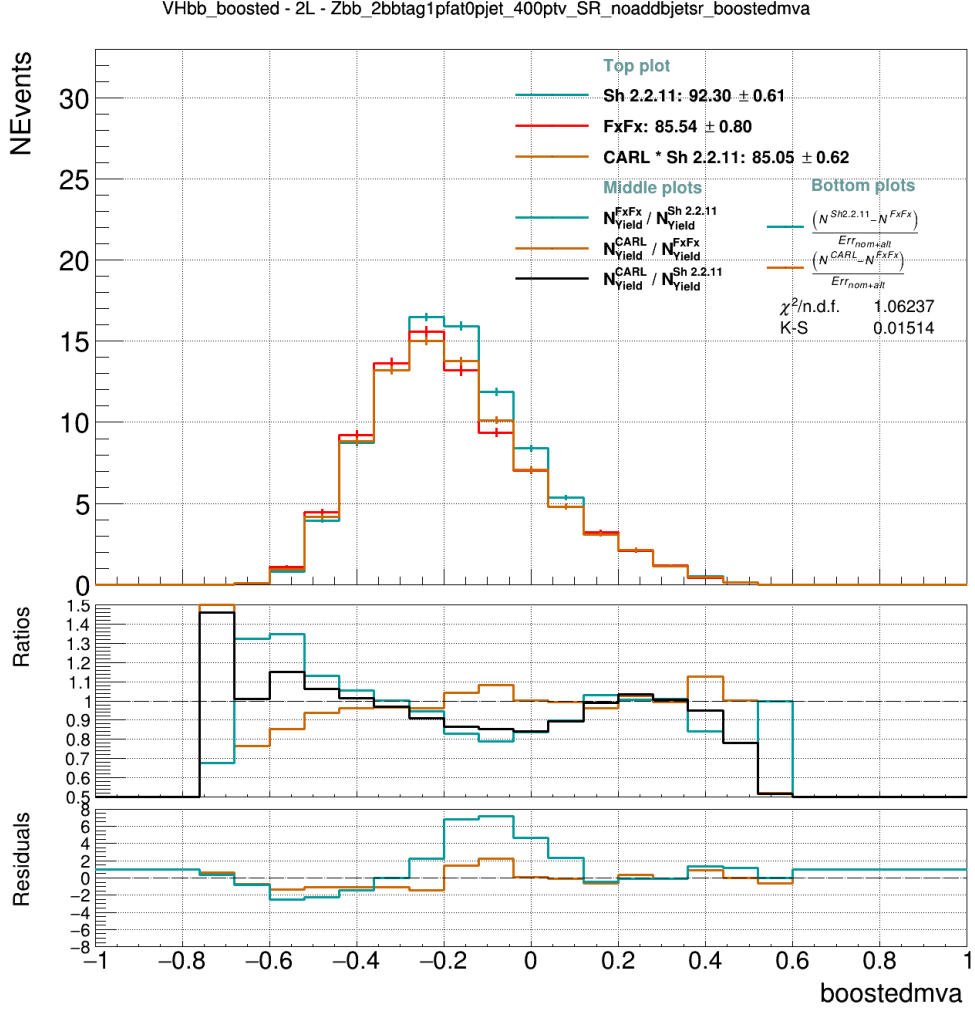


Figure 6.32: Distribution of the BDT discriminant variable computed on 2L $Z+b$ -jets events with $p_T^V > 400$ GeV selected in the SR of the $VH, H \rightarrow b\bar{b}$ boosted regime; simulated with the nominal SHERPA 2.2.11 generator, with the alternative MADGRAPH FxFx generator, and with the CARL-reweighted sample. The top panel displays the distributions, the medium panel the ratios between the distributions, and the bottom panel the residuals between the nominal vs alternative and between the CARL-reweighted vs alternative distributions.

Uncertainties	resolved $VH, H \rightarrow b\bar{b}/c\bar{c}$	boosted $VH, H \rightarrow b\bar{b}$
VH		
$WH/q\bar{q}ZH/ggZH, H \rightarrow c\bar{c}$ normalisation/acceptance	6.2%, 5.8%, 31%	
$H \rightarrow b\bar{b}$ BR	1.61%	
$H \rightarrow c\bar{c}$ BR	+5.53% -1.99%	
Diboson		
$WW/ZZ/WZ$ normalisation	16%/17%/19%	16%/17%/27%
$ggVV$ normalisation	30%	
Lepton channel acceptance	2-23%	7%
N_{jet} acceptance	10%-30%	-
p_{T}^V acceptance	3%-16%	8%-40%
SR/CR acceptance	6%-16%	-
Z+jets		
$Z + hf$ normalisation	Floating	
$Z + mf$ normalisation	Floating	35%
$Z + lf$ normalisation	Floating	35%
$Z + hf$ flavour composition ratio	8%-12%	6%-9%
$Z + mf$ flavour composition ratios	4%-10%	6%-9%
High- ΔR CR to SR	5%-30%	-
topCR-SR extrapolation ratio	-	15%-25%
2- to 0L acceptance ratio	2%-10%	3%
p_{T}^V extrapolation	-	15%
W+jets		
$W + hf$ normalisation	Floating	
$W + mf$ normalisation	Floating	36%
$W + lf$ normalisation	Floating	38%
$W + hf$ flavour composition ratios	4%-25%	11%
$W + mf$ flavour composition ratios	14-29%	9%-15%
$W + lf$ flavour composition ratios	9%	-
High/Low- ΔR CR-SR extrapolation ratios	2-63%	-
topCR-SR extrapolation ratio	-	16%-27%
1- to 0L acceptance ratio	3%-30%	20%
p_{T}^V extrapolation	-	3%
N_{jet} extrapolation	12-20%	-
Top (0 and 1L resolved, no single top s/t)		
Top($b\bar{b}$) normalisation	Floating	-
top(others) normalisation	Floating	-
Flavour acceptance ratios	5% - 10%	-
1- to 0L acceptance ratio	2% - 8%	-
High/Low- ΔR CR-SR extrapolation ratios	2% - 10%	-
$Wt / t\bar{t}$ ratio	12-48%	-
Top (2L resolved)		
Normalisation	Floating ($VH, H \rightarrow c\bar{c}$), 0.08% ($VH, H \rightarrow b\bar{b}$)	-
Single-top s-/t-channel (0L and 1L for resolved)		
Normalisation	4-10%	-
High/Low- ΔR CR-SR extrapolation ratios	3% - 17%	-
p_{T}^V extrapolation ratios	7% - 15%	-
N_{jet} acceptance	15%	-
1- to 0L acceptance ratio	6%	-
$t\bar{t}$ and single-top (for boosted regime)		
$t\bar{t}$ normalisation	-	Floating
single-top normalisation	-	4.6%-25%
1- to 0L acceptance ratio	-	40%
topCR-SR acceptance ratio	-	20%
single-top p_{T}^V extrapolation	-	20%
Multi-jet (1L)*		
Normalisation	20%-100%	

Table 6.10: Summary of the normalisation and relative acceptance uncertainties considered in the analysis. The values given refer to the size of the uncertainty affecting the yield of each sample.

SHERPA 2.2.11 generator in modelling events with high p_T^V values, detailed in Sec. 6.9.3.

I used the code to compare the nominal SHERPA 2.2.11 samples with the alternative MADGRAPHxFx sample, providing valuable insights on regions associated with large relative acceptance uncertainties in the $VH, H \rightarrow b\bar{b}$ boosted regime. These regions were eventually merged with others, more populated and better-modelled, to reduce their impact on the stability of the fit.

Finally, I derived multiple times preliminary numbers for the relative acceptance uncertainties affecting the $Z + \text{jets}$ events in the $VH, H \rightarrow b\bar{b}$ regions, as described in Sec. 6.9.2¹².

6.9.1 Nominal and alternative MC samples

The nominal MC sample Nominal $V + \text{jets}$ events are simulated with the SHERPA 2.2.11 generator [149], interfaced with the Hessian NNPDF3.0NNLO [158, 159] PDF set to describe the partonic structure of the colliding protons. The following setup is used:

- b and c -quarks are treated as massless at the matrix-element level and massive in the parton shower;
- $m_{\ell\ell} > 40 \text{ GeV}$ in $Z + \text{jets}$ events;
- $m_{l\nu} > 20 \text{ GeV}$ in $W + \text{jets}$ events;
- $Z + \text{jets}$ electroweak input scheme is determined by m_Z , the QED coupling constant at the Z -boson resonance $\alpha(m_Z)$, and the effective weak mixing angle $\sin^2(\theta_{\text{eff}})$;
- $W + \text{jets}$ electroweak input scheme is determined by the Fermi constant G_F , m_Z , and m_W ;
- samples are corrected to an approximate NLO electroweak accuracy by using weights generated internally in the SHERPA samples, further described later in this section.

The matrix elements are calculated at NLO accuracy in α_s for events with up to two jets and at LO for those with up to five jets using the Comix [160] and OPENLOOPS [161] libraries. Events are further processed by the SHERPA internal parton shower algorithm based on Catani-Seymour dipole factorisation and the cluster hadronisation model [162, 163]. The algorithm employs a dedicated set of tuned parameters optimized by the SHERPA authors to operate with the NNPDF3.0NNLO PDF set.

Simulating events using matrix-elements computations and parton-shower algorithms requires a procedure to avoid double-counting. For instance, a $V + 2 \text{ jets}$ event such as that schematized in Fig. 6.12a is directly accounted for in the matrix-element computation, but

¹²The final numbers, computed by two colleagues, reflect the final configuration of the analysis categories, slightly updated while I was writing this thesis.

it can also arise from a $V+1$ gluon-jet event, computed at the matrix-element level, where the gluon subsequently emits the quarks during the parton-showering step¹³.

To deal with this, the NLO matrix elements of a given jet multiplicity are matched to the parton shower using a colour-exact variant of the MC@NLO algorithm [164]. Different jet multiplicities are then merged into an inclusive sample using an improved CKKW matching procedure [165, 166], which is extended to NLO accuracy using the MEPS@NLO prescription [167], with a merging cut set to $Q_{\text{cut}} = 20$ GeV. The MEPS@NLO prescription matches fixed-order perturbative calculations to parton showers by probabilistically identifying parton-shower topologies through a k_T -type clustering algorithm [166]. In this approach, the first or hardest scale at which the parton shower starts must be identified. This is done through a backwards declustering process, which assumes that the k_T scale diminishes at every subsequent branching. Cases exist in extreme regions of the phase space where this assumed hierarchy is violated. The violation results in the erroneous omission of Sudakov-type suppression factors in the algorithm and may lead to large enhancements of the differential cross-sections and of their uncertainties. In SHERPA 2.2.11, this problem is solved by vetoing unordered parton-shower histories. If an unordered emission is found, the previous maximal scale is used as described in Sec. 3.4 of [168].

The $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis includes regions of the phase spaces corresponding to high p_T^V values and to large jet multiplicities, where the V +jets production cross-section is small; therefore, it is important to ensure that statistical fluctuations of the background prediction are small compared to those in data, to be able fully to exploit the sensitivity of these regions. This is achieved by producing the V +jets samples using a continuous phase-space biasing technique that enhances analytically the $\max(H_T, p_T^V)$ variable, where H_T is defined as the scalar sum of all the transverse momenta of the partonic jets. Fig. 6.33 shows the effect of this enhancement. Simulating the events using the cross-section predicted by the SM would provide the distribution shown with the red line. The analytic enhancement modifies the cross-section to better populate the regions corresponding to large values of p_T^V or H_T . The samples are then normalized to the QCD NNLO-accurate theory predictions of single boson production cross-sections [169].

In the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis, final states contain flavour-tagged jets; therefore, it is important to enhance the statistics of simulated V + heavy-flavour jets events to provide a robust MC prediction to compare to data. To achieve this goal while saving time for production¹⁴ V +jets samples are filtered at the particle level to increase the relative ratio of V +jets events with heavy flavour jets. For similar reasons, kinematic filters are used in

¹³The two ways of simulating the events differ in accuracy. Quoting from [149]: “Consider a well-defined inclusive n -jet type observable. A tree-level calculation at $\mathcal{O}(\alpha_s)$ will typically provide the lowest-order prediction. Subsequent emissions from a parton shower provide a (leading) logarithmic approximation for the higher jet rates, preserving the leading-order n -jet rate. In contrast, an exact NLO QCD calculation, *i.e.* including the virtual and real corrections, yields an NLO accurate prediction for the n -jet cross-section, while the $(n+1)$ jet rate is approximated to leading order”.

¹⁴The production of the SHERPA 2.2.11 samples required more than a year to complete, notwithstanding the large amount of CPU resources dedicated to the task by the ATLAS Collaboration.

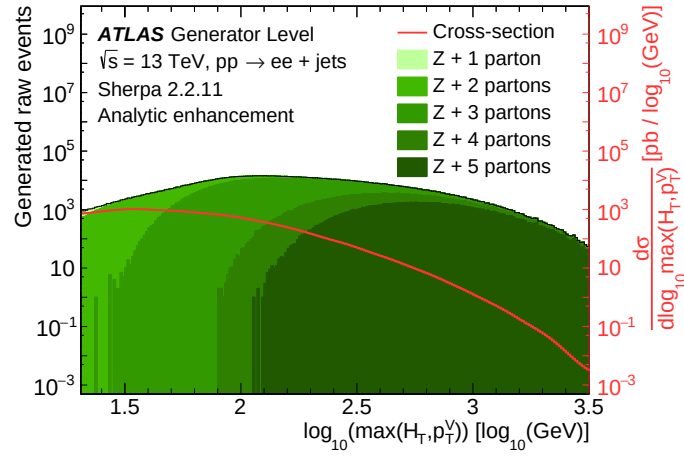


Figure 6.33: Distribution of unweighted MC sampled events showing the analytic enhancement as a function of $\log_{10}(\max(H_T, p_T^V))$ used to improve the simulated statistics of events with large p_T^V values in the SHERPA 2.2.11 sample. Simulating the events using the cross-section predicted by the SM would provide the distribution shown by the red line. The analytic enhancement successfully modifies the cross-section to better populate the regions corresponding to large values of p_T^V or H_T . The samples are then normalized to the QCD NNLO-accurate theory predictions of single boson production cross-sections. From. [149].

Configuration	SHERPA2.2.1	SHERPA2.2.11
PDF set	NNPDF3.0NNLO	NNPDF3.0NNLO
EW input scheme	Effective	$\sin^2\theta_{\text{eff}}$
QCD accuracy	0-2j@NLO+3,4j@LO	0-2j@NLO+3,4,5j@LO
NLO EW _{virt} corrections	No	Yes
Subtraction scheme	Default Catani–Seymour	Modified Catani–Seymour
Unordered histories allowed	Yes	No
Phase-space strategy	Sliced in $\max(H_T, p_T^V)$	Analytic enhancement

Table 6.11: Summary of the SHERPA2.2.1 and SHERPA 2.2.11 configurations.

the $W \rightarrow \tau_{\text{had}}\nu$ samples to increase the relative ratio of events with high E_T^{miss} .

Sherpa 2.2.11 vs Sherpa 2.2.1 $VH, H \rightarrow b\bar{b}/c\bar{c}$ is the first ATLAS analysis to use the SHERPA 2.2.11 samples. The previous $VH, H \rightarrow c\bar{c}$ and $VH, H \rightarrow b\bar{b}$ analyses relied on its predecessor SHERPA 2.2.1. A summary of the setup of the most relevant parameters of the two SHERPA versions from the standpoint of this analysis is provided in Tab. 6.11 [149].

The difference between the samples can be appreciated in the distributions displayed in Fig. 6.34, which shows the p_T^V and m_{CC} distributions of 1L W +jets events in the $VH, H \rightarrow c\bar{c}$ regime. The two generators model the mass peak slightly differently, and the SHERPA 2.2.11 sample has a softer p_T^V spectrum. Furthermore, statistical uncertainty is reduced by 40 – 50% in the SHERPA 2.2.11 sample, an effect of the analytic enhancement and of the greater number of simulated events. This results in smoother distributions in the less populated regions of the phase space, as exemplified in Fig. 6.35.

Differences in the modelling of the kinematic distributions result in differences in the discriminant score of the BDT, the variable used during the fit. These differences, exemplified in Fig. 6.36 were monitored carefully across all the analysis’ regions.

The alternative MC samples Alternative samples serve multiple purposes for the analysis. Firstly, they are used for estimating the modelling uncertainties associated with the nominal prediction. Secondly, they provide valuable cross-checks for the nominal prediction during the analysis design. Comparing the nominal and alternative predictions across different variables and phase space regions offers insights into the factors influencing the final state features; and ensures that the analysis strategy is not fine-tuned based on the nominal sample, in the potential cases where the alternative prediction deviates from the nominal. Additionally, these deviations can hint at possible discrepancies between observed and simulated data, and help identify areas for improvement in the simulation.

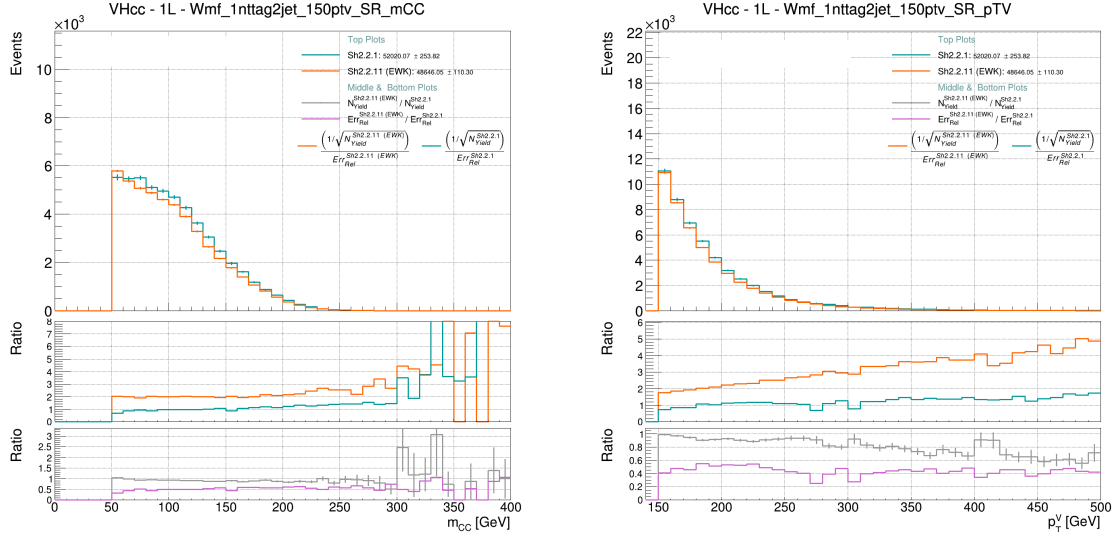


Figure 6.34: Comparison of the m_{CC} (a) and p_T^V (b) distributions of $W+2$ jets events belonging to the TN-tag SR region of the $VH, H \rightarrow c\bar{c}$ regime, obtained using the SHERPA 2.2.11 and SHERPA 2.2.1 samples. SHERPA 2.2.11 exhibit a slight shift in the mass peak and a softer p_T^V spectrum. The bottom panels show in grey the ratio of the yields of the two histograms, and the ratio of their relative uncertainty in magenta. The latter shows a reduction of $\sim 50\%$ in the relative error associated to the SHERPA 2.2.11 distributions. The middle panel displays the ratio between the uncertainty that would be associated to the distribution if it were obtained with unweighted events, and the real uncertainty that takes into account the weights.

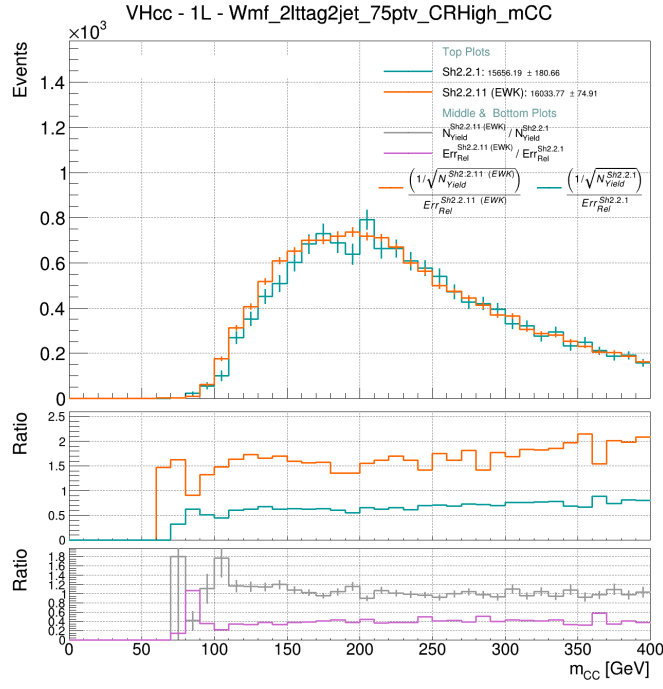


Figure 6.35: m_{CC} distribution of $W+2\text{jets}$ events belonging to the low- p_T^V High- ΔR CR region of the $VH, H \rightarrow c\bar{c}$ regime. The distribution obtained using the SHERPA 2.2.11 sample is smoother, and has smaller relative uncertainty, thanks to the analytic enhancement and to the improved statistics.

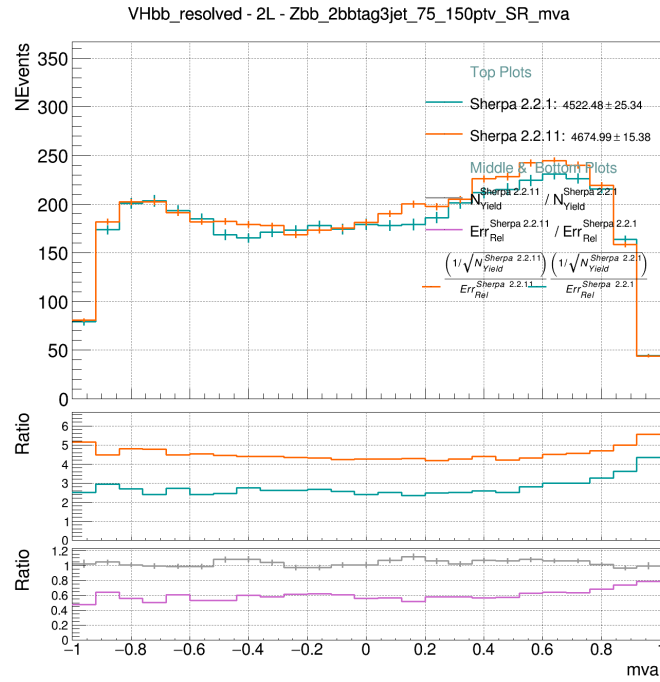


Figure 6.36: Comparison of the discriminant variable obtained when training the BDT with the SHERPA 2.2.11 and SHERPA 2.2.1 $V + jets$ samples.

Alternative V +jets events are simulated with the so-called MADGRAPH FxFx setup, that uses a MADGRAPH5_aMC@NLO 2.6.5 generator [170], interfaced with the NNPDF3.1NNLO PDF set supplemented with the LUXqed photon PDFs using $\alpha_s = 0.118$ [171]. The matrix elements are calculated at NLO accuracy in α_s for events with up to three jets. The parton shower and subsequent hadronization is performed using the PYTHIA 8.240 algorithm [172], with the A14 set of tuned parameters [135] interfaced with the NNPDF2.3LO PDF set [173] with $\alpha_s = 0.130$, and where the factorization and renormalization scales are set to half the final state partons' transverse mass. The different jet multiplicities are merged using the FxFx NLO matrix-element and parton-shower merging prescription [174], with a merging cut set at $Q_{\text{cut}} = 20 \text{ GeV}$. Additionally:

- QED radiation is added by PYTHIA 8;
- b and c -quarks are treated as massless at the matrix-element level and as massive during showering;
- final state quarks are required to have p_T larger than 10 GeV at the matrix element level.

To enable proper comparison, MADGRAPH FxFx samples are normalized to the same cross-sections as the nominal SHERPA 2.2.11 V + jets samples and are produced using the same heavy-flavour jet filtering strategy.

Fig. 6.37 compares the p_T^V distribution of $Z + 3$ jets event, where the two tagged jets are b -jets, simulated with the SHERPA 2.2.11 and MADGRAPH[FxFx] generators in the SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime. Albeit the total yields of the distributions differ by $\sim 20\%$, no shape is visible in their ratio. Analogous plots were monitored in all the regions to gain a complete picture of the different modellings achieved by the two generators throughout the analysis phase space. Particular attention was dedicated to the modelling of the p_T^V distributions since the nominal prediction underestimates the data at large p_T^V values. This problem, which also pertains to the alternative MADGRAPH FxFx generator but which is mitigated in the less precise SHERPA 2.2.1 samples, is discussed in Sec. 6.9.3. Other regions that were carefully monitored are those associated with large acceptance uncertainties.

In addition to the MADGRAPH FxFx sample, alternative samples produced with different setups of the SHERPA 2.2.11 generator are considered by adding internal weights to the events. These samples are used to estimate the QCD scale systematic variations, the PDF+ α_s uncertainties, and the effects of NLO EWK corrections.

QCD scale variations QCD scale variations are used to assess uncertainties related to the choice of the renormalization scale μ_r and of the factorisation scale μ_f . The scales are varied by a factor of 2 up and down coherently in the matrix elements and the parton

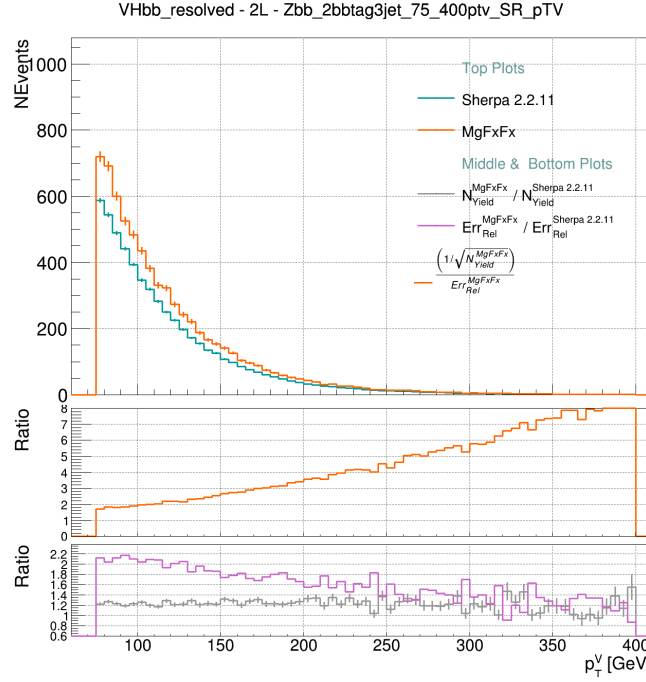


Figure 6.37: p_T^V distribution of Zbb events in the 3-jet SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime, as modelled by the SHERPA 2.2.11 and MADGRAPH FxFx generators. The greater statistics of the SHERPA 2.2.11 sample is reflected in the ratios between the generators' relative errors (in magenta in the bottom pad), larger than 40 % in the full spectrum. No shape effect can be seen in the ratio between the yields of the histograms (in grey in the bottom panel).

shower¹⁵ [175], vetoing the combinations involving opposite directions (*i.e.*, $2\mu_r$ with $1/2\mu_f$ or vice-versa). The QCD scale uncertainty is defined as the maximum shift from the nominal value of the envelope of the variations. Figure 6.38 provides an example of the effects of the QCD scale variations by showing the p_T^V distribution of $Z + 3$ jets events, where the two b -tagged jets are b -jets, in the SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime.

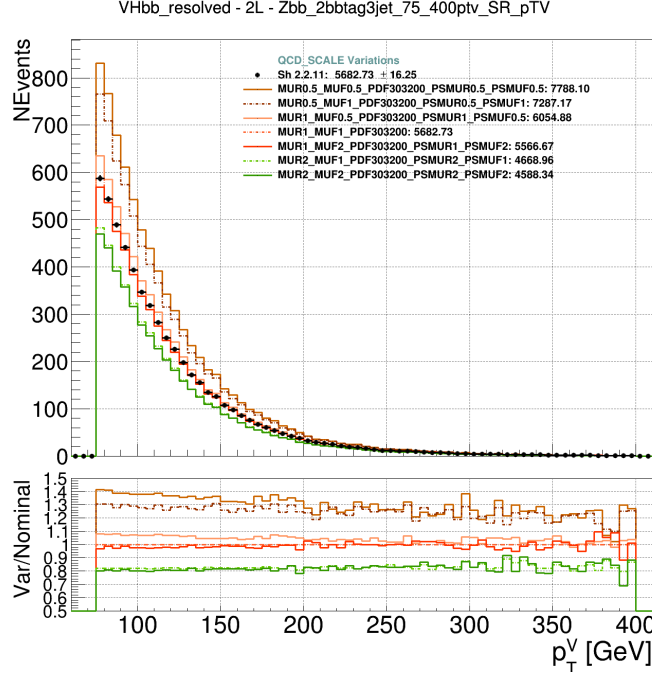


Figure 6.38: p_T^V distributions of $Z + 3$ jets $Z+bb$ events in the SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime obtained by varying the μ_R and μ_F renormalization and factorisation scales by $1/2$ and by 2 .

The SHERPA 2.2.11 and SHERPA 2.2.1 samples were generated with a different scale choice. Fig. 6.38 exemplifies how varying the QCD scales leads to relatively large differences in the normalization of the p_T^V distributions in the analysis phase space. Comparing these distributions obtained with the two versions of SHERPA shows that the differences between their yields fall within the SHERPA 2.2.11 QCD scale uncertainties in every bin. However, all the QCD variations yield a very similar p_T^V shape as the nominal sample and are thus unable to account for the overall difference in the p_T^V shape predicted with the two SHERPA versions. The additional set of uncertainties derived from the SHERPA 2.2.11 vs SHERPA 2.2.1 comparison to cover for this difference is detailed in Sec. 6.9.3.

¹⁵Less conservative ME-only QCD scale variations are used to derive the shape uncertainties because statistical fluctuations in the ME+PS variations are too large. The two sets of variations are everywhere consistent within statistical uncertainty.

PDF+ α_s variations PDF+ α_s variations are used to assess systematic uncertainties arising from limitations in the data statistics used to derive the PDFs, from the choice of the method used to derive them, and from the precision to which $\alpha_s(m_Z)$ is known. The variations are 100 replicas from the Hessian eigenvector set of the nominal NNPDF3.0NNLO PDFs, and the central values of the alternative CT18NNLO [176] and MSHT2020NNLO [177] PDF sets. Furthermore, two variations of α_s at ± 0.001 from its nominal value of 0.118 are included, which move the strong constant within its experimental constraints. The uncertainties are calculated using the LHAPDF tool [178]. First, distributions are produced $N_{\text{mem}} = 104$ additional times, where N_{mem} is the number of variations considered. Then, the uncertainty on the yield $\sigma^{(0)}$ of every bin is computed at 68% CL via

$$\delta^{\text{pdf}}\sigma = \sqrt{\sum_{k=1}^{N_{\text{mem}}} (\sigma^{(k)} - \sigma^{(0)})^2}. \quad (6.6)$$

In this analysis, PDF+ α_s variations have limited impact. This is exemplified in Fig. 6.39, showing the PDF+ α_s uncertainties in the p_T^V distribution of $Z + 3$ jets events, where the two b -tagged jets are b -jets, in the SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime.

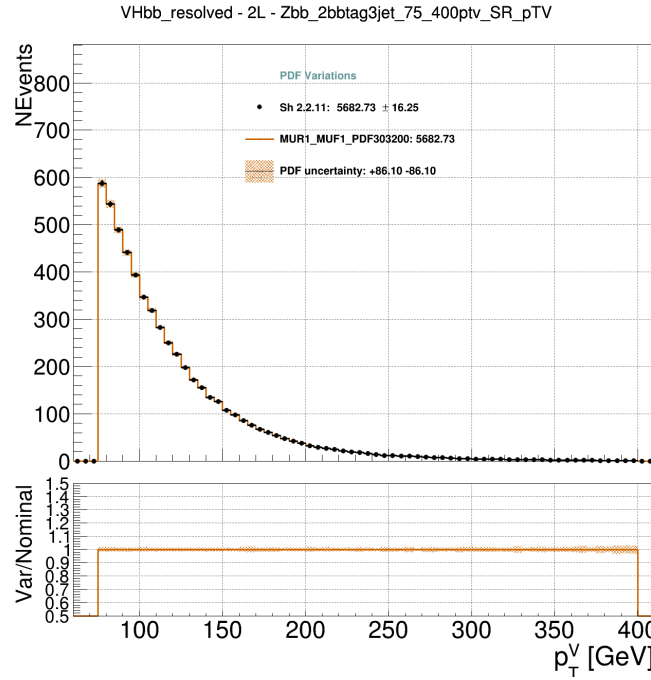


Figure 6.39: p_T^V distributions of $Z + 3$ jets $Z+bb$ events in the SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime displaying the uncertainty related to the PDF+ α_s variations, computed as in Eq. 6.6.

NLO EW Corrections The full NLO EW calculations for $V + \text{jets}$ events contain contributions from virtual loops and from real emissions of gauge bosons. The SHERPA 2.2.11 samples only contain the virtual-loop contributions, neglecting real emission both for simplicity and because real emissions of vector bosons would lead to fundamentally different experimental signatures that are irrelevant for the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis. The NLO EW corrections are combined with the NLO QCD terms using the *additive*, *multiplicative* and *exponentiated* schemes [179].

In the additive scheme, the cross-section calculated at NLO QCD and NLO EW accuracy for a $2 \rightarrow n$ final state can be expressed by summing together the QCD and EW NLO terms:

$$\sigma_{n,\text{QCD}+\text{EW}}^{\text{NLO}} = \sigma_{n,\text{LO}} + \Delta\sigma_{n,\text{QCD}}^{\text{NLO}} + \Delta\sigma_{n,\text{EW}_{\text{virt}}}^{\text{NLO}},$$

where $\sigma_{n,\text{LO}}$ denotes the Born-level cross-section and $\Delta\sigma_{n,\text{QCD}}^{\text{NLO}}$ the NLO QCD corrections to the Born process. The $\Delta\sigma_{n,\text{EW}_{\text{virt}}}^{\text{NLO}}$ term only contains the virtual-loop components.

The multiplicative scheme performs the combination according to:

$$\sigma_{n,\text{QCD}\times\text{EW}}^{\text{NLO}} = (\sigma_{n,\text{LO}} + \Delta\sigma_{n,\text{QCD}}^{\text{NLO}}) \times (1 + \Delta\sigma_{n,\text{EW}_{\text{virt}}}^{\text{NLO}}),$$

while, in the exponentiated scheme,

$$\sigma_{n,\text{QCD}\times e^{\text{EW}_{\text{virt}}}}^{\text{NLO}} = (\sigma_{n,\text{LO}} + \Delta\sigma_{n,\text{QCD}}^{\text{NLO}}) \times e^{\Delta\sigma_{n,\text{EW}_{\text{virt}}}^{\text{NLO}}}$$

is used. The multiplicative and exponentiated schemes are motivated by the factorization of the QCD and EW Sudakov corrections at high energies [179].

Comparisons of the NLO QCD-only prediction with and without the virtual electroweak corrections are shown in Fig. 6.40. The corrections reach 30% around $p_{\text{T}}^V \sim 1$ TeV and qualitatively follow the expected logarithmic behaviour owing to the appearance of EW Sudakov logarithms¹⁶ of the form $\log(s/m_V^2)$. The multiplicative and exponentiated contributions show the largest deviations from the NLO-only prediction and are in good agreement with each other.

The nominal samples use the additive combination scheme since it provides distributions closest to the NLO QCD-only prediction. The difference between the predictions obtained using the additive and multiplicative schemes and the additive and exponentiated schemes are interpreted as two distinct sources of systematic uncertainty for missing higher orders resulting from mixed QCD–EW contributions. Around $p_{\text{T}}^V \sim 1$ TeV, they account to approximately 10% for all final states.

Fig. 6.41 compares the NLO QCD-only prediction with and without the virtual electroweak corrections as seen in Z+jets events in some SRs of the $VH, H \rightarrow c\bar{c}$ regime. It shows the additive scheme yielding the distribution closest to the NLO QCD-only prediction.

¹⁶*i.e.*, logarithms that arise as NLO EW correction due to the emission and reabsorption of electroweak gauge bosons during high-energy scattering processes.

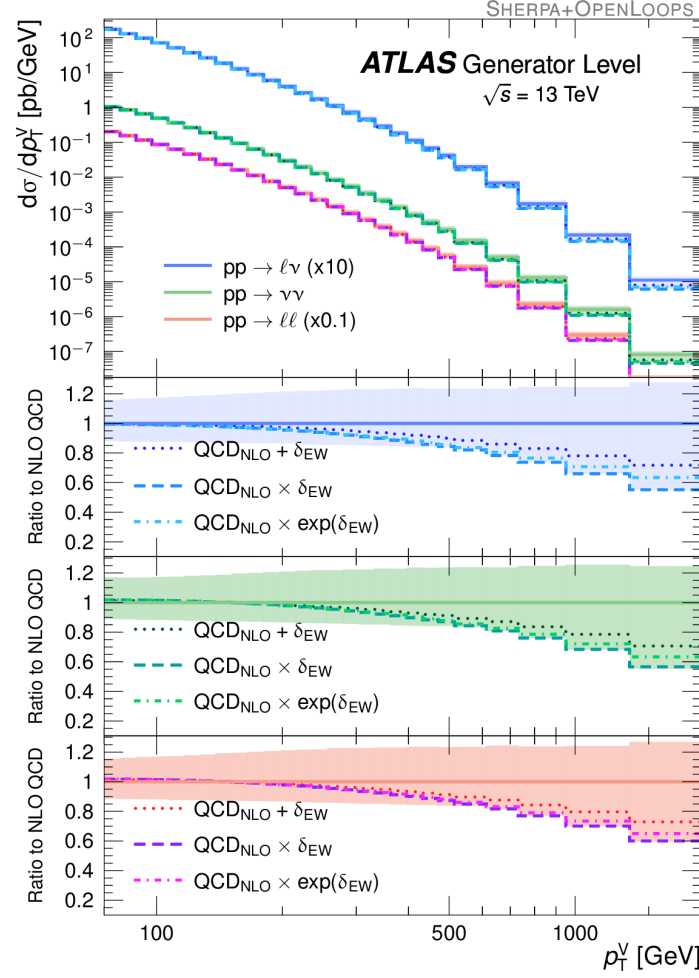


Figure 6.40: Differential distributions of the transverse momentum of the vector boson for $pp \rightarrow \ell\ell, \nu\nu, \ell\nu$ processes [179]. The QCD accuracy of all predictions is 0–2j@NLO+3–5j@LO. The electroweak virtual corrections are denoted by Δ_{EW} . The $\ell\ell$ and $\ell\nu$ final states have been averaged from electron, muon and τ -lepton final states. The shaded bands include scale uncertainties calculated from the envelope of renormalization and factorization variations.

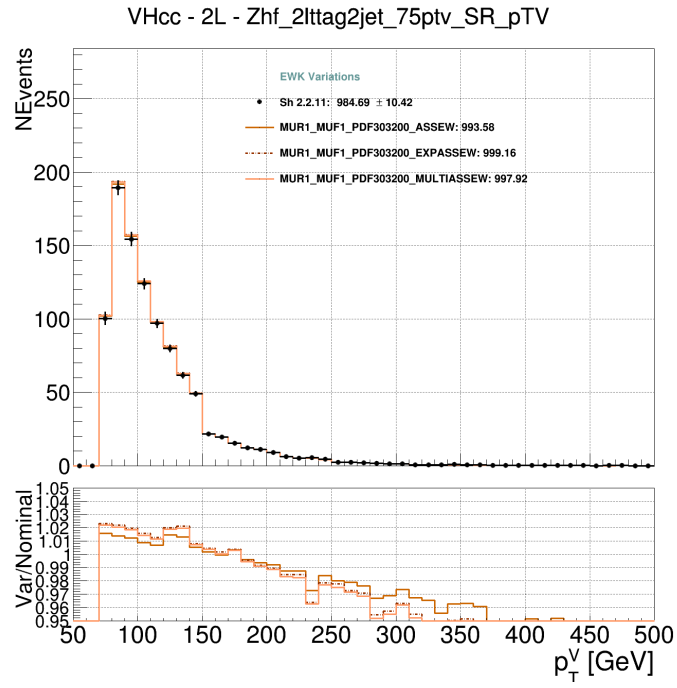


Figure 6.41: p_T^V distribution of $Z + 2$ heavy-flavour jets events in some SRs of the $VH, H \rightarrow c\bar{c}$ regime, comparing the NLO QCD-only prediction with and without the virtual electroweak corrections. The distribution obtained using the additive combination scheme is the one closest to the NLO QCD-only distribution.

6.9.2 Deriving normalisation and relative acceptance uncertainties

The strategy for deriving modelling uncertainties and configuring the contributions of the various processes in the fit is to group similar processes together. The W and Z +jets samples are divided into six different flavour components according to the jet flavour of the jets forming the Higgs-boson candidate: $W/Z+bb$, $W/Z+cc$, $W/Z+bc$, $W/Z+bl$, $W/Z+cl$ and $W/Z+l$. To account for 0L W +jets events containing a hadronic τ whose jet is not identified as a τ -jet, three additional flavour components are available in the 0L channel: $W+b\tau$, $W+c\tau$ and $W+l\tau$.

Under the assumption that different flavour components originating from the same physics process are expected to be similar, $W/Z+bb$ and $W/Z+cc$ events, that are generated from a vector boson produced in association with a gluon splitting into a pair of heavy-flavour quarks (Fig. 6.12a) or from the splitting of two gluons into heavy-flavour quark pairs and the subsequent fusion of two of the heavy-flavour quarks into a vector boson (Fig. 6.12b) are grouped together in the $W+hf$ and $Z+hf$ components.

$W/Z+bc$, $W/Z+cl$ and $W/Z+bl$ events are merged together as $W+mf$ and $Z+mf$, where mf stands for *mixed flavour*, because the invariant mass of the pair of tagged jets they contain is similar. In the 0L channel, $W+mf$ include also the $W+b\tau$ and $W+c\tau$ components.

The remaining flavours, namely $W/Z+l$, form the $W+lf$ and $Z+lf$ components, with $W+lf$ also including $W+l\tau$ events in the 0L channel. Here, lf stands for *light flavour*.

With the exception of the subdominant $W/Z+mf$ and $W/Z+lf$ components in the $VH, H \rightarrow b\bar{b}$ boosted regime, which are scaled to the SM prediction, all the V +jets components are left freely floating in the fit. Hence, they are not assigned any normalization uncertainty from the theory cross-section prediction. In the boosted $VH, H \rightarrow b\bar{b}$ regime, these uncertainties are 36% and 38% for the $W+mf$ and $W+lf$ components, respectively, and 35% for $Z+mf$ and $Z+l$.

In the fit, as far as $VH, H \rightarrow c\bar{c}$ is concerned, the $W/Z+lf$ components are mostly constrained in the LN regions. $W/Z+hf$ and $W/Z+mf$ events are mostly constrained in the signal and High- ΔR CRs. $W/Z+hf$ events are dominated by the $W/Z+cc$ components, and are constrained in the TT regions. $W/Z+mf$ events are dominated by the $W/Z+cl$ components, and are mostly constrained in the TL and TN regions. When it comes to $VH, H \rightarrow b\bar{b}$, the contribution of the $W/Z+mf$ and $W/Z+lf$ components is negligible. $W/Z+hf$ events are dominated by the $W/Z+bb$ components, which are constrained in the two b -tag regions.

For deriving the relative acceptance uncertainties, only the most dominant contribution in each group is used¹⁷, in contrast to the derivation of the shape uncertainties, where all the components of each group are used.

¹⁷This choice simplifies the framework while yielding no significant impact on the results.

Events	Variation	$W + cc/W + bb$	$W + cl/W + bc$	$W + bl/W + bc$
$W + \text{jets},$ $VH, H \rightarrow b\bar{b}$ boosted, 0L	MADGRAPH FxFx	$-2.4 \pm \mathbf{4.7} \%$	$\mathbf{9.4} \pm 7.0 \%$	$3.5 \pm \mathbf{6.8} \%$
	$1/2\mu_R \ \& \ 1/2\mu_F$	$\mathbf{-4.8} \%$	7.3 %	10.7 %
	$1/2\mu_R \ \& \ \mu_F$	-4.7 %	$\mathbf{8.6} \%$	$\mathbf{11.9} \%$
	$\mu_R \ \& \ 1/2\mu_F$	0.4 %	2.4 %	2.5 %
	$\mu_R \ \& \ \mu_F$	0.0 %	0.0 %	0.0 %
	$\mu_R \ \& \ 2\mu_F$	0.5 %	-1.1 %	-0.8 %
	$2\mu_R \ \& \ \mu_F$	-3.4 %	4.1 %	4.7 %
	$2\mu_R \ \& \ 2\mu_F$	-3.6 %	2.8 %	3.7 %
	PDF+ α_s up	5.1 %	-8.1 %	-7.9 %
	PDF+ α_s down	$\mathbf{-5.9} \%$	$\mathbf{11.3} \%$	$\mathbf{11.0} \%$
	EWK Exp.	$\mathbf{-0.1} \%$	$\mathbf{-1.1} \%$	$\mathbf{-0.8} \%$
	EWK Mult.	$\mathbf{0.3} \%$	$\mathbf{-1.3} \%$	$\mathbf{-1.1} \%$
	Total	$\mathbf{9.0} \%$	$\mathbf{17.1} \%$	$\mathbf{17.6} \%$

Table 6.12: Contribution of all the groups of alternative samples to the total relative acceptance uncertainty on the flavour components of $W + \text{jets}$ events in the 0L channel of the boosted $VH, H \rightarrow b\bar{b}$ regime. Only the bold values are included in the pythagorean sum of Eq. 6.5.

Relative acceptance uncertainties are computed as detailed in Sec. 6.8.2 by combining all the available alternative samples. Tab. 6.12, which serves as an example of many similar tables, shows the contributions of all the groups of alternative samples to the relative acceptance uncertainty on the flavour components of $W + \text{jets}$ events in the $VH, H \rightarrow b\bar{b}$ boosted regime. The total value for each uncertainty is computed with the pythagorean addition of the bold numbers in each column as indicated in Eq. 6.5. The first row displays the contribution of the MADGRAPH FxFx alternative sample, with its statistical uncertainty. When the magnitude of the statistical uncertainty surpasses the double ratio computed with Eq. 6.4, the statistical uncertainty is used in the squared sum. Since the MADGRAPH FxFx generator does not include NLO EWK corrections, when deriving the relative acceptance uncertainties they are also switched off in the nominal SHERPA 2.2.11 sample. The second group of alternative samples is composed of the QCD scale variations. Only the variation with the larger magnitude is considered. The same approach is used to deal with the PDF+ α_s variations. The contributions of both the NLO EW corrections obtained with the exponentiated and multiplicative combination schemes are considered.

Categories differ in the resolved and boosted regimes. Accordingly, the set of relative acceptance uncertainties to be calculated in the two cases varies.

Relative acceptance uncertainties in W +jets events For boosted W +jets events, the relative acceptance uncertainties taken into account are:

- **1L to 0L:** Applied to 0-lepton events, they amount to 20% for $W + hf$, $W + mf$ and $W + l$ events.
- **400 GeV $< p_T^V < 600$ GeV to $p_T^V \geq 600$ GeV:** Applied to events with $p_T^V \geq 600$ GeV, they amount to 3% for $W + hf$, $W + mf$ and $W + l$ events.
- **SR to TopCR:** Applied to the top CRs of the boosted analysis, they amount to 27% for $W + hf$ events, 20% for $W + mf$, and 16% for $W + l$.
- **Flavour Composition Uncertainties**, namely:
 - **$W + bb$ to $W + cc$:** Applied to $W + cc$ events, which float together with $W + bb$ events as $W + hf$, amounts to 11%.
 - **$W + bc$ to $W + cl$ and $W + bc$ to $W + bl$:** Applied to $W + cl$ and $W + bl$ events respectively, which float together with $W + bc$ events as $W + mf$, they amount to 15% and 9%.

The resolved regimes are more complicated, as the anchor regions change when performing the full $VH, H \rightarrow b\bar{b}/c\bar{c}$ combined fit or when fitting the $VH, H \rightarrow b\bar{b}$ or $VH, H \rightarrow c\bar{c}$ regions as standalone analyses. Furthermore, when statistics allows, the flavour components are decorrelated in p_T^V and jet multiplicity. This additional complexity largely increases the number of uncertainties to derive due to combinatory. In the following values, ranges appear when the same type of uncertainty is computed in different regions.

- **1L to 0L:** Applied to 0-lepton events, they amount to 13-30% for $W + hf$ events, 3-10% for $W + mf$ events, and 4% for $W + l$ events.
- **CR to SR**, namely:
 - **SR to High- ΔR CR:** Applied to $W + hf$ and $W + lf$ in the High- ΔR CRs, it amounts to 3-50% for $W + hf$ and 5-17% for $W + lf$.
 - **High- ΔR CR to SR:** Applied to $W + mf$ events in the SRs, it amounts to 2-16%.
 - **Low- ΔR CR to SR:** Applied to 1L $W + hf$ events in the SRs, it amounts to 3-65%.
- **Flavour Composition Uncertainties**, namely:
 - **$W + bb$ to $W + cc$:** Applied to $W + cc$ events, which float together with $W + bb$ events as $W + hf$, amounts to 4-25%.

- **$W + cl$ to $W + bc$ and $W + cl$ to $W + bl$:** Applied to $W + bc$ and $W + bl$ events respectively, which float together with $W + cl$ events as $W + mf$, they amount to 14-24% and 12-29%.
- **$W + l$ to $W + l\tau$ and $W + l$ to $W + \tau\tau$:** Applied to 0L $W + l\tau$ and $W + \tau\tau$ events, which float together with $W + l$ events as $W + lf$, they amount to 9%.

Relative acceptance uncertainties in Z +jets events For boosted Z +jets events, the relative acceptance uncertainties taken into account are:

- **2L to 0L:** Applied to the 0-lepton $Z + hf$ and $Z + mf$ events, they amount to 3%.
- **400 GeV $< p_T^V < 600$ GeV to $p_T^V \geq 600$ GeV:** Applied to $Z + hf$ and $Z + mf$ events with $p_T^V \geq 600$ GeV, they amount to 15%.
- **SR to TopCR:** Applied to the top CRs of the 0-lepton events, they amount to 15% for $Z + hf$, and to 25% for $Z + mf$.
- **Flavour Composition Uncertainties**, namely:
 - **$Z + bb$ to $Z + cc$:** Applied to $Z + cc$ events, which float together with $Z + bb$ events as $Z + hf$, amounts to 9% and 6% for the 0-lepton and 2-lepton channels respectively.
 - **$Z + bc$ to $Z + cl$ and $Z + bc$ to $Z + bl$:** Applied to $Z + cl$ and $Z + bl$ respectively, which float together with $Z + bc$ events as $Z + mf$, they amount to 9% and 6% in the 0-lepton and 2-lepton channels respectively.

In the resolved regime, they are:

- **2L to 0L:** Applied to 0-lepton events, they amount to 2-8% for $Z + hf$ events, 3-8% for $Z + mf$ events, and 4-10% for $Z + l$ events.
- **CR to SR**, namely:
 - **SR to High- ΔR CR:** Applied to $Z + hf$ the High- ΔR CRs, it amounts to 5%.
 - **High- ΔR CR to SR:** Applied to $Z + hf$ and $Z + mf$ in the SRs, it amounts to 7-30% for $Z + hf$ and 5-7% for $Z + mf$.
- **Flavour Composition Uncertainties**, namely:
 - **$Z + bb$ to $Z + cc$:** Applied to $Z + cc$ events, which float together with $Z + bb$ events as $Z + hf$, it amounts to 8-12%.
 - **$Z + cl$ to $Z + bc$ and $Z + cl$ to $Z + bl$:** Applied to $Z + bc$ and $Z + bl$ events respectively, which float together with $Z + cl$ events as $Z + mf$, they amount to 10% and 4%.

6.9.3 $V + \text{jets}$ shape uncertainties

Shape uncertainties factor in the residual differences in the shape of distributions derived from nominal and alternative samples after normalizing the alternative distribution to the nominal yield. During the fit procedure described in the next section, templates vary within their available uncertainties to best describe the data. A template using the degree of freedom provided by a normalisation uncertainty either increases or decreases its total yield. When it uses the degree of freedom provided by a shape uncertainty, its total yield is fixed, but its shape changes according to the interpolation procedure described in Sec. 6.10.2. The shape uncertainties for $V + \text{jets}$ uncertainty consider the QCD scale, $\text{PDF} + \alpha_s$ and NLO EWK variations using directly the distributions obtained by varying the internal weights of the SHERPA 2.211 nominal generator. Shape uncertainties from the comparison with the MADGRAPH FxFx alternative samples are derived by reweighting the nominal distributions to the alternatives with the CARL method, described in Sec. 6.8.2.

Additional shape uncertainties are introduced to account for the different p_T^V distributions obtained when comparing the SHERPA 2.2.1 and SHERPA 2.2.11 generators. This is done because the SHERPA 2.2.1 generator better describes the p_T^V distribution observed in data in the 100–350 GeV interval, as illustrated in Fig. 6.42. The plot compares predictions from the MADGRAPH FxFx, SHERPA 2.2.1, and SHERPA 2.2.11 generator with and without NLO EWK corrections to a measurement of the transverse momentum p_T^H of Drell–Yan lepton pairs with invariant mass in the range 66–116 GeV, which targets the Z -boson peak [180]. The red uncertainty band, shown on the NLO QCD-only prediction from the SHERPA 2.2.11 generator, arises from the QCD-scale variations. The blue uncertainty band, shown on the NLO QCD+EW prediction, is the envelope of the multiplicative and exponentiated NLO EW combination schemes. Albeit the nominal SHERPA 2.2.11 predictions are consistent with data within the QCD-scale uncertainties, it was already shown in Fig. 6.38 that they cannot account for the shape effects observed in the 100–350 GeV p_T^V range, that result in a $\sim 15\%$ underestimation of the data further worsened by the application of the NLO EW corrections.

The p_T^V shape uncertainties are derived by fitting the ratio of the normalized SHERPA 2.2.1 over the SHERPA 2.2.11 p_T^V distributions with a polynomial function, which is then used to compute reweighted distributions of the BDT scores both in the $VH, H \rightarrow c\bar{c}$ and $VH, H \rightarrow b\bar{b}$ regimes. The procedure is performed separately in the three lepton channels, for $W + \text{jets}$ and $Z + \text{jets}$ samples, in the different hf , mf and lf flavour components; inclusively before applying the flavour tagging selection. An example of the comparison between the p_T^V shapes of $Z + mf$ events in the 2L channel simulated with the two SHERPA versions is shown in Fig. 6.43. The SHERPA 2.2.11 prediction is displayed with and without the NLO EWK predictions. The uncertainty bands are the shape-only effects produced by the variations of the internal QCD scale and NLO EW weights, and are not able to account for the shape differences with SHERPA 2.2.1. The black line is the polynomial function that best fits the ratio.

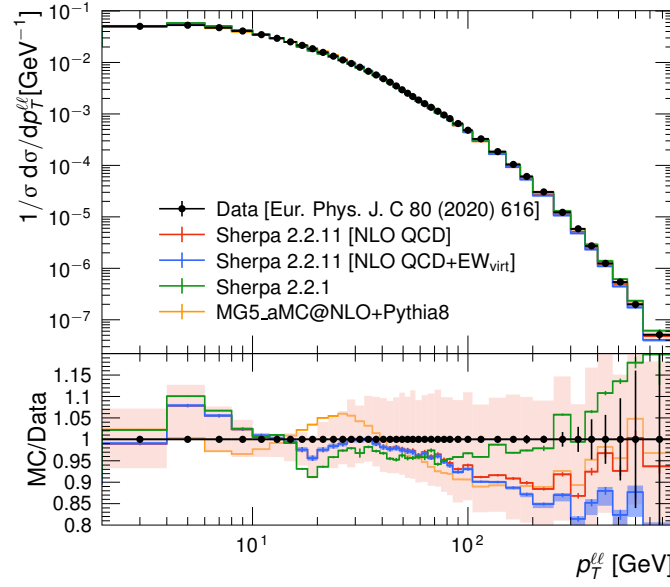


Figure 6.42: Unit normalized cross-sections as a function of the transverse momentum of the dilepton system [179]. Predictions of the MADGRAPH FxFx, SHERPA 2.2.1, and SHERPA 2.2.11 generator with and without NLO EWK corrections are compared with data points taken from the ATLAS measurement published in [180]. The red uncertainty band corresponds to the QCD-scale variations in the NLO QCD-only SHERPA 2.2.11 setup. The blue uncertainty band is the envelope of the multiplicative and exponentiated NLO EW combination schemes.

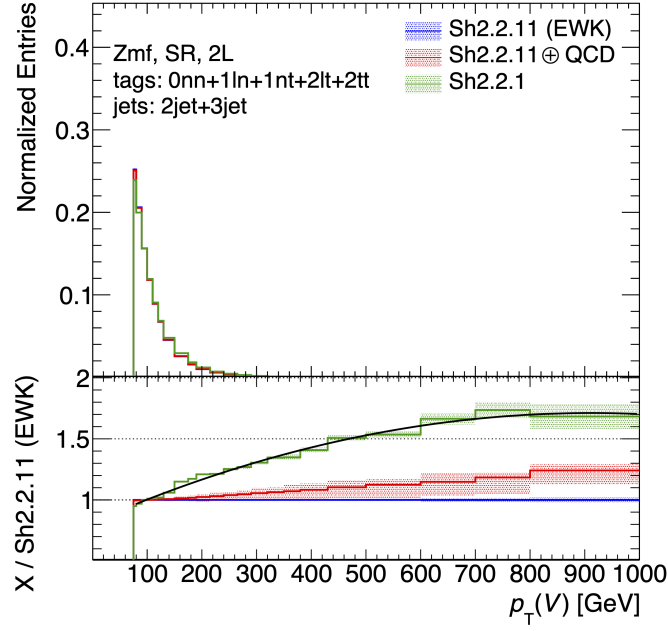


Figure 6.43: The p_T^V shape uncertainty derived from the SHERPA 2.2.11 and SHERPA 2.2.1 comparison for the $Z + mf$ events in the 0L channel. The blue line shows the distribution obtained with the nominal SHERPA 2.2.11 sample with the virtual EW corrections applied. The red line shows it for the same sample without the EW corrections. The green line shows it for the SHERPA 2.2.1 sample. The black line in the bottom panel is the polynomial function which best fits the data, which is used to reweight the distributions of the BDT scores to provide the shape uncertainties.

6.10 Statistical treatment and inference

To measure the $VH, H \rightarrow b\bar{b}/c\bar{c}$ signals, a *profile binned likelihood fit* is performed in all the SRs and CRs, together in all three lepton channels. To cross-check the strategy of the analysis, the procedure is repeated by measuring the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ and $VW, Z \rightarrow (b/c)q$ SM candles. In all the SRs and most CRs, the distribution which is fitted is the output score of the BDT described in Sec. 6.7. In some of the CRs, the H_{cand} mass, p_{T}^V , the output of the BDT trained to differentiate between the $t\bar{t}$ and $W + \text{jets}$ backgrounds in the low- ΔR CR (Fig. 6.23a), or single-bin distributions are fitted instead. The parameters of interest (POIs) in the fit are the signal strengths of the $VH, H \rightarrow b\bar{b}/c\bar{c}$, $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ and $VW, Z \rightarrow (b/c)q$ signals. The implementation of the fit builds on HISTFACTORY [181] and on the ROOFIT [182] and ROOSTATS [183] software packages.

6.10.1 The profile binned likelihood function

The fit procedure consists in maximising a profile binned likelihood function with respect to both the signal strength and the systematic uncertainties. The likelihood function is the product of three terms. The first term is the poissonian probability of the observed data given the predicted signal and background distributions, assuming a counting experiment in each bin and disregarding any source of uncertainty other than the statistical uncertainty on data:

$$\mathcal{L}_{\text{phys}}(\text{data}|\vec{\mu}) = \prod_{i \in \text{bins}} \text{Pois}(N_i|\vec{\mu} \cdot \vec{s}_i + b_i) = \prod_{i \in \text{bins}} \frac{(\vec{\mu} \cdot \vec{s}_i + b_i)^{N_i}}{N_i!} e^{-(\vec{\mu} \cdot \vec{s}_i + b_i)}, \quad (6.7)$$

where N_i is the number of data events observed in bin i , b_i is the total count of the predicted backgrounds, $\vec{s}_i$ is the vector containing the predicted counts of the VH , VZ and VW signal processes in their targeted decay channels, and $\vec{\mu}$ contains their signal strengths.

Systematic uncertainties affect the likelihood function in the form of Nuisance Parameters (NPs) $\vec{\theta}$, which can modify the predicted signal and background yields: $\{s_i, b_i\} \rightarrow \{s_i(\vec{\theta}), b_i(\vec{\theta})\}$. The central value and uncertainty associated with a NP are typically determined in advance through auxiliary measurements, as it happens, for instance, for the jet flavour-tagging efficiency calibrations or for the determination of the relative acceptance uncertainties. In such cases, the information from the auxiliary measurement is incorporated into the likelihood by multiplying $\mathcal{L}_{\text{phys}}$ by a normal gaussian term $\mathcal{L}_{\text{Aux}}$:

$$\mathcal{L}_{\text{Aux}}(\vec{\theta}) = \prod_{\theta \in \vec{\theta}} \text{Gauss}(\theta|0, 1) = \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}. \quad (6.8)$$

$\mathcal{L}_{\text{Aux}}$ parametrizes the auxiliary measurements of the NPs, which are rescaled to zero mean and unitary standard deviation. The fit process consists in maximising the likelihood function with respect to the POIs and NPs. Maximising the product $\mathcal{L}_{\text{phys}} \times \mathcal{L}_{\text{aux}}$ with respect

to a NP θ corresponds to a joint measurement of θ , where $\mathcal{L}_{\text{phys}}$ can provide additional constraining power, if the analysis is particularly sensitive to θ . The variation from zero of the best value of a NP after the fit is called the *pull*. A decrease in its variance after performing the fit is called a *constraint*¹⁸.

Some NPs, called *free-floating* parameters and labelled with the symbol $\vec{\tau}$, are not associated with any auxiliary measurement in the likelihood function. They are used to modify the overall normalisations of the main backgrounds in the fit and are determined purely from $\mathcal{L}_{\text{phys}}$, which leverages the constraining power of the data in the dedicated CRs. Before the fit procedure, the values of free-floating parameters are set to 1.

Finally, uncertainties on the yields of the predicted events due to the limited size of the MC samples must be addressed [184]. The yields of the simulated signal and background events in every bin are sampled from the process' underlying probability distribution via the MC technique. To parameterize their uncertainty in the likelihood function, they are treated as the results of counting experiments from poissonian distributions whose expected values are determined by NPs, labelled with the γ symbol, that must be determined in the fit. In this conceptual framework, the likelihood of observing a yield b_i^p of simulated events from a process p in the bin i is

$$\mathcal{L}_{\text{MC yield}}(\gamma_i^p) = \text{Pois}(b_i^p | \gamma_i^p b_i^p) \quad (6.9)$$

Assigning a γ parameter to every process in every bin would explode the number of parameters to determine in the fit. To simplify the model, the Barlow and Beeston procedure is followed [184], which consists in introducing a single γ in every bin. These γ can modify the predicted yields, *i.e.* $b_i(\vec{\theta}) \rightarrow \gamma_i b_i(\vec{\theta})$, and are constrained by

$$\mathcal{L}_{\text{BkgStat}}(\vec{\gamma}) = \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}), \quad (6.10)$$

with

$$\beta_i = \frac{1}{\sigma_{\text{rel},i}^2}, \quad (6.11)$$

where σ_{rel} is the relative statistical uncertainty on the predicted total background yield in the bin. The analytical form of $\mathcal{L}_{\text{BkgStat}}$, which is not poissonian and where the predicted yields b_i do not appear, follows from two subtleties. Firstly, whereas, according to first principles, it should be described with the product of poissonian terms, using gaussians decreases the time needed for the fit to converge without affecting the final results. Secondly, poissonian distributions with expected values λ have a standard deviation of $\sqrt{\lambda}$. Consequently, their

¹⁸While constraints on NPs are not inherently problematic, they are approached with suspicion. Auxiliary measurements are meticulously optimized to yield the most accurate information about the targeted NPs. Combining these measurements with others is not expected to enhance their precision, as these auxiliary measurements are already designed to provide the best possible knowledge about the specific NPs they focus on. Hence, when developing the fit model, significant efforts are dedicated to understanding any additional information contributed by the analysis, if constrained NPs are present. In many cases, unexpected constraints hint at erroneous implementations of NPs in the fit model, resulting in overestimates.

relative statistical uncertainty $\sigma_{\text{rel}} = \frac{1}{\sqrt{\lambda}}$ is related to their expected value λ by $\lambda = \left(\frac{1}{\sigma_{\text{rel}}}\right)^2$. This means that β_i is the yield that would be expected in a bin with relative statistical uncertainty $\sigma_{\text{rel},i}$ if the sampling was poissonian. Using β_i instead than b_i in $\mathcal{L}_{\text{BkgStat}}$ better suits the poissonian description of the constraint, factoring in the fact that the MC statistical uncertainties actually arise from Eq. 5.1, that takes into account the weights.

γ parameters are not applied to the VH signal processes, which, for their special role, are simulated with enough statistic to disregard their relative statistical uncertainty; nor to the data-driven backgrounds, whose statistical uncertainty is parametrized with the θ parameter associated to their auxiliary measurement.

The complete profile binned likelihood function is hence:

$$\mathcal{L}(\vec{\mu}, \vec{\theta}, \vec{\gamma}, \vec{\tau}) = \prod_{i \in \text{bins}} \text{Pois}(N_i | \vec{\mu} \cdot \vec{s}_i(\vec{\theta}) + \gamma_i b_i(\vec{\theta}, \vec{\tau})) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \times \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}). \quad (6.12)$$

The best estimates of the POIs and their uncertainties are obtained by minimising $-\ln \mathcal{L}$ with respect to all parameters $\vec{\mu}$, $\vec{\theta}$, $\vec{\tau}$ and $\vec{\gamma}$ through a numerical gradient descent implemented in the MINUIT package [185]. The likelihood is also used to derive the significance of the observed signal with respect to the background only hypothesis $\vec{\mu} = 0$, using the profile likelihood ratio

$$\lambda(\vec{\mu} = 0) = \frac{\mathcal{L}(0, \hat{\vec{\theta}}, \hat{\vec{\tau}}, \hat{\vec{\gamma}})}{\mathcal{L}(\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\tau}}, \hat{\vec{\gamma}})} \quad (6.13)$$

where $\{\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\tau}}, \hat{\vec{\gamma}}\}$ are the best estimates of the POIs and NPs, and $\{\hat{\vec{\theta}}, \hat{\vec{\tau}}, \hat{\vec{\gamma}}\}$ are the best estimates of the NPs given the background-only hypothesis $\vec{\mu} = 0$. By construction, $\lambda(\vec{\mu}) \in (0, 1]$. Values close to 1 indicate good agreement between the data and the $\vec{\mu} = 0$ hypothesis, while values approaching 0 allow to reject it with increasing CLs.

From the likelihood ratio, a test statistic q_μ can be defined:

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0 \\ 2 \ln \lambda(0) & \hat{\mu} < 0 \end{cases} \quad (6.14)$$

where the second case is included to deal with the possibility that a negative fluctuation in the data favours a negative contribution of the signal strength [154]. In the following, the second case is omitted for simplicity.

Test statistics such as q_0 are crucial because the probability distributions to which they tend given enough data, called *asymptotic distributions*, can be derived from first principles. For sample sizes large enough to satisfy the Wilks and Wald approximation [154], the probability density of q_0 in the background only hypothesis, $f(q_0 | \mu = 0)$, is a χ^2 distribution. It can

be used to derive the p -value p_0 , *i.e.* the probability, under the background only hypothesis H_0 , of finding data of equal or greater incompatibility with the prediction of H_0 compared to the observed data. p_0 is defined as:

$$p_0 = \int_{q_{\vec{\mu}, \text{obs}}}^{\infty} f(q_0 | \vec{\mu} = 0) dq_0 \quad (6.15)$$

$q_{0, \text{obs}}$ is the value of q_0 observed in data.

p -values are frequently expressed in terms of the *equivalent gaussian significance* Z , defined in such a way that a variable following a gaussian distribution, when found Z standard deviations above its mean, has an upper-tail probability equal to p :

$$Z = \Phi^{-1}(1 - p), \quad (6.16)$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard gaussian. Under the assumptions of the Wilks and Wald theorem [154], $Z = \sqrt{q_0}$.

In the particle physics community, for claiming a new discovery, the background-only hypothesis must be rejected with a significance $Z \geq 5\sigma$, which translates in a p -value $p \leq 2.87 \times 10^{-7}$.

For the $VH, H \rightarrow c\bar{c}$ an upper limit at 95% CL on the signal strength is calculated using a modified frequentist CL_s method [154, 186]. Such limit indicates the value of the signal strength above which the $s + b$ hypothesis given the observed data has a p -value smaller than 0.05¹⁹.

The Asimov dataset An important tool to evaluate the expected sensitivity of an analysis, introduced in [154], is the *Asimov dataset*. It enables a computationally efficient method to calculate the expected median²⁰ significance with which data distributed according to a specified signal+background hypothesis could reject the background-only hypothesis, or the expected median upper limit that the data would set on the strength of the signal. This is crucial because optimizing these values is one of the main goals to pursue when designing the analysis.

The brute-force approach to computing expected significances and upper limits would require producing the MC samples multiple times, extracting the resulting significances and upper limits with respect to the distributions hypothesized on data, and then calculating their means and variances. The Asimov dataset method allows the extraction of these values by means of a single fit to a dataset of *pseudo-data*, the so-called Asimov dataset.

¹⁹*i.e.*, the value of the signal strength above which the probability of observing the data given the signal plus background hypothesis is less than 5% .

²⁰The method proposed on [154] does not allow to predict expected *mean* significances and upper limits due to the non-linear relations between the best estimates of the POIs, the test statistics and the equivalent significance. Nevertheless, for samples large enough for gaussian behaviour to take over, median and mean converge to the same value.

Such a dataset is built by adding the nominal signal and background templates setting the NPs to the values estimated via a fit to the observed data where the POIs are set to 1 and the NPs associated with the signal are fixed to their initial values; and by assigning poissonian uncertainties to every bin. An Asimov dataset produced in this way is labelled as *post-fit*, and is used for computing the expected significances and upper limits, as well as studying the NPs constraints and correlations²¹. In contrast, a pre-fit Asimov dataset, built by also setting all the NPs associated with the backgrounds to their pre-fit priors, is used to assess preliminary impacts of sub-sets of the NPs on the best estimates of the POIs and on their uncertainties (Sec. 6.12.1).

6.10.2 $VH, H \rightarrow b\bar{b}/c\bar{c}$ fit model

Fit inputs The profiled binned likelihood function is constructed from the one-dimensional histograms of the variables chosen in the SRs and CRs schematized in Fig. 6.19. Their pre-fit distributions are compared to the observed data in Fig. 6.21-6.25. The fit variable used in the V +light CR is p_T^V . This choice helps determine from data the p_T^V shapes of the V +jets backgrounds, which suffer from the modelling issues detailed in Sec. 6.9.3.

Inter- and extrapolation To assess the impact of experimental and modelling uncertainties, the corresponding NPs are varied to their $\pm\sigma$ values, and the resulting distributions are stored in additional templates. However, to maximise the likelihood function, a continuous parametrization of its dependence upon each POI and NP is necessary. To achieve this, a polynomial interpolation between the central values and the $\pm\sigma$ bounds and an exponential extrapolation to variations larger than 1σ are used. This approach guarantees the continuity of both the first and second derivatives of the likelihood function with respect to its parameters, while preventing the occurrence of negative yields.

Symmetrisation Some systematic uncertainties are associated only with one-sided variations. These asymmetries are caused by technical limitations in the auxiliary measurements. For instance, to derive the experimental systematic uncertainties related to the resolution of the jet energies, the resolution predicted by the MC simulation is smeared to that achieved by the detector. There is no second alternate template because it is not possible to unsmear the resolution of the simulation to make it better. In the unrealistic situation where the detector achieved a better resolution than its simulated counterpart, the MC resolution would be maintained in the simulated jets.

In instances where only a one-sided variation exists, the impact of the NP on the fitted distribution is symmetrized to derive the variation on the alternate side.

²¹This description holds for the Asimov dataset used in the blinded stages of the analysis, as described in the following section. After unblinding, the POIs and NPs related to the signal are not fixed when building the dataset.

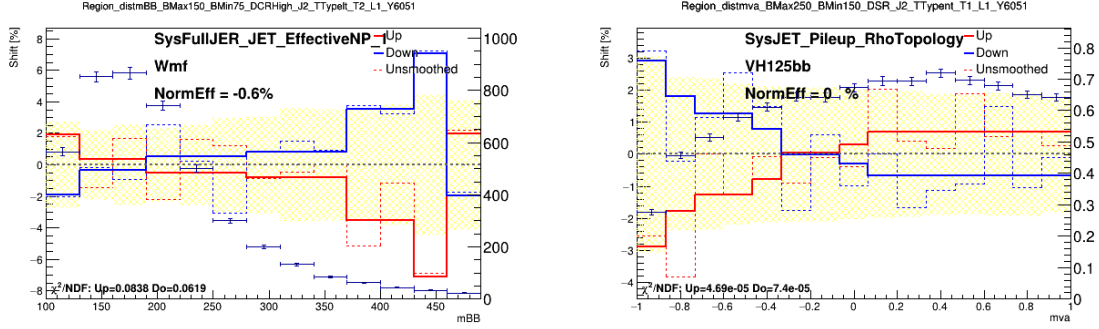


Figure 6.44: Effect of the parabolic (a) and monotonic (b) smoothing algorithms. Solid lines show the shift from 1 of the rebinned ratios associated with $\pm 1\sigma$ shape variations over the nominal template. The dotted lines display the ratios before the smoothing. In the parabolic case, one local extreme is allowed: in (a), it is found in the second to last rebinned bin. Cross markers show the nominal template, associated with the scale on the right y axis. The yellow bands represent the statistical uncertainties associated with the nominal template.

Smoothing Statistical fluctuations affecting alternative templates that model shape uncertainties in poorly populated regions of the analysis phase space can lead to artificial constraints in the fit. This problem is reduced by using techniques like CARL reweighting, and the residual fluctuations are addressed with smoothing iterative algorithms.

Smoothing algorithms modify the alternative templates as follows. They start by rebinning the templates so that the number of local extremes in the shape of the ratio of the alternative over the nominal template does not exceed a given limit. Parabolic smoothing, performed in the regions where p_T^V or the mass of the Higgs-boson candidate are fitted, allows for at most one local extreme. Monothonic smoothing, performed where the fit uses the BDT score, forbids local extremes. The templates and their ratio are then further rebinned until no bin of the nominal has a relative statistical uncertainty exceeding 5%. The yields of the alternative template are then adjusted based on the rebinned ratio, by substituting them with the multiplication of the the yields of the nominal template, taken before rebinning, by the value of the corresponding bin in the rebinned ratio.

The effect of the monotonic and parabolic smoothing procedures can be appreciated in Fig. 6.44, where rebinned ratios associated with shape variations of the mass of the Higgs-boson candidate and of the score of the BDT are compared to the ratios before rebinning. In the plots, the nominal distributions are overlaid with cross markers, associated to the scale on the right axis. The yellow band corresponds to the MC statistical uncertainty of the nominal bins.

Pruning A pruning technique is used to simplify the fit model by considering only the limited subset of NPs that significantly impact the result. This approach enhances the stability of the fit and reduces the time it needs to converge. It consists in discarding:

- normalisation uncertainties smaller than 0.5%, or inducing up and down variations having the same sign.
- shape variations where none of the bins varies more than 0.5%, or where either the $+1\sigma$ or the -1σ variation has no impact.
- normalisation and shape uncertainties associated with background samples in regions where their yield is less than 2% of the signal or less than 0.5% of the total background.

6.11 GNN truth-tagging application in $VH, H \rightarrow b\bar{b}/c\bar{c}$

The GNN-based truth flavour-tagging technique, detailed in Chap. 5, is tailored for integration into the analysis. This marks the inaugural application of the technique to simulated samples employed in a physics measurement.

Two distinct sets of GNN models are trained to deal with background samples in the resolved and merged topologies, due to the different jet collections used to reconstruct H_{cand} and to the different jet flavour-tagging schemes which are applied in the two cases. Multiple models are needed within the resolved and boosted regimes because the MC generators simulating the background samples differ for different processes: the models are trained to learn the parametrization of the jet flavour-tagging efficiency functions $\epsilon_f(\mathbf{x}|\theta)$, and $\epsilon_f(\mathbf{x})$ differs in different MC generators.

In principle, a single model achieving a good parametrization of the dependency of ϵ_f over the full set of *relevant* parameters (Sec. 5) can be successfully used to weight events arising from different physics processes. Nevertheless, in the resolved topology, training separate models for separate processes improves the performance. This observation hints at an imperfect parametrization of ϵ_f , a hypothesis further corroborated by the persistence of a few cases in which the distributions obtained when truth flavour-tagging the jets do not close perfectly on those obtained with pass-or-fail flavour-tagging. Albeit imperfect, the parametrization achieved by the GNN-based models is much more accurate than that achieved by means of the two-dimensional (p_T, η) maps, which were used in the past $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ analyses published by the ATLAS Collaboration. In the previous $VH, H \rightarrow c\bar{c}$ analysis, such inferior parameterization forced including additional systematic uncertainties ranking as high as seventh among those causing the highest impact on the final results.

Tab. 6.13 lists all the models used to truth flavour-tag the events in the different analysis regimes.

Process	Events in training sample	
	Resolved	Boosted
PYTHIA $t\bar{t}$	PYTHIA $t\bar{t}$	All PYTHIA events
PYTHIA single-top	PYTHIA single-top	
PYTHIA signal	PYTHIA signal	
SHERPA 2.2.11 W + jets	SHERPA 2.2.11 V + jets	SHERPA 2.2.11 V + jets
SHERPA 2.2.11 Z + jets		
SHERPA 2.2.11 diboson		

Table 6.13: GNN-based models trained to apply truth jet-flavour tagging in the analysis. The models are trained inclusively in all 3 lepton channels

Hybrid tagging As detailed in Chap. 5, the truth jet-flavour tagging technique is especially relevant when dealing with simulated background samples where H_{cand} is formed by a pair of mistagged light-flavour jets, since most of the simulated events would not meet the pass-or-fail flavour-tagging criteria. In some cases, the improvement brought about by applying the technique to simulated events which contain true heavy-flavour jets is negligible, since the flavour-tagging scheme of the analysis is tailored to select them with high efficiency. In such scenarios, it is advantageous to employ a hybrid-tagging approach, in which simulated b or c -jets, contingent upon whether the cut pertains to the $VH, H \rightarrow c\bar{c}$ or $VH, H \rightarrow b\bar{b}$ regime, are selected through pass-or-fail flavour-tagging, while the other classes of jets are weighted with truth flavour-tagging. When hybrid-tagging, the negligible loss of simulated events is compensated by more direct modelling of the dominant class of jets, which does not rely on the parametrization of $\epsilon_f(\mathbf{x}|\theta)$.

Truth-flavour tagging strategy Truth flavour-tagging is conjugated in different ways at different steps of the analysis. Most noticeably, it is used in the selection of the nominal events that populate the templates that float in the fit, to reduce their statistical fluctuations and the impact of the γ NPs on the final results.

In this context, a complex strategy was devised, that deals differently with different backgrounds in the $VH, H \rightarrow c\bar{c}$ and boosted and resolved $VH, H \rightarrow b\bar{b}$ regimes. Such strategy, schematized in Tab. 6.14, was designed to effectively utilize the statistical power of the simulated samples while avoiding cases where the parametrization of $\epsilon_f(\mathbf{x}|\theta)$ is imperfect.

Signal events are selected with pass-or-fail flavour-tagging, to rule-out the risk of introducing bias through truth-flavour tagging; and because truth-flavour tagging does not increase their statistics substantially. All the remaining samples pertaining to the boosted regime are truth flavour-tagged. In the resolved $VH, H \rightarrow b\bar{b}$ regime, only V + jets and single-top events

in the s and t channel are selected with hybrid flavour-tagging, where the b -jets undergo pass-or-fail flavour tagging.

In $VH, H \rightarrow c\bar{c}$, H_{cand} can be reconstructed by a tight-tight, tight-loose or loose-tight c -tagged pair of jets (Fig. 6.7b). In this case, the combination that benefits most from the truth flavour-tagging technique is tight-tight, due to the small c -jet efficiency and the large b and light-flavour jet rejection associated with the tight c -tagging working point. Truth flavour-tagging in the c -tag area of Fig. 6.7b, though, raises an additional problem. By construction, when applying truth flavour-tagging, all the available events are weighted and utilized to populate the targeted flavour-tagging regions. If all the available statistics of a simulated sample were utilized to populate the tight-tight c -tagging region, none would remain to populate the tight-loose, loose-tight, and tight-no-tag regions. A seeming solution would be using all the available statistics once more to populate the regions in which pass-or-fail flavour-tagging is applied. This approach is not conceptually sound because it would falsify the assumption that the counts in the different bins of the templates used in the fit are independent, which is foundational for building a likelihood function as a product of poissonian terms. Hence, it was decided to dedicate 8% of the available statistics to truth flavour-tagging in the tight-tight region, and to select the remaining events with pass-or-fail flavour-tagging ²².

This complicated picture emerged from a careful trial and error procedure. Initially, aiming at an extensive application of the technique to all the main backgrounds' templates, the closure of the distributions obtained by applying truth-flavour tagging on those relying on pass-or-fail flavour tagging was monitored everywhere. New GNN-models were trained exclusively on the background samples where regions of non-closure were observed. Eventually, no further improvement on the residual non-closing regions was to be gained from more exclusive trainings.

To quantify the performance of the method over all the regions involved in the fit, the templates obtained by means of truth flavour-tagging were fitted to a pre-fit Asimov pseudodataset (Sec. 6.10) obtained by summing the templates obtained with pass-or-fail flavour-tagging. In a scenario where a perfect parametrization is realized, no NP is pulled in the fit. The few pulled NPs hinted at the regions where the residual non-closures had non-negligible impact, and truth flavour-tagging was not applied to the background processes associated with the non-closing templates.

The amount of statistics to dedicate to truth-flavour tagging in the tight-tight c -tag region of the $VH, H \rightarrow c\bar{c}$ regime was chosen by trading off the statistical uncertainty in the tight-tight c -tag region, decreasing for increasing fractions of truth flavour-tagged events, with its corresponding increase in the remaining regions.

More aggressive approaches dedicating more events to truth flavour-tagging and applying

²²A residual overlap that causes double-counting of events in the resolved $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ regions, introduced by truth flavour-tagging, remains. Due to the very different weights that truth flavour-tagging applies in the two regions, the effects of this residual overlap were disregarded.

	$VH, H \rightarrow b\bar{b}$ resolved	$VH, H \rightarrow b\bar{b}$ boosted	$VH, H \rightarrow c\bar{c}$
Truth-tag WP combination	70% & 70% b -tag	85% & 85% b -tag	tight & tight c -tag
% of truth-tag events	100%	100%	8%
V + jets	hybrid tag	truth tag	truth tag
single-top s/t	hybrid tag	truth tag	truth tag
single-top Wt	pass-or-fail tag	truth tag	pass-or-fail tag
$t\bar{t}$	pass-or-fail tag	truth tag	pass-or-fail tag
diboson	pass-or-fail tag	truth-tag	pass-or-fail tag
signal	pass-or-fail tag	pass-or-fail tag	pass-or-fail tag

Table 6.14: Final strategy for the application of GNN-based truth flavour-tagging to the MC templates used in the fit.

the method also to the loose c -tag WP were considered. They were not pursued because they brought about negligible reductions of the impact of the MC statistical uncertainty over the uncertainty on the expected $\mu_{VH, H \rightarrow c\bar{c}}$. This was not surprising, because the uncertainty on $\mu_{VH, H \rightarrow c\bar{c}}$ is dominated by the statistical uncertainty of the observed data, given the predicted rarity of the $VH, H \rightarrow c\bar{c}$ process,.

Fig. 6.45 compares two distributions obtained when applying the truth flavour-tagging strategy to the distributions obtained with pass-or-fail flavour tagging in one of the SRs and one of the CRs of the $VH, H \rightarrow c\bar{c}$ regime. The distributions obtained by applying the truth-flavour tagging strategy parametrizing $\epsilon_f(\mathbf{x}|\theta)$ with (p_T, η) efficiency maps are overlaid to show the improvements brought about by the GNN models.

The distribution of the BDT score is showcased in Fig. 6.45a for $Z + mf$ 2-jet 2L events in the $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ SR. The jets must be c -tagged in the tight-tight, loose-tight or tight-loose categories, which are merged together in the templates used in the SRs during the fit. In this case, given the dominant contribution of the pass-or-fail flavour-tagged loose-tight and tight-loose events, the truth flavour-tagging method does not significantly improve the statistical uncertainties in the bins. Nevertheless, the superior parametrization achieved by the GNN allows to solve the mismodelling that would be introduced by using map-based truth-flavour tagging.

Scenarios in which the method expresses its potential are exemplified in Fig. 6.45b, showcasing the H_{cand} mass distribution of $Z + l$ events in the high- ΔR CR. In the CRs, the templates of tight-tight c -tag events are not merged with those of loose-tight and tight-loose c -tag events. Hence, the significant improvement brought about by truth-flavour tagging the events is glaring. The GNN model, which captures the flavour-tagging efficiency much better than its map-based counterpart, strongly reduces the statistical fluctuations arising when applying pass-or-fail flavour-tagging.

Truth flavour-tagging was also applied to all the events used to train the BDT described

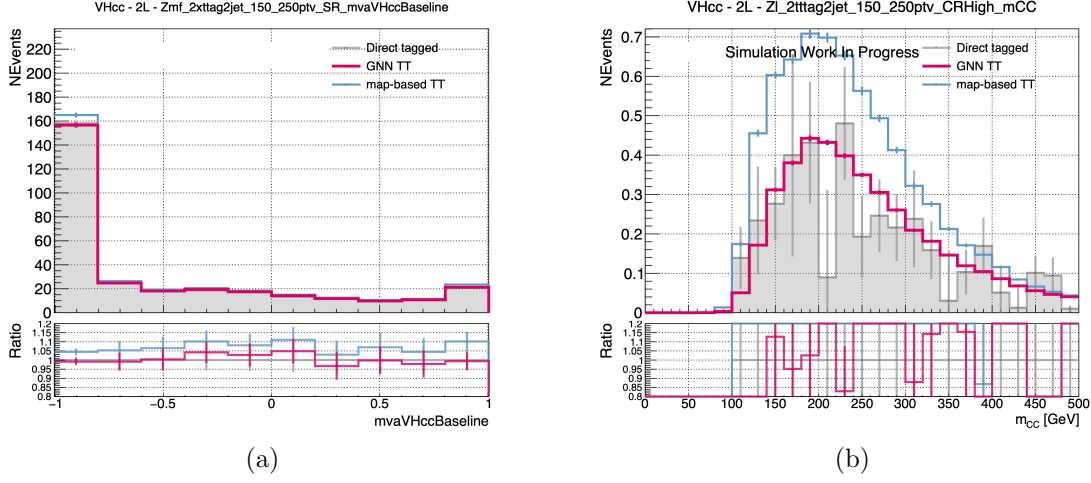


Figure 6.45: Two of the closure plots produced to monitor the performance of truth-flavour tagging in the SRs (a) and CRs (b) of the $VH, H \rightarrow c\bar{c}$ regime. In the SRs, the templates of processes flavour-tagged in the tight-tight, loose-tight or tight-loose categories are merged together during the fit. Therefore, the impact of truth flavour-tagging the subdominant tight-tight category is negligible. In the CRs, where the tight-tight templates and the tight-loose and loose-tight templates are separated, the success of the truth flavour-tagging method in reducing the statistical uncertainties of the MC sample is glaring.

in Sec. 6.7, after checking the closure in the distributions of the input variables (Fig. 6.46). The truth flavour-tagging method significantly increases the size of the training samples.

Furthermore, in cases benefitting from increased simulated statistics, truth flavour-tagging is used when modelling the systematic uncertainties. In particular, it was applied to all the simulated events composing the nominal and alternative samples used in the training of some of the CARL models, and to derive relative acceptance uncertainties in some regions.

In both scenarios, the GNN-models trained on the events composing the nominal sample were used to infer the truth-flavour tagging weights of the events generated with the alternative MC as well. This approach proves advantageous because it allows the use of the alternative MC generators to model the systematic uncertainties without having to take into account the different efficiencies achieved by the flavour-tagging algorithms when tagging jets belonging to the alternative sample compared to those belonging to the nominal. Tab. 6.15 exemplifies the impact of using truth flavour-tagging when deriving the relative acceptance uncertainties. It shows the flavour composition uncertainties in 0L $W + \text{jets}$ events in the resolved $VH, H \rightarrow b\bar{b}$ regime. Truth flavour-tagging reduces the total uncertainties on all the flavour ratios. Additionally, it sizeably reduces the MC statistical uncertainties associated to the MADGRAPH FxFx alternative sample.

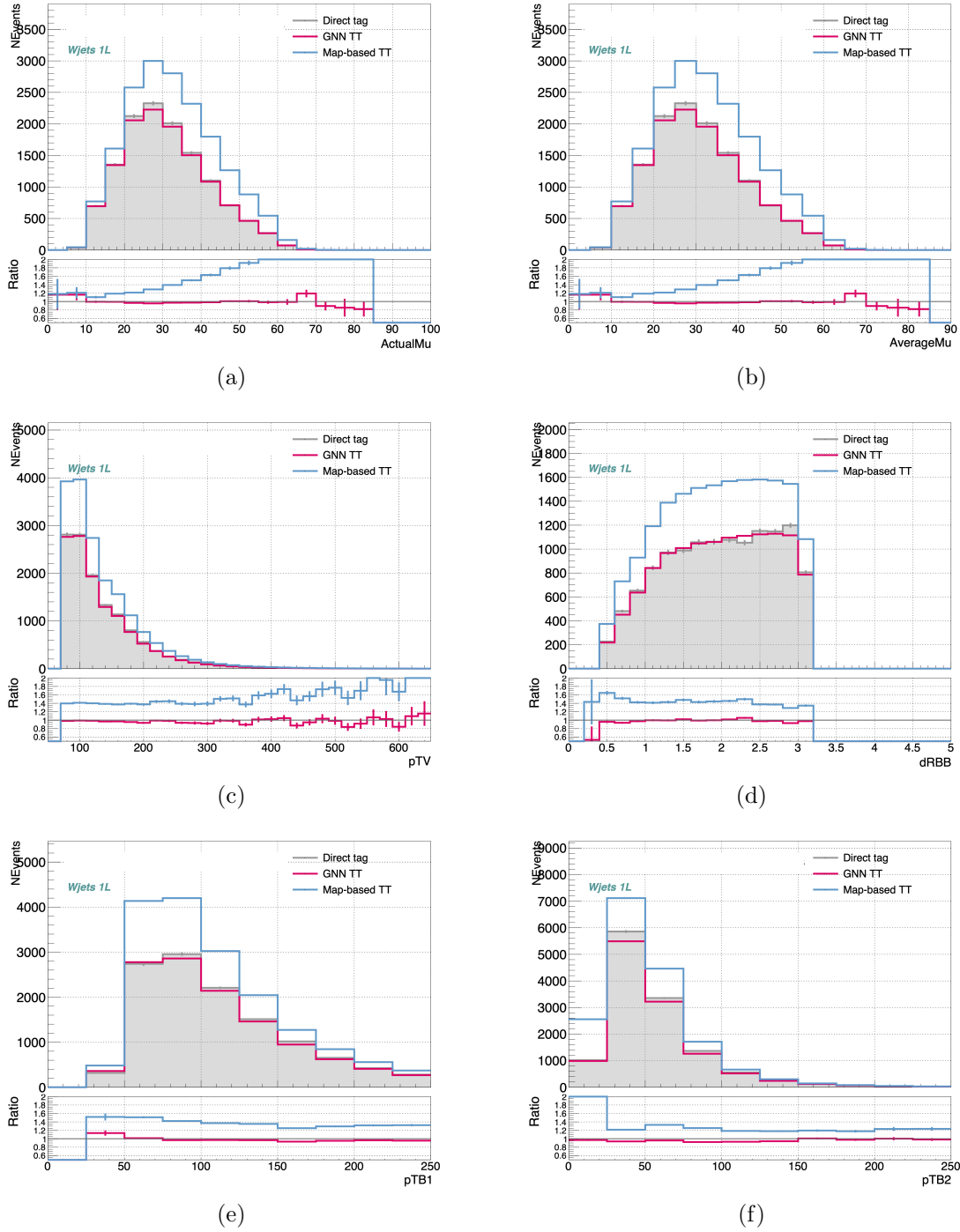


Figure 6.46: Examples of the plot monitoring the performance of the truth-flavour tagging method on the variables used to train the BDT. In this case, performance is monitored inclusively over all the W +jets events of the 1L channel.

Events	Variation	Truth flavour-tagging			Pass-or-fail flavour-tagging		
		$Z + cc/Z + bb$	$Z + cl/Z + bc$	$Z + bl/Z + bc$	$Z + cc/Z + bb$	$Z + cl/Z + bc$	$Z + bl/Z + bc$
Z + jets , VH, H → b $\bar{b}$ resolved, 0L	MadGraph FxFx	-7.6 ± 0.5 %	16.2 ± 1.1 %	7.1 ± 1.3 %	-14.6 ± 3.4 %	7.3 ± 10.7 %	11.2 ± 4.9 %
	1/2 μ_R & 1/2 μ_F	-4.6 %	2.5 %	3.4 %	-3.5 %	9.9 %	0.7 %
	1/2 μ_R & μ_F	-2.9 %	1.5 %	2.2 %	-1.5 %	1.3 %	1.6 %
	μ_R & 1/2 μ_F	-1.8 %	1.9 %	0.5 %	-0.9 %	6.0 %	-0.5 %
	μ_R & μ_F	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	μ_R & 2 μ_F	-0.9 %	-2.6 %	-0.7 %	-1.1 %	2.3 %	-0.7 %
	2 μ_R & μ_F	-0.8 %	0.4 %	2.7 %	-1.7 %	5.0 %	4.0 %
	2 μ_R & 2 μ_F	-1.5 %	-2.6 %	1.9 %	-2.0 %	4.5 %	3.1 %
	PDF+ α_s up	0.6 %	-0.8 %	-0.6 %	6.5 %	14.2 %	1.7 %
	PDF+ α_s down	-0.6 %	0.8 %	0.6 %	-6.7 %	-15.5 %	-1.8 %
	EWK Exp.	-0.1 %	-0.1 %	-0.0 %	-0.2 %	0.3 %	0.1 %
	EWK Mult.	0.0 %	-0.0 %	0.0 %	0.0 %	0.5 %	0.1 %
	Total	8.9 %	16.5 %	7.9 %	16.5 %	21.3 %	12.0 %

Table 6.15: Example displaying the impact of truth flavour-tagging on the calculation of relative acceptance uncertainties. The table shows the flavour composition uncertainties on 0L Z + jets events. Using truth flavour tagging reduces the total uncertainties on all the flavour ratios. Additionally, it sizeably reduces the MC statistical uncertainties associated to the MADGRAPH FxFx alternative sample.

6.12 Blinded Results

While this thesis is being concluded, the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis is in the final stages of its *blinded* validation. In the context of this measurement, *blinding* refers to a set of self-imposed rules that forbid monitoring real data in the most sensitive regions as well as pulls and constraints of NPs related to the signal, and that restrict the types of fit that can be performed while still in the designing stages of the analysis. These rules help reduce the *experimenter's bias*, *i.e.*, the unintended influence on a measurement towards prior results or theoretical expectations²³.

Real data is not monitored in bins where signal processes contribute more than 10% to the overall yield and in a $80 \text{ GeV} < m_{H_{\text{cand}}} < 140 \text{ GeV}$ window of the H_{cand} mass distribution. Additionally, data is not monitored in the upper bins of the BDT score distributions that collectively contain 60% of the predicted signal, as seen in Fig. 6.51, 6.52 and 6.53.

Different types of fits are used for various purposes. *Conditional* fits set all the POIs to 1 and the NPs related to the signals to their central values. *Unconditional* fits allow all parameters to vary during minimization. While data is blinded, the following types of fit are used:

- **conditional fits to observed data**, used to build the post-fit Asimov dataset;
- **unconditional fits to the post-fit Asimov dataset**, used to determine expected signal strengths, equivalent median gaussian significances, and upper limits;
- **conditional fits to the post-fit Asimov dataset**, used to study the pulls and constraints of the NPs unrelated to the signal.
- **unconditional fits to the pre-fit Asimov dataset**, used for preliminary estimates of the impact of groups of NPs on the expected signal strengths and their uncertainties.

I derived the results presented in the following sections respecting all the rules imposed for the current blinded status of the analysis.

Unblinded results, obtained by the team after the defense of this thesis and published in [2], are added for completeness in App. A.

6.12.1 $(W/Z)H, H \rightarrow b\bar{b}/c\bar{c}$ combined fit

Expected results for the $VH, H \rightarrow b\bar{b}$ process The expected best estimate of the signal strength $\hat{\mu}_{VH, H \rightarrow b\bar{b}}^{\text{exp}}$ of the $VH, H \rightarrow b\bar{b}$ process is:

$$\hat{\mu}_{VH, H \rightarrow b\bar{b}}^{\text{exp}} = 1.00 \pm 0.13; \quad (6.17)$$

²³An excellent, very short introduction to blinding techniques is found in [187].

	$\hat{\mu}_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$	Equivalent $Z_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$
0L	1.00 ± 0.20	4.8σ
1L	1.00 ± 0.19	5.3σ
2L	1.00 ± 0.22	4.5σ
Combination	1.00 ± 0.13	8.0σ

Table 6.16: Best estimates of the signal strength and equivalent median gaussian significance of the $VH, H \rightarrow b\bar{b}$ process, obtained when assigning one POI per lepton channel and when assigning only one POI shared across the channels. Sensitivities are similar in the three lepton channel, with 1L slightly surpassing the others.

corresponding to an expected significance of the signal over the background-only hypothesis of:

$$Z_{VH,H \rightarrow b\bar{b}}^{\text{exp}} = 8.0\sigma. \quad (6.18)$$

This result can be compared with the analysis of the $VH, H \rightarrow b\bar{b}$ process performed by the CMS collaboration [188], measuring

$$Z_{VH,H \rightarrow b\bar{b}}^{\text{exp, CMS}} = 5.6\sigma; \quad (6.19)$$

and with the former most significant result within the ATLAS Collaboration:

$$Z_{VH,H \rightarrow b\bar{b}}^{\text{exp, old ATLAS}} = 6.7\sigma, \quad (6.20)$$

measured in the previous resolved $VH, H \rightarrow b\bar{b}$ analysis [96].

This comparison underscores the great improvement in the sensitivity to the $VH, H \rightarrow b\bar{b}$ process brought about by the innovations introduced in this analysis and detailed in the previous sections. Performing the measurement on the same Run-2 dataset used in [96] (previously the world's most significant result), these innovations enable improving $Z_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$ by 20%.

Tab. 6.16 showcases the expected best estimates of the signal strengths of the $VH, H \rightarrow b\bar{b}$ process obtained when assigning one POI per lepton channel, together with the combined result, where only one POI is shared across the channels. The expected significance is similar in all the lepton channels, with 1L being the most significant and 2L the least.

Tab. 6.17 showcases the best estimates of the signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained when assigned separate POIs, corresponding to expected median gaussian significances of

$$Z_{WH,H \rightarrow b\bar{b}}^{\text{exp}} = 5.2\sigma \text{ and } Z_{ZH,H \rightarrow b\bar{b}}^{\text{exp}} = 5.7\sigma. \quad (6.21)$$

	$\hat{\mu}^{\text{exp}}$	Equivalent Z^{exp}
$WH, H \rightarrow b\bar{b}$	1.00 ± 0.19	5.2σ
$ZH, H \rightarrow b\bar{b}$	1.00 ± 0.17	5.7 σ
Combination	1.00 ± 0.13	8.0 σ

Table 6.17: The expected best estimates of the signal strength of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained when assigned separate POIs, along with the combined result where $VH, H \rightarrow b\bar{b}$ is assigned a single POI. If the unblinded results confirm the expected 5.2 σ significance of the $WH, H \rightarrow b\bar{b}$ signal over the background only hypothesis, they will mark the first observation of the process.

Both values exceed the 5 σ limit required to claim observation within the particle physics community. Most noticeably, the $Z_{WH, H \rightarrow b\bar{b}}^{\text{exp}}$ value indicates the first expected observation of the $WH, H \rightarrow b\bar{b}$ process.

Fig. 6.47 displays the best estimates of the signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained by assigning 15 POIs for final states featuring different p_{T}^V and jet multiplicities. Albeit the significance of the individual POIs is much reduced in this scenario, the increased granularity is useful to test new physics models predicting deviations from the SM at specific p_{T}^H values or jet multiplicities²⁴.

Expected results for the $VH, H \rightarrow c\bar{c}$ process The expected 95% CL median upper limit set on the signal strength of the $VH, H \rightarrow c\bar{c}$ process is:

$$\text{Exp. 95\% CL upper limit on } \mu_{VH, H \rightarrow c\bar{c}} = 11_{-3}^{+4}. \quad (6.22)$$

This result can be compared with the latest measurement of the $VH, H \rightarrow c\bar{c}$ process performed by the CMS collaboration [115], achieving

$$\text{Exp. 95\% CL upper limit on } \mu_{VH, H \rightarrow c\bar{c}}^{\text{CMS}} = 7.6_{-2.3}^{+3.4}, \quad (6.23)$$

and with the former most stringent limit set within the ATLAS Collaboration, namely

$$\text{Exp. 95\% CL upper limit on } \mu_{VH, H \rightarrow c\bar{c}}^{\text{old ATLAS}} = 31_{-8}^{+12}, \quad (6.24)$$

measured in the previous $VH, H \rightarrow c\bar{c}$ analysis [95].

²⁴The predictions of sets of new-physics models are factored-in by means of Effective Field Theory operators with dimensions larger than 4, multiplied by dimensional scalar coefficients and added to the SM lagrangian density. Different operators induce different deviations of the SM predictions, and granular signal strength measurements help constrain their coefficients.

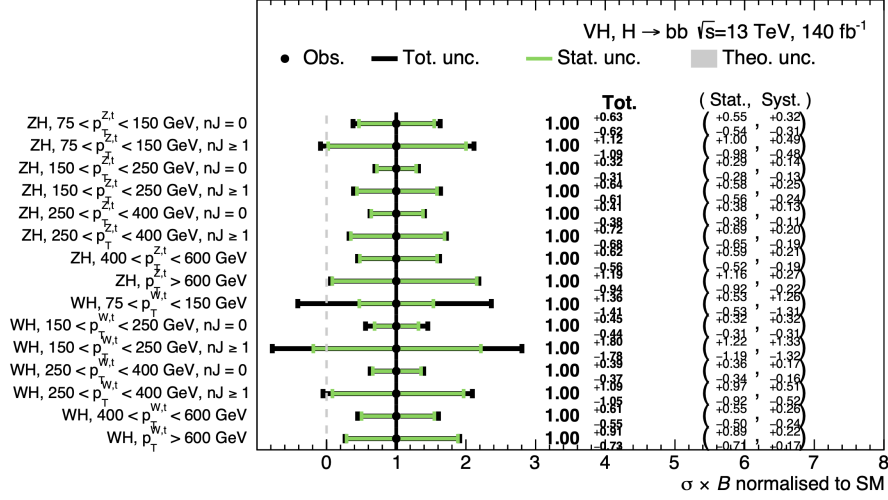


Figure 6.47: Best estimates of the signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained from an unconditional fit to the pre-fit Asimov dataset when assigning 15 POIs for final states featuring different p_T^V and jet multiplicities

As for the $VH, H \rightarrow b\bar{b}$ process, this comparison underscores the impact of the innovations introduced in this analysis compared to its predecessor, which reduces by 65% the expected upper limit on the $VH, H \rightarrow c\bar{c}$ process by using the same Run-2 proton-proton collision dataset.

Additionally, it emphasizes the excellent expected limit that the colleagues of the CMS Collaboration were able to set with their measurement, 30% more stringent than the one set by this analysis. One of the main factors that enabled such a result was the application of the **ParticleNet** algorithm [189] to merged jets. **ParticleNet** is a $H \rightarrow c\bar{c}$ tagging algorithm based on GNNs, operating on large- R jets to identify those containing c -quark pairs stemming from the decay of a Higgs-boson. Preliminary studies performed within the ATLAS Collaboration indicate that using the **GN2** algorithm to identify small- R c -jets and the **GN2X** algorithm [117] to identify large- R jets containing c -quark pairs stemming from the decay of a Higgs-boson has the potential to bring the sensitivity of the ATLAS experiment to the $VH, H \rightarrow c\bar{c}$ process at least to the same level achieved by the CMS Collaboration.

Fig. A.2 showcases the expected 95% CL median upper limits on the $VH, H \rightarrow c\bar{c}$ process when assigning one POI per lepton channel, together with the combined result. The tree lepton channels have similar sensitivities, with 0L contributing the most to the limit.

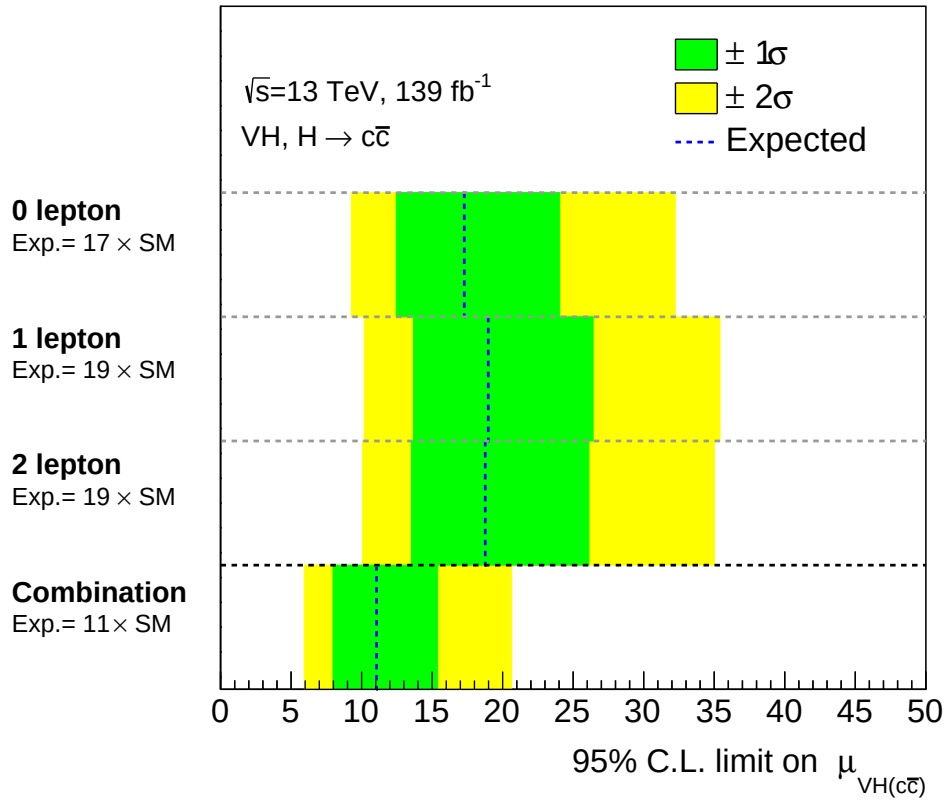


Figure 6.48: Expected 95% CL median upper limits on the $VH, H \rightarrow c\bar{c}$ process when assigning one POI per lepton channel, together with the combined result. The tree lepton channels have similar sensitivities, with 0L contributing the most to the limit.

Blinded fit validation A breakdown of the contributions of the different sources of uncertainties to the total uncertainties on the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ signal strength measurements, estimated from an unconditional fit to the pre-fit Asimov dataset, is displayed in Tab. 6.18.

The contribution of a group of uncertainties is determined through a fit where all NPs belonging to that specific group are held fixed at their central values. This fit yields a new best-fit value $\hat{\mu}'$ and its associated uncertainty $\sigma_{\hat{\mu}'}$. The difference between $\sigma_{\hat{\mu}'}$ and the nominal uncertainty $\sigma_{\hat{\mu}}$ defines the contribution, expressed as:

$$\text{contribution} = \sqrt{\sigma_{\hat{\mu}}^2 - \sigma_{\hat{\mu}'}^2} \quad (6.25)$$

The contribution of the statistical uncertainties is computed with a fit where all NPs are fixed except for the floating normalizations. The total systematic uncertainty is the quadratic difference between the statistical and total uncertainties. The data statistical contribution is obtained by fixing all the NPs. The MC statistical contributions is obtained by fixing all the γ parameters. The sum in quadrature of individual contributions may not match the total uncertainty due to correlations among the NPs.

Statistical and systematic sources contribute similarly to the uncertainties of both the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ processes, with statistical sources contributing $\sim 15\%$ more than systematic. The statistical component is dominated by the limited size of the Run-2 dataset, contributing almost three times more than the floating normalisations. The systematic uncertainties are further broken down into theoretical and experimental uncertainties. Uncertainties related to the modelling of $V + \text{jets}$ events dominate the theoretical contributions, with $W + \text{jets}$ and $Z + \text{jets}$ contributing similarly to $VH, H \rightarrow b\bar{b}$, and $Z + \text{jets}$ dominating $VH, H \rightarrow c\bar{c}$. The uncertainties related to small- R jets contribute the most to the experimental uncertainties.

A ranking of the impact on $\hat{\mu}^{\text{exp}}$ of the fifteen most impactful NPs, estimated from an unconditional fit to the pre-fit Asimov dataset, is displayed in Fig. 6.49 for the $VH, H \rightarrow b\bar{b}$ process and Fig. 6.50 for the $VH, H \rightarrow c\bar{c}$ process.

These impacts are obtained by fixing the values of the NPs to their $\pm 1\sigma$ variations and computing their impact $\Delta\hat{\mu}$ on the POIs as

$$\Delta\hat{\mu} = \hat{\mu} - \hat{\mu}'. \quad (6.26)$$

$\Delta\hat{\mu}$ is then used to rank the NPs.

The NPs impacts are indicated in the figures via the coloured boxes, using darker shades for the $+1\sigma$ and lighter shades for the -1σ variations. These boxes refer to the scale written in grey on the upper side of the plot, and different colours are used to identify experimental uncertainties, normalisation and relative acceptance uncertainties (Bkg. Modelling) and shape uncertainties on the background templates (Bkg. Shape), and signal modelling uncertainties. The corresponding NP constraints are shown with black markers (red for the

Source of Uncertainty		$VH, H \rightarrow b\bar{b}$	$VH, H \rightarrow c\bar{c}$
Total		0.127	5.1
Statistical		0.096	3.8
Systematic		0.085	3.3
Statistical uncertainties			
Data statistical		0.088	3.5
top($e\mu$) CR		0.011	0.13
Floating normalisations		0.030	1.2
Theoretical and modelling uncertainties			
Signal		0.021	0.2
W + jets		0.037	1.1
Z + jets		0.035	1.5
$t\bar{t}$		0.013	0.6
Diboson		0.029	0.6
Multijet		0.005	0.7
Single top		0.008	0.1
MC statistical		0.020	1.4
Experimental uncertainties			
Small- R jets		0.027	0.8
Large- R jets		0.009	0.2
Leptons		0.004	0.1
E_T^{miss}		0.008	0.2
Small- R jet flavour-tagging calibration	b -jets	0.015	0.3
	c -jets	0.008	0.7
	light-flavour jets	0.002	0.7
	extrapolation	< 0.001	< 0.1
Variable- R jet flavour-tagging calibration	b -jets	0.004	0.1
	c -jets	0.001	< 0.1
	ljets	< 0.001	< 0.1
	extrapolation	0.001	< 0.1
PU		0.005	0.1
Luminosity		0.007	< 0.1

Table 6.18: Breakdown of the contributions to the uncertainty on $\hat{\mu}^{\text{exp}}$ for the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ signal strength measurements, estimated from an unconditional fit to the pre-fit Asimov dataset. The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. In cases where the upward and downward systematic variations have different values, the mean of the absolute values is shown. The most prominent contribution of every set of uncertainties is emboldened.

floating normalizations), and refer to the scale written in black on the lower side of the plot. Since the fit is to the pre-fit Asimov dataset, no pull is observed, as the dataset is built using the central values of the POIs and NPs.

The NPs that have the largest impact on $\hat{\mu}_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$ (Fig. 6.49) are relative acceptance uncertainties on the distribution of $V + \text{jets}$ events in the low and high- ΔR CRs. Both of these uncertainties are related to extrapolating from a CR, and they are applied in the SR only, which explains why they have such a large impact. Following is the experimental uncertainty on the large- R jet energy scales related to the jets' flavour composition for the signal process. Also in this case, the uncertainty is directly applied to the signal, explaining its large impact. NPs governing the shapes of the diboson distribution are also highly ranked, due to the sizeable correlations of the diboson NPs with the signal strengths, owing to the similarities in the final states of the processes.

When it comes to $\hat{\mu}_{VH,H \rightarrow c\bar{c}}^{\text{exp}}$, shape effects on the dominant $t\bar{t}$ and $Z + cc$ background are more important, probably due to the smallness of the $VH, H \rightarrow c\bar{c}$ signal. Most of the other parameters highly ranked refer to the relative acceptance uncertainties of $V + \text{jets}$ components and the $V + \text{jets}$ floating normalizations. The experimental NP ranked the highest relates to the calibration of the flavour-tagging efficiencies on light-flavour jets, an observation still under investigation.

Monitoring studies are ongoing to pin down the sources of constraints on the uncertainties of the NPs appearing in these plots.

The picture emerging from the breakdown of the uncertainties and the ranking of the NPs highlights the importance of the aspects of this analysis in which I was most directly involved.

$V + \text{jets}$ modelling uncertainties contribute the most to the total uncertainty. Among these, as confirmed by the ranking plots, relative acceptance uncertainties are very significant. Carefully quantifying their size is hence crucial for the accuracy of the analysis results, a task to which I largely contributed and that, in this measurement, benefitted from the application of truth-flavour tagging based on the GNNs that I trained.

In the previous rounds of the $VH, H \rightarrow b\bar{b}$ boosted and resolved analyses [96, 105], MC statistical uncertainties, parameterized in the likelihood via the γ parameters, contributed more to the total uncertainty than any modelling uncertainties of individual background processes. In this measurement, their contribution to the uncertainty on $\hat{\mu}_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$ is smaller than the modelling uncertainties on $V + \text{jets}$ and diboson events. The reduced contribution of MC statistical uncertainties is explained by the increased statistics of the SHERPA 2.2.11 samples and the application of GNN-based truth flavour-tagging.

Finally, in the previous $VH, H \rightarrow c\bar{c}$ analysis [95], dedicated systematic uncertainties were introduced to deal with the mismodellings introduced by map-based truth flavour-tagging, contributing 3.3 and 1.7 to the total uncertainties, compared, for instance, to the 7.0 contribution of modelling uncertainties of $Z + \text{jets}$ events. These NPs ranked high, two of

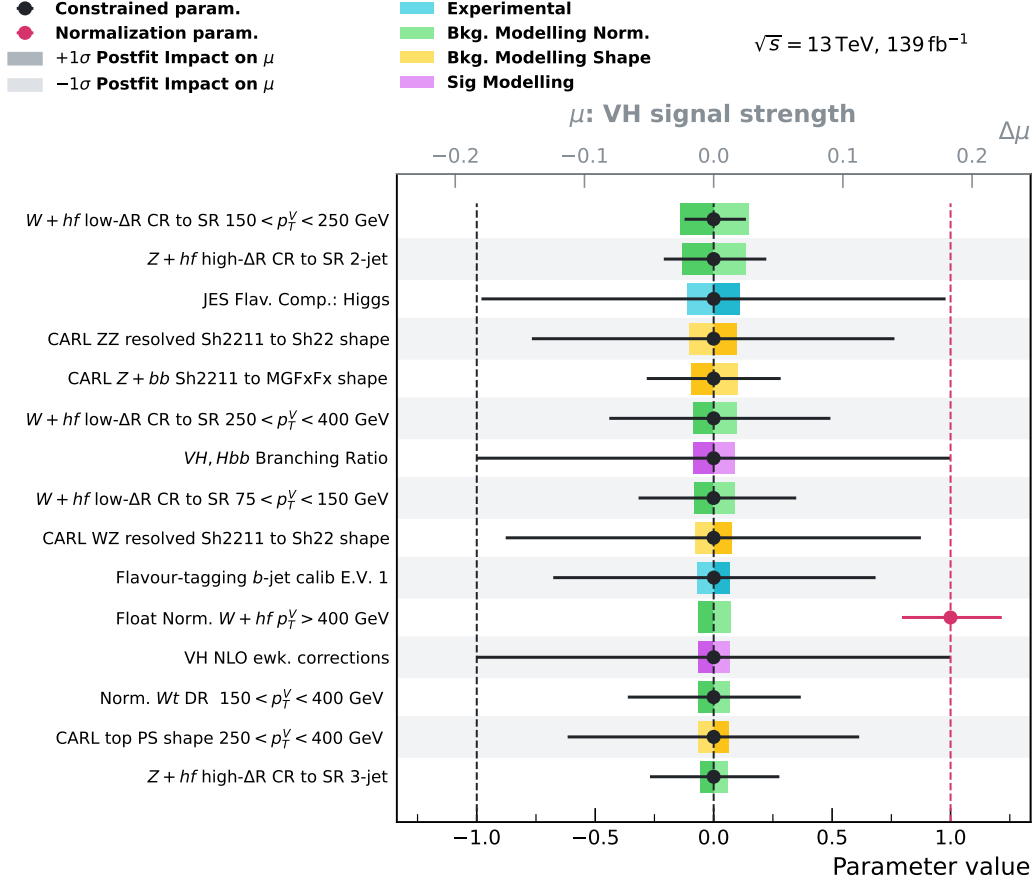


Figure 6.49: The impact of individual NPs on the best estimate of the $\hat{\mu}_{VH,H \rightarrow b\bar{b}}^{\text{exp}}$ parameter. The impacts are indicated via the coloured boxes, using darker shades for the $+1\sigma$ and lighter shades for the -1σ variations. These boxes refer to the scale written in grey on the upper side of the plot. Different colours are used to identify experimental uncertainties, normalisation and relative acceptance uncertainties (Bkg. Modelling) and shape uncertainties on the background templates (Bkg. Shape), and signal modelling uncertainties. The corresponding NP constraints are shown with black markers (red for the floating normalizations), and refer to the scale written in black on the lower side of the plot. Since the fit is to the pre-fit Asimov dataset, no pull is observed, as the dataset is built using the central values of the POIs and NPs.

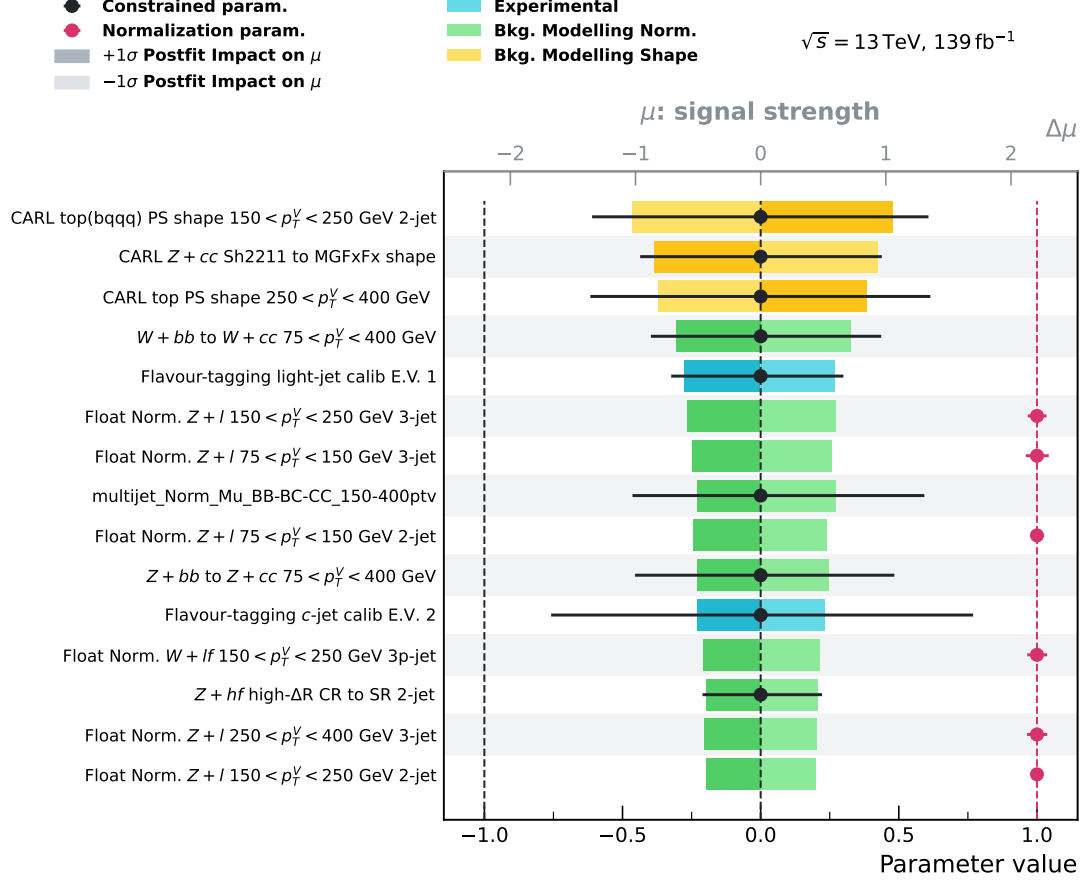


Figure 6.50: The impact of individual NPs on the best estimate of the $\hat{\mu}_{VH,H \rightarrow c\bar{c}}^{\text{exp}}$ parameter. The impacts are indicated via the coloured boxes, using darker shades for the $+1\sigma$ and lighter shades for the -1σ variations. These boxes refer to the scale written in grey on the upper side of the plot. Different colours are used to identify experimental uncertainties, normalisation and relative acceptance uncertainties (Bkg. Modelling) and shape uncertainties on the background templates (Bkg. Shape), and signal modelling uncertainties. The corresponding NP constraints are shown with black markers (red for the floating normalizations, and refer to the scale written in black on the lower side of the plot). Since the fit is to the pre-fit Asimov dataset, no pull is observed, as the dataset is built using the central values of the POIs and NPs.

them appearing among the fifteen highest-ranked. In this analysis, the much-improved parametrization of the jet flavour-tagging efficiencies achieved by the GNN models eliminates the need to introduce NPs related to truth flavour-tagging.

6.12.2 Expected results for the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ process

The expected median equivalent gaussian significances of the $VZ, Z \rightarrow b\bar{b}$ and $VZ, Z \rightarrow c\bar{c}$ processes are extracted from an unconditional fit to the post-fit Asimov dataset, and are showcased in Tab. 6.19. This measurement serves the purpose of validating the $VH, H \rightarrow b\bar{b}/c\bar{c}$ analysis strategy by measuring the much more frequent $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ processes, which, by virtue of their similar final states, are used as SM candles for $VH, H \rightarrow b\bar{b}/c\bar{c}$.

Most noticeably, the expected 5.5σ significance of the $VZ, Z \rightarrow c\bar{c}$ signal over the background only hypothesis indicates the first expected observation of the process within ATLAS.

	Equivalent $Z_{VZ, Z \rightarrow b\bar{b}}^{\text{exp}}$	Equivalent $Z_{VZ, Z \rightarrow c\bar{c}}^{\text{exp}}$
0L	12.6σ	4.4σ
1L	7.3σ	2.8σ
2L	9.6σ	3.7σ
Combination	16.2σ	5.5σ

Table 6.19: The expected median equivalent gaussian significances of the $VZ, Z \rightarrow b\bar{b}$ and $VZ, Z \rightarrow c\bar{c}$ processes, obtained when assigning separate POIs in the different lepton channels, along with the combined results where $VZ, Z \rightarrow b\bar{b}$ and $VZ, Z \rightarrow c\bar{c}$ are assigned a single POI each. If the unblinded results confirm the expected 5.5σ significance of the $VZ, Z \rightarrow c\bar{c}$ signal over the background only hypothesis, they will mark the first observation of the process within ATLAS.

6.12.3 Postfit Distributions

In the previous sections, various data-MC pre-fit closure plots were shown. Their post-fit distributions from the conditional fit to the observed data are displayed here.

Fig. 6.51 shows the post-fit BDT distributions in data and MC samples belonging to the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ SRs of the resolved $VH, H \rightarrow b\bar{b}$ regime. Fig. 6.52 displays those in the 2-jet $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ TL+TT SRs of the $VH, H \rightarrow c\bar{c}$ regimes; and Fig. 6.53 those in the $400 \text{ GeV} < p_T^V < 600 \text{ GeV}$ SRs of the boosted regime.

Fig. 6.54 exemplifies the post-fit distributions in the high- ΔR CRs, Fig. 6.55 those in the top(bc) CRs, Fig. 6.56 those in the top($e\mu$) CRs, Fig. 6.23 those in the the low- ΔR CRs,

Fig. 6.58 those in the V +light CRs, and Fig. 6.59 those in the top CRs of the boosted regime.

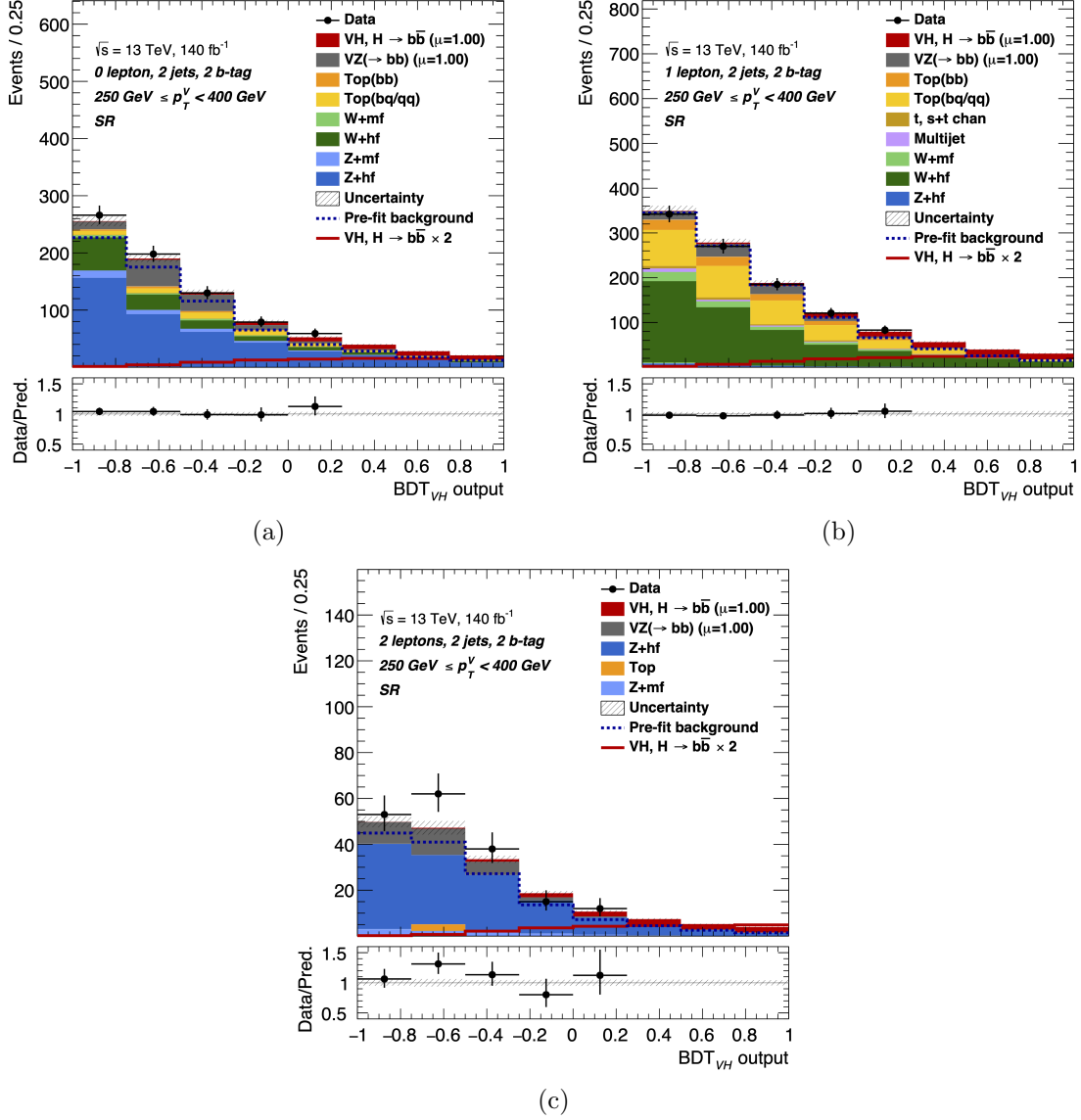


Figure 6.51: Data and post-fit MC distributions of the output score of the BDT evaluated on events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ SR in the resolved $VH, H \rightarrow b\bar{b}$ regime.

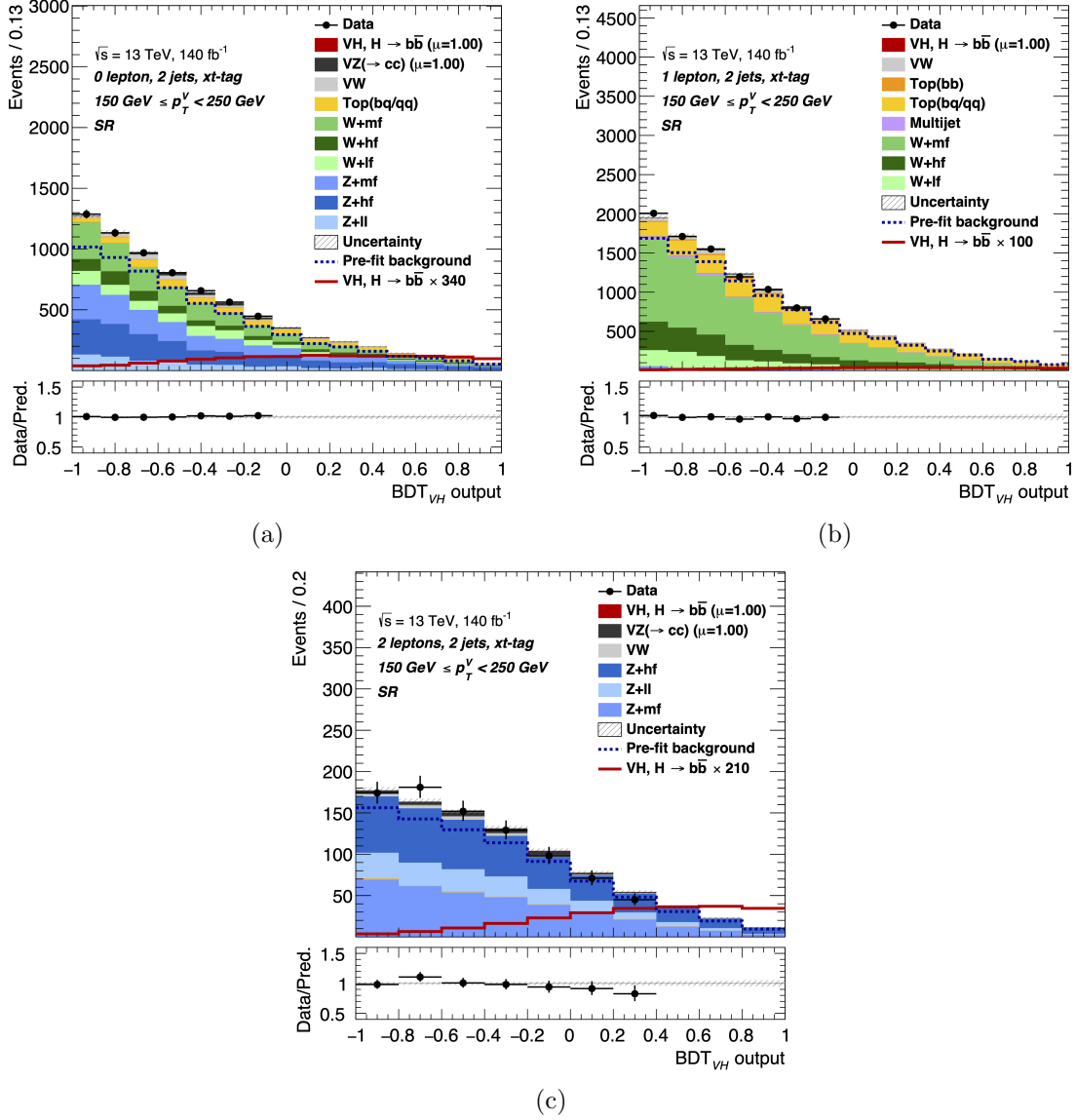


Figure 6.52: Data and post-fit MC distributions of the output score of the BDT evaluated on events entering the 2-jet $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ SRs in the $VH, H \rightarrow c\bar{c}$ regime. *xt-tag*, a name used in the legend, indicates TT+TL *c*-tag regions.

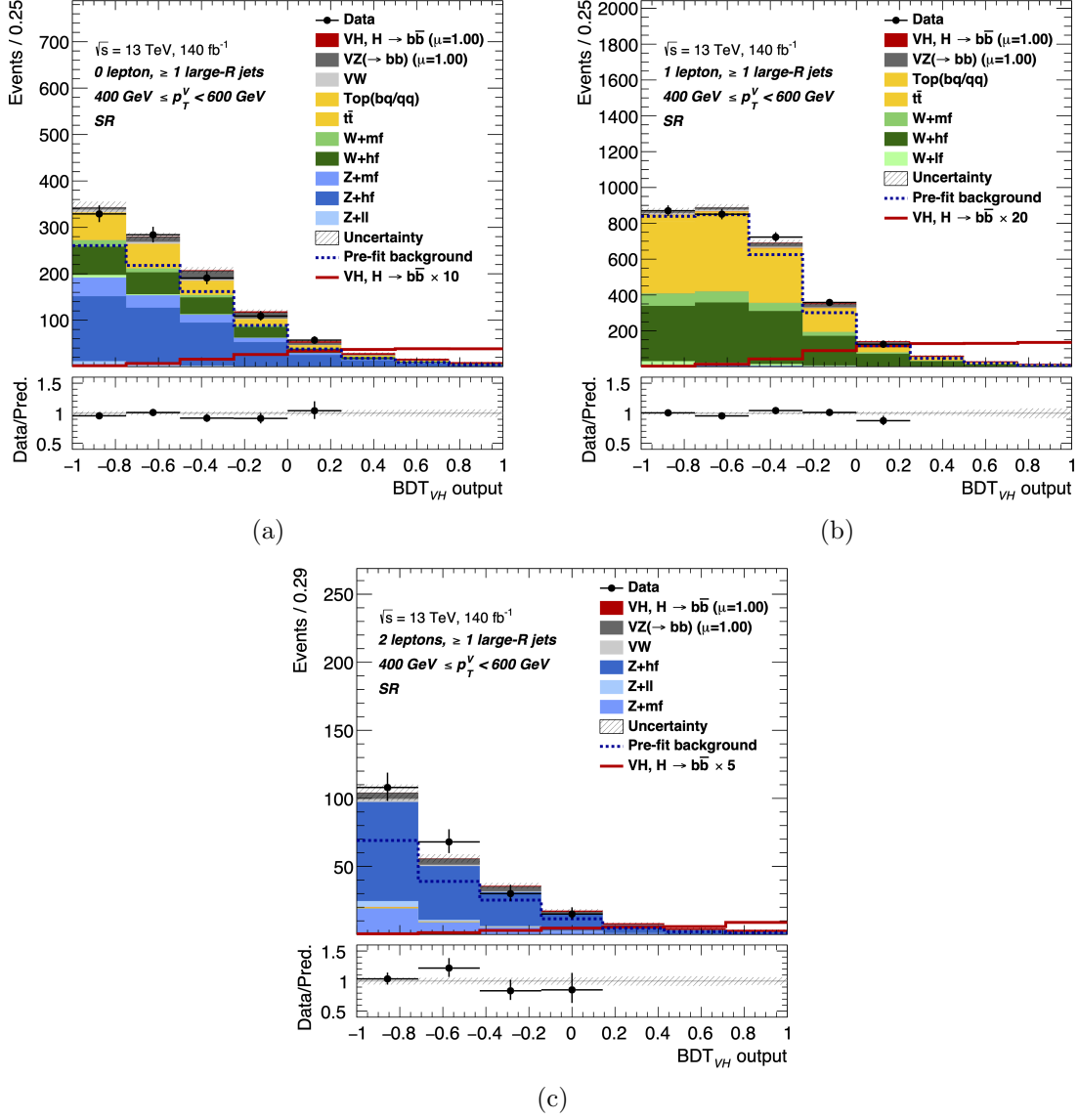


Figure 6.53: Data and post-fit MC distributions of the output score of the BDT evaluated on events entering the SRs in the boosted $VH, H \rightarrow b\bar{b}$ regime.

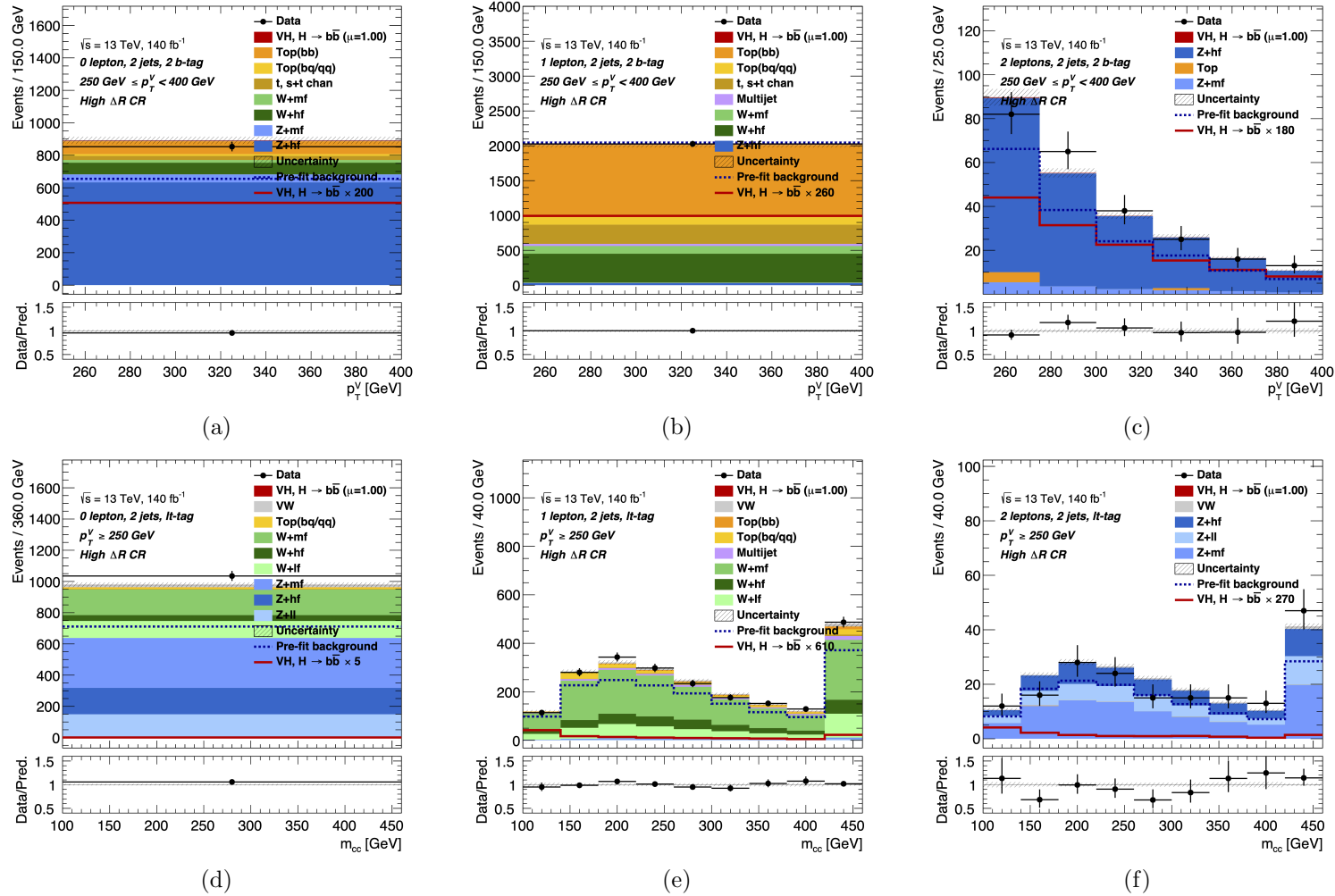


Figure 6.54: (a, b and c) Data and post-fit MC distributions of events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ high- ΔR CR in the resolved $VH, H \rightarrow b\bar{b}$ regime. (d, e, f) Data and post-fit MC distributions of events entering the 2-jet $p_T^V > 250 \text{ GeV}$ LT c -tag high- ΔR CR in the $VH, H \rightarrow c\bar{c}$ regime. The regions are enriched in V +jets and top(bb) events. The distributions that are displayed are those used as templates in the fit. The trend in the ratio was observed in the data also in the previous

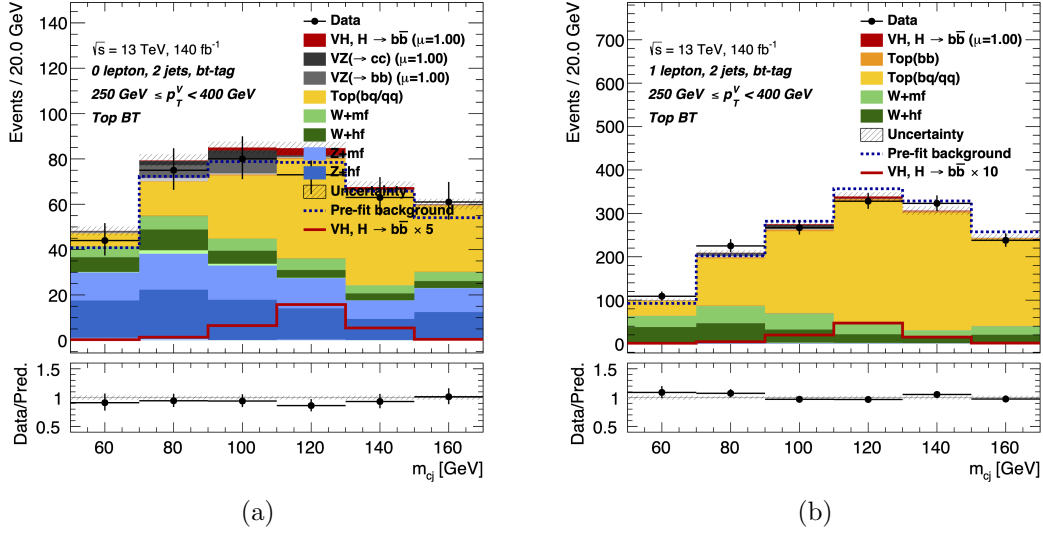


Figure 6.55: Data and post-fit MC distributions of events entering the 2-jet $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ top(bc) CR. The regions are enriched in top(bq) and top(qq) events. The distributions that are displayed are those used as templates in the fit.

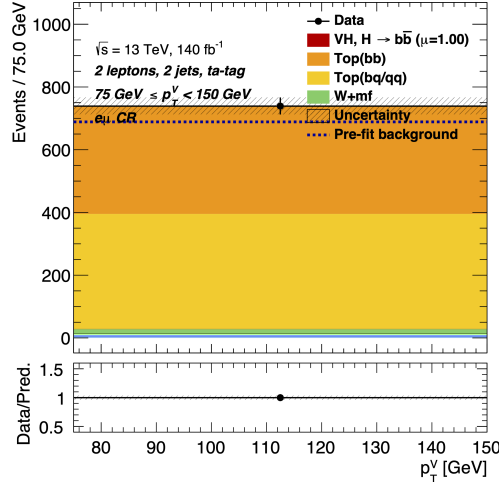


Figure 6.56: Data and post-fit MC distributions of events entering the 2-jet $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ top($e\mu$) CR in the $VH, H \rightarrow c\bar{c}$ regime. The region is enriched in top(bb) and top(bq) events. The distribution that is displayed is used as a template in the fit.

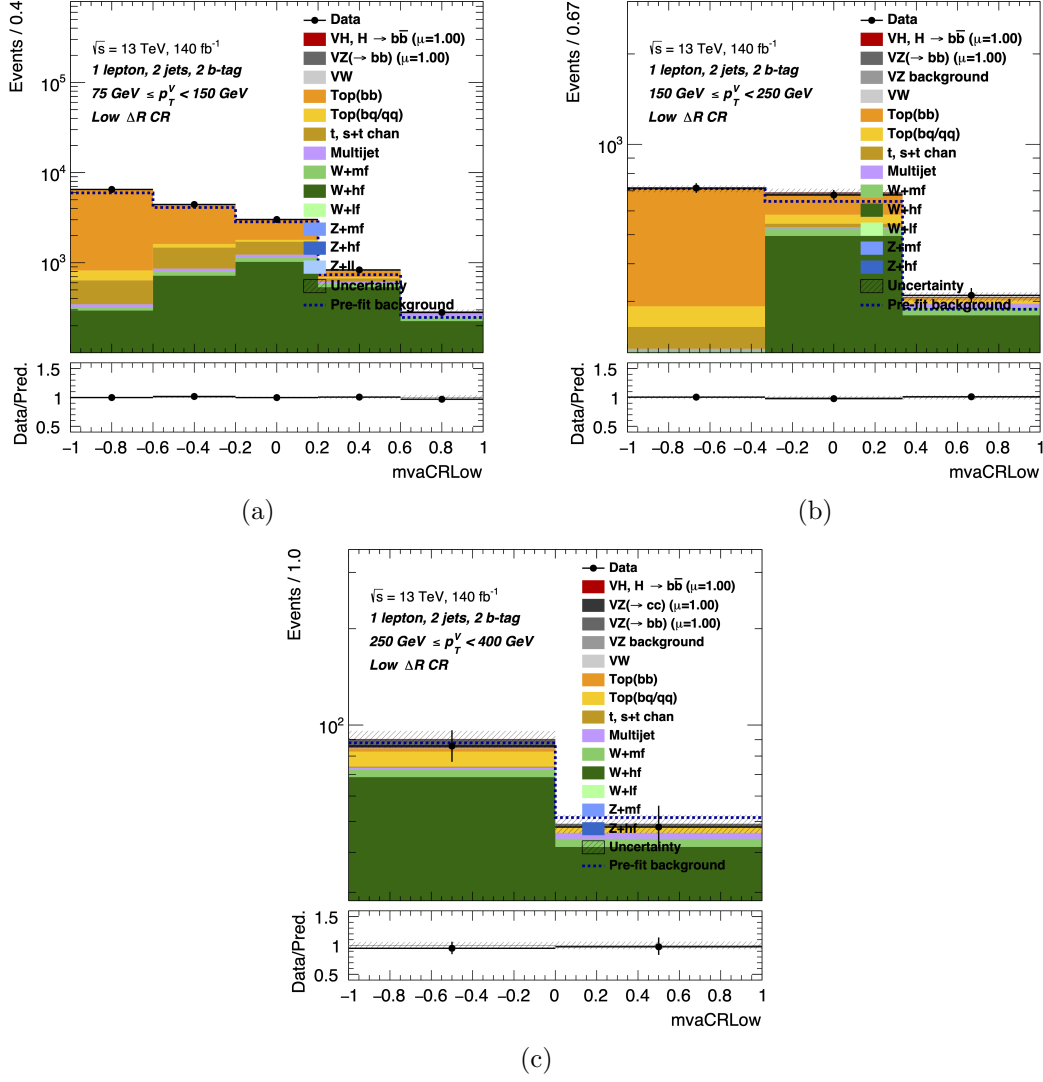


Figure 6.57: Data and post-fit MC distributions of events entering the 2-jet $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ low- ΔR CR. In this region, a BDT is trained to separate $W + hf$ events from the other processes. As a result, the bins with a higher score of the BDT's output distribution are very pure in $W + hf$ events. The distributions that are displayed are those used as templates in the fit.

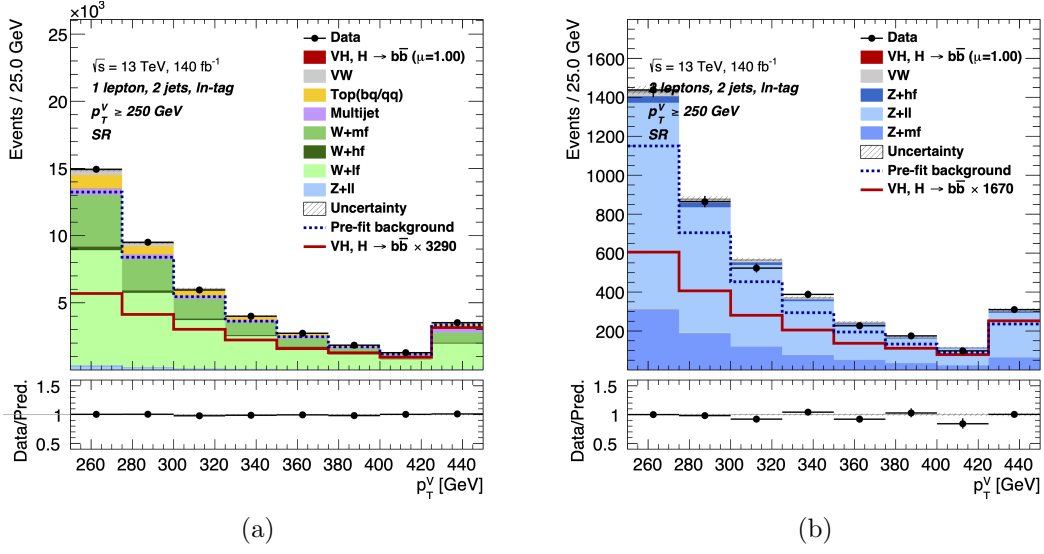


Figure 6.58: Data and post-fit MC distributions of events entering the 2-jet $p_T^V > 250$ GeV V +light-flavour jets CR. The regions are enriched in V +light-flavour jets events. The distributions that are displayed are those used as templates in the fit.

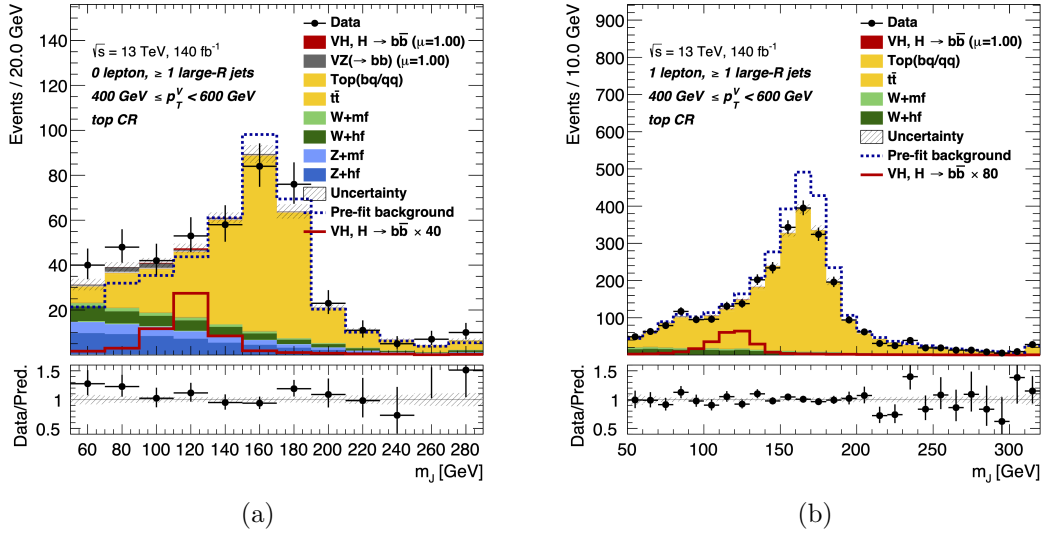


Figure 6.59: Data and post-fit MC distributions of events entering the $400 \text{ GeV} < p_T^V < 600 \text{ GeV}$ Top CRs of the boosted $VH, H \rightarrow b\bar{b}$ analysis. The regions are enriched in $t\bar{t}$ events. The distributions that are displayed are those used as templates in the fit.

The data/MC ratios in the lower panel of the plots demonstrate an excellent description of the data in most of the regions. The only exception is found in the 2L channel of the $VH, H \rightarrow c\bar{c}$ regime, where the ratio exhibits a clear trend (Fig. 6.52c). The opposite trend is observed in the distribution of the mass of H_{cand} in the corresponding high- ΔR CR, displayed in Fig. 6.60b. A similar mismodelling was observed in the previous $VH, H \rightarrow c\bar{c}$ analysis [95] (Fig. 6.60c). In that context, it was addressed by refining the binning of the H_{cand} mass and identifying two consecutive bins most likely affected by a downwards fluctuation of data. Then, the compatibility of the post-fit distributions obtained in the mismodelled region with an unconditional and a conditional fit to the post-fit Asimov dataset was performed. Observing compatibility between the two resulting distributions proved that the signal templates were not used to correct for the mismatch, therefore no bias was being introduced in the measurement of the POIs. A similar validation could be performed in the context of this measurement.

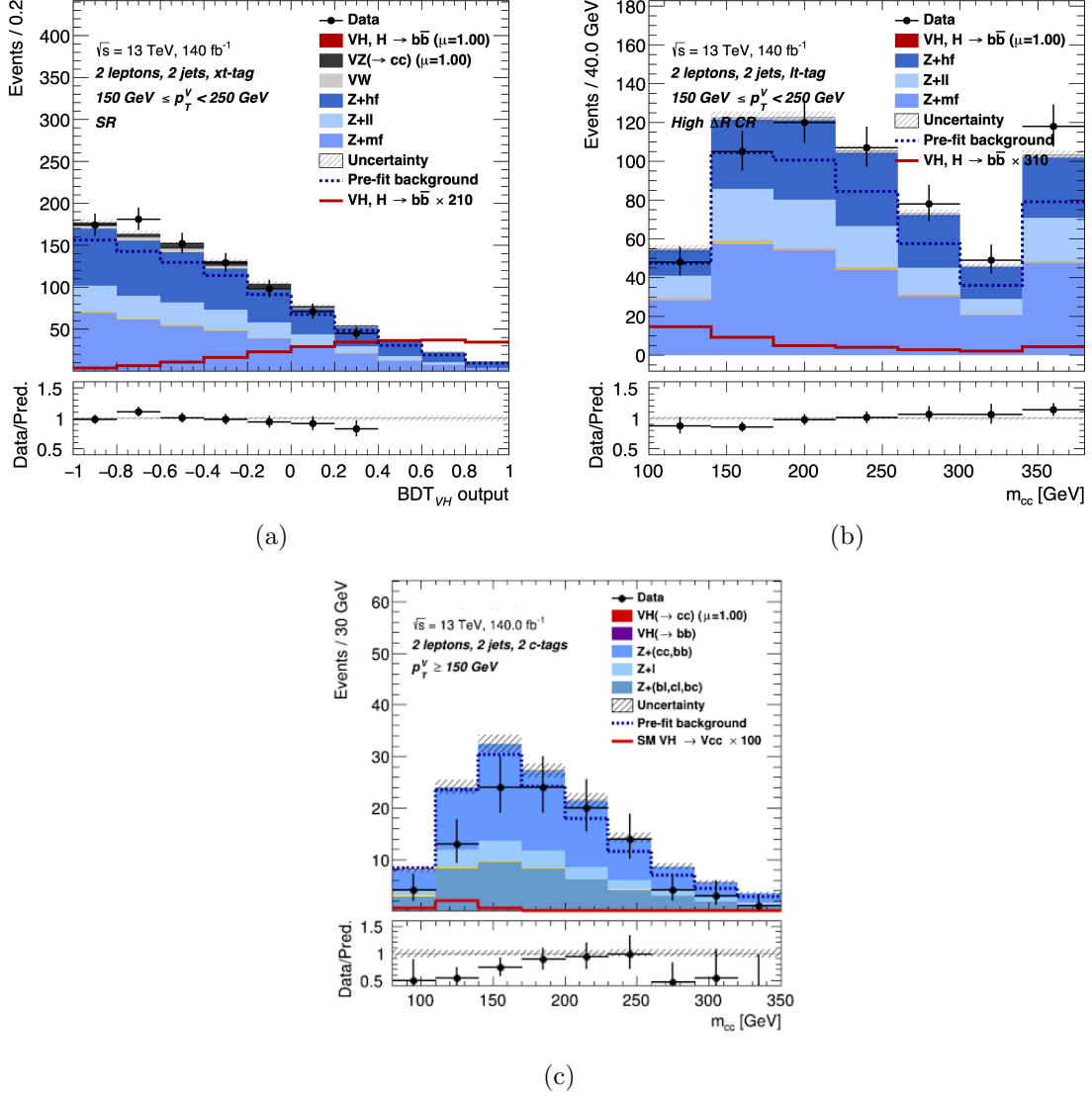


Figure 6.60: Data and conditional to-data post-fit MC distributions of the output score of the BDT evaluated on events entering the 2-jet 150 GeV < p_T^V < 250 GeV SRs in the VH, H → c $\bar{c}$ regime (a), of the H_{cand} mass distribution entering the corresponding high-ΔR CR (b), and data and unconditional to-data post fit MC distribution of the H_{cand} mass from the in the p_T^V > 150 GeV high-ΔR CR of the previous VH, H → c $\bar{c}$ analysis [95]. The trend in the data-MC ratio was observed already in [95].

Conclusions

In 1608, Hans Lippershey, a Dutch eyewear manufacturer, realized that two lenses, placed at the right distance, allowed the magnification of images.

Soon later, on the 14th of April 1611, on the occasion of a great celebration organized on the highest point of the Gianicolo hill in Rome to honour his co-optation in the recently born Accademia dei Lincei, Galileo Galilei demonstrated to the scientists of his time the potential of his upgraded version of Lippershey's invention.

First, he pointed it towards the church of San Giovanni in Laterano, where it was possible to read the inscription carved in its stone by Pope Sisto V on the first year of its pontificate. Then, he pointed it in the direction of the Tuscolan hills, 25 km away, where prince Altamps had a villa, the smallest windows of which everyone could see through Galileo's instrument.

But all these tricks were nothing more than an interlude of the real spectacle awaiting those patient enough to wait for the Sun to set. That night, those who looked through Galileo's lenses, pointed upwards into the dark sky, could see the satellites of Jupiter, and realized that Saturn was not a star.

On that occasion, the prince Federico Cesi, who had organised the celebration, rebaptized the instrument, which Galileo had called *perspicillum*, with a neologism derived from Greek: telescope, a crasis of the terms $\tau\eta\lambda\epsilon$ (*tēle*) and $\sigma\kappa\omicron\pi\epsilon\omega$ (*skopeo*), literally “*which sees far*”.

Galileo would put the telescope to good use: by discovering Venus's phase cycle, analogous to the Moon's, he concluded that the planet orbited around the Sun, much like the Moon orbits Earth. In doing so, he falsified the ptolemaic model of the solar system (Fig. 6.61). He also used it to discover the four galilean satellites of Jupiter, Sun spots, the granularity of the Milky Way comprising countless stars, and the details of the surface of the Moon.

The instrument would be passed on to subsequent generations of scientists, many of them introducing their own upgrades. In such a way, the history of Science unfolded, through the eyes of Kepler, Huygens, Cassini, Newton, and countless others.

In 2018, the Planck Collaboration released an overview of the cosmological legacy of its results, focusing on observing the anisotropies of the cosmic microwave background at microwave and infrared frequencies [190]. The Collaboration harnessed data measured by two



Figure 6.61: Galileo Galilei using his telescope to challenge the ptolemaic model of the solar system.

telescopes mounted on a space observatory stationed at the Earth/Sun L2 point from 2009. These two telescopes, direct descendants of the one Galileo first used to study the solar system, precisely measured the first photons ever to navigate the Cosmos when it first became transparent. *There is nothing further away to be seen with light*: before then, everything was opaque.

Such is the scientific method, which proceeds by developing new instruments to falsify its most established models and stare into the unseen landscapes at the edges of the unknown²⁵.

The research documented in this thesis aligns with such tradition.

In the ATLAS experiment at the LHC, we witness collisions between particles at the highest energies ever achieved by humankind. Much like Galileo did from the highest point of the Gianicolo hill, when he pointed his telescope into the sky on a starry night in the spring of 1611, we stare into the unknown.

Our instrument is the ATLAS detector, described in Chap. 2, one of the most sophisticated scientific apparatus ever built by humans; the model that we want to falsify is the SM, introduced in Chap. 1, arguably the most successful theory in the history of Science.

We have good reasons to believe that new insights may be attained by studying the properties of the Higgs boson, the last fundamental particle discovered. Among these reasons, the one that I consider the most compelling is that most parameters of the SM listed in Sec. 1.2 are linked to it. Many, namely the Higgs boson self-coupling, the nine Yukawa couplings

²⁵A figure that symbolizes this process is displayed in Fig. 62.

of the Higgs boson with the fermions, and the coupling of the Higgs boson with the W and Z bosons are unobserved as of today or measured with coarse precision. Refining such precision is crucial to detect possible deviations from the SM predictions, or to restrict the pool of theories beyond the SM that are still compatible with observations.

Some of these properties, most noticeably the y_b and y_c Yukawa couplings, influence the rates at which the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ processes occur and can be analysed using, among many others, the instrument of jet flavour-tagging algorithms. The most recent measurement of the $VH, H \rightarrow b\bar{b}/c\bar{c}$ processes, currently being performed within the ATLAS Collaboration, was presented in Chap. 6. Using **DL1r**, a DNN-based jet flavour-tagging algorithm developed in 2019, the measurement, still in the final stages of its validation, is expected to yield the most precise measurement of the signal strength of the $VH, H \rightarrow b\bar{b}$ process, namely $\mu_{VH, H \rightarrow b\bar{b}}^{\text{exp}} = 1.00 \pm 0.13$, corresponding to a significance of the $VH, H \rightarrow b\bar{b}$ signal over the background-only hypothesis of 8σ ; and to provide the most stringent 95% CL upper limit on the signal strength of the $VH, H \rightarrow c\bar{c}$ process set by the ATLAS Collaboration, namely 11_{-3}^{+4} . It is also expected to mark the first observation of the $WH, H \rightarrow b\bar{b}$ process, with a significance of 5.2σ over the background-only hypothesis; and the first observation of the $VZ, Z \rightarrow c\bar{c}$ process by the ATLAS Collaboration, with an expected significance of 5.5σ .

In the context of this measurement, a novel event weighting technique based on the parametrization of the jet flavour-tagging efficiencies using GNNs, described in Chap. 5, marked its inaugural application. It allows for better usage of the simulated data when producing the templates of the dominant background processes entering the final fit, when deriving relative acceptance uncertainties, and when training the BDTs used to enhance the separation of the signal from the background processes and the CARL models used to estimate shape uncertainties linked to the choice of the MC generators. Previous measurements of the $VH, H \rightarrow c\bar{c}$ and $VH, H \rightarrow b\bar{b}$ processes in ATLAS adopted a simpler parametrization of the flavour-tagging efficiencies using (p_T, η) maps, whose inferior performance forced including high-ranking additional systematic uncertainties. The good parametrization achieved using the GNNs surpasses the need for such additional uncertainties.

Similar to Galileo's lenses, whose upgrades allowed him to see further away than Hans Lippershey, GNNs greatly surpass the performance of DNNs. This was demonstrated for jet flavour-tagging in Chap. 3, where the **GN2** algorithm was described. Its superior expressivity allows it better to grasp the internal structure of the hadronic jets, improving the c and light-flavour jets rejection rates of factors of ~ 4 when used to identify b -jets at a 70% efficiency WP, as showcased in Fig. 3.34.

GN2 sees far, but we want to see further. In Chap. 4, **GN2+Clus**, a novel jet flavour-tagging algorithm that I developed harnessing the information on the clusters of silicon hits in the ATLAS Inner Detector in addition to the tracks, was presented. Introducing the clusters of silicon hits enables to better identify the flavour of the jets when they are produced with high p_T , where the efficiency of the track reconstruction algorithms decreases for the reasons

detailed in Sec. 4. **GN2+Clus** succeeds in significantly improving the background jet rejection rates of **GN2**, with peaks of $\sim 80\%$ ($\sim 10\%$) of additional rejection when used to identify b -jets (c -jets) with $1\text{ TeV} < p_T < 2\text{ TeV}$, even if approximately ten times more statistics was used to train **GN2**. Moreover, GNN-based jet flavour-tagging algorithms utilizing *only* the clusters of silicon hits were introduced, whose promising performance opens the door to future applications in the context of fast trigger pre-selection of jets, where tracking algorithms drain most of the available computational resources.

This new generation of GNN-based jet flavour-tagging algorithms is the instrument we pass on for future measurements of processes featuring heavy-flavour jets in the ATLAS experiment at the LHC. My interests especially lay in further measurements of the $VH, H \rightarrow c\bar{c}$ process also harnessing the available Run-3 proton-proton collision data, as well as processes such as $HH \rightarrow b\bar{b}\tau^-\tau^+$, which is sensitive to the Higgs boson self-coupling and is unobserved as of today.

Our goal is now to calibrate these instruments and to put them to good use.

New insights lay surely ahead.

“Eppur si muove...”

These personal considerations solely express the opinion of the author.

Science proceeds by trying to falsify its models, which aim to describe as precisely as possible the established experimental facts. The deontology of the scientist requires him to be maximally aware of the models he uses and to ponder the established facts without altering them for his means in any way.

Unfortunately, political pressures have been interfering with the work of scientists from the primes of Science, as the human experience of Galileo Galilei testifies. Due to external interests, Galileo was forced to recant its superior model of the solar system.

I believe it is time to invert this paradigm, and, as scientists, advocate for our representatives to comprehend their models, honestly assessing their alignment with established facts.

In the international platform, no fact is as compelling as a scientific observation. Nevertheless, shared values that our societies agreed upon are written in the Universal Declaration of Human Rights [191]. As I write, blatant violations of these Rights are being committed in various arenas, but the model of righteousness assumed by our institutions realizes in opposing only some of them, while others are ignored or enabled.

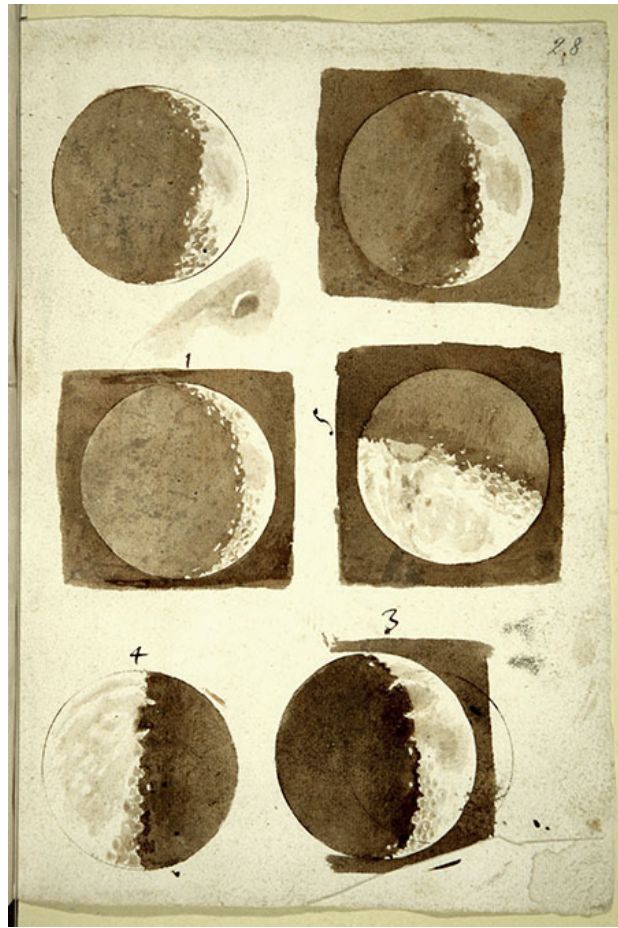
I advocate that the justest model for geopolitical action should realize in opposing all the violations of human rights. Nevertheless, scientists in our academies and institutions are being pressured to condemn only some of these, and to overlook the others.

“Eppur si muove...”,

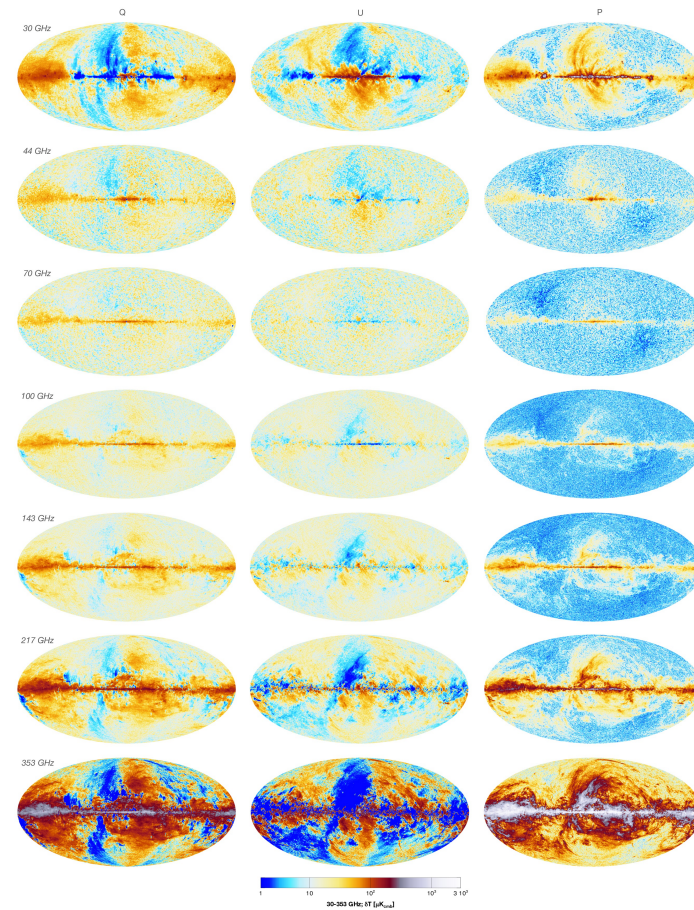
thought Galileo in the secret of his mind, when men of influence tried to force upon him an inferior model of the Cosmos. The resistance of its thought and of the generations of scientists that grew upon his shoulders ignited the scientific revolution. May our pressure towards our representatives be the spark that fires a revolution of Justice.

This Thesis is dedicated to A. and M., my Palestinian friends and colleagues. I hope your loved ones survive our relentless bombing.

Appendices



(a)



(b)

Figure 62: The scientific method proceeds by developing new instruments to falsify its most established models and stare into the unseen landscapes at the edges of the unknown. (a) Galileo Galilei, phases of the Moon, watercolour, 1610. The Moon is not a smooth sphere. (b) Sky polarization in seven polarized frequency bands of *Planck*, 2019 [190]. The Universe expands. Isotropy of the Universe is violated at large scales. Causally separated regions of the sky show correlations.

Appendix A

Unblinded results of the (W/Z) $H, H \rightarrow b\bar{b}/c\bar{c}$ measurement

I defended this thesis on the 16th of May 2024 at the University of Genova. The (W/Z) $H, H \rightarrow b\bar{b}/c\bar{c}$ measurement described in Chapter 6 was later unblinded and published in [2]. I added this Appendix after my defense for completeness, to present the final results obtained by the team.

As anticipated in Sec. 6.12 and Sec. 6.10, after the blinded stage, the final observed and expected results are obtained via unconditional fits to data and unconditional fits to a post-fit Asimov dataset built by adding the nominal signal and background templates setting the NPs to the values estimated via the fit to the observed data.

A.1 (W/Z) $H, H \rightarrow b\bar{b}/c\bar{c}$ combined fit

Results for the $VH, H \rightarrow b\bar{b}$ process The measured signal strength $\hat{\mu}_{VH, H \rightarrow b\bar{b}}$ of the $VH, H \rightarrow b\bar{b}$ process is:

$$\hat{\mu}_{VH, H \rightarrow b\bar{b}} = 0.91^{+0.16}_{-0.15} = 0.91 \pm 0.10 \text{ (stat.)}^{+0.13}_{-0.11} \text{ (syst.)}, \quad (\text{A.1})$$

corresponding to an observed (expected) significance of the signal over the background-only hypothesis of:

$$Z_{VH, H \rightarrow b\bar{b}} = 7.4 \text{ (8.0)} \sigma. \quad (\text{A.2})$$

Fig. A.1 showcases the observed signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained when assigned separate POIs, corresponding to observed (expected) significances of

$$Z_{WH, H \rightarrow b\bar{b}} = \mathbf{5.3 \text{ (5.5)}} \sigma \text{ and } Z_{ZH, H \rightarrow b\bar{b}} = 4.9 \text{ (5.7)} \sigma. \quad (\text{A.3})$$

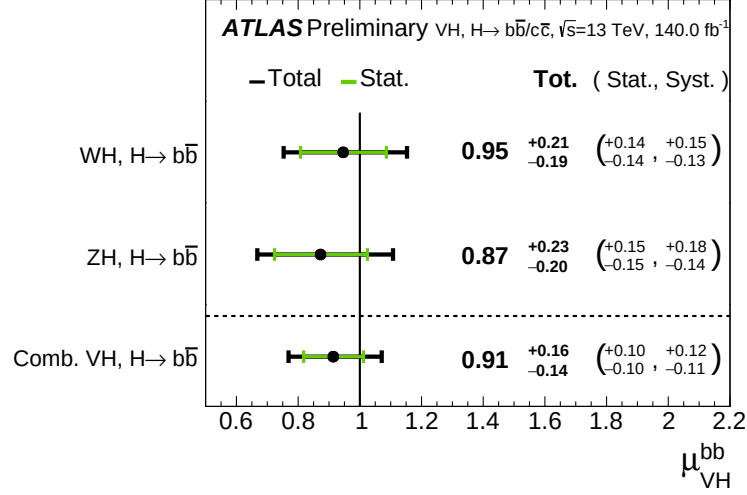


Figure A.1: The observed signal strengths of the $WH, H \rightarrow b\bar{b}$ and $ZH, H \rightarrow b\bar{b}$ processes obtained when assigned separate POIs, along with the combined result where $VH, H \rightarrow b\bar{b}$ is assigned a single POI.

The observed $Z_{WH, H \rightarrow b\bar{b}}$ value exceeds the 5σ limit required to claim observation within the particle physics community. Therefore, this is the first observation of the $WH, H \rightarrow b\bar{b}$ process.

Results for the $VH, H \rightarrow c\bar{c}$ process The measured signal strength $\hat{\mu}_{VH, H \rightarrow c\bar{c}}$ of the $VH, H \rightarrow c\bar{c}$ process is:

$$\hat{\mu}_{VH, H \rightarrow c\bar{c}} = 1.0^{+5.4}_{-5.2} = 1.0^{+4.0}_{-3.9} (\text{stat.})^{+3.7}_{-3.5} (\text{syst.}), \quad (\text{A.4})$$

corresponding to an observed (expected) 95% upper limit set on the signal strength of the $VH, H \rightarrow c\bar{c}$ process of:

$$95\% \text{ CL upper limit on } \mu_{VH, H \rightarrow c\bar{c}} = 11.3 (10.4). \quad (\text{A.5})$$

Fig. A.2 showcases the 95% CL upper limits on the $VH, H \rightarrow c\bar{c}$ process when assigning one POI per lepton channel, together with the combined result. The tree lepton channels have similar sensitivities, with 0L contributing the most to the limit.

Fig. A.3 shows the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ measured signal strengths together with the expected and observed 68% and 95% CL contours. Both measurements show good agreement with the SM and their correlation is 5%.

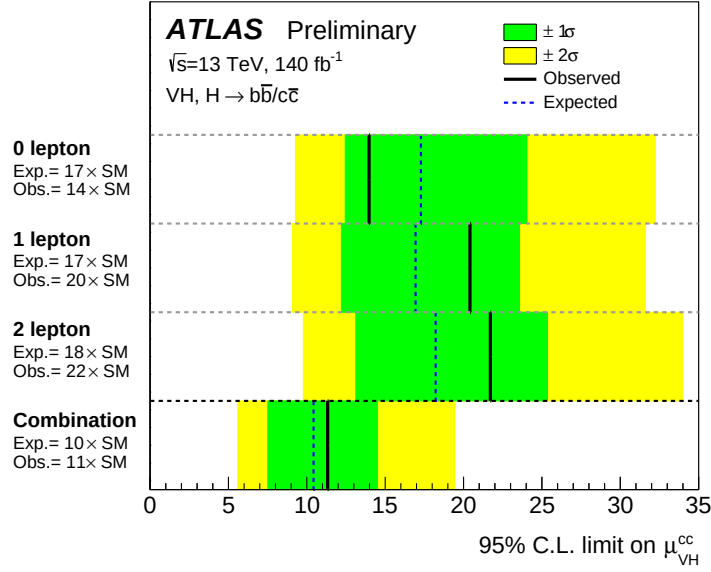


Figure A.2: 95% CL upper limits on the $VH, H \rightarrow c\bar{c}$ process when assigning one POI per lepton channel, together with the combined result. The tree lepton channels have similar sensitivities, with 0L contributing the most to the limit.

A.2 Observed results for the $VZ, Z \rightarrow b\bar{b}/c\bar{c}$ process

The observed (expected) signal strengths of the $VZ, Z \rightarrow b\bar{b}$ and $VZ, Z \rightarrow c\bar{c}$ processes are

$$\mu_{VZ}^{bb} = 0.92^{+0.13}_{-0.11} = 0.92 \pm 0.05 \text{ (stat.)}^{+0.12}_{-0.10} \text{ (syst.)} \quad (\text{A.6})$$

and

$$\mu_{VZ}^{cc} = 0.98^{+0.25}_{-0.22} = 0.98 \pm 0.13 \text{ (stat.)}^{+0.22}_{-0.18} \text{ (syst.)}, \quad (\text{A.7})$$

in good agreement with the SM predictions. For the $VZ, Z \rightarrow c\bar{c}$ process, the observed (expected) significance over the background-only prediction is 5.2 (5.3) σ , marking the first observation of the process within ATLAS. The $VZ, Z \rightarrow b\bar{b}$ process is observed with a significance greater than 10σ .

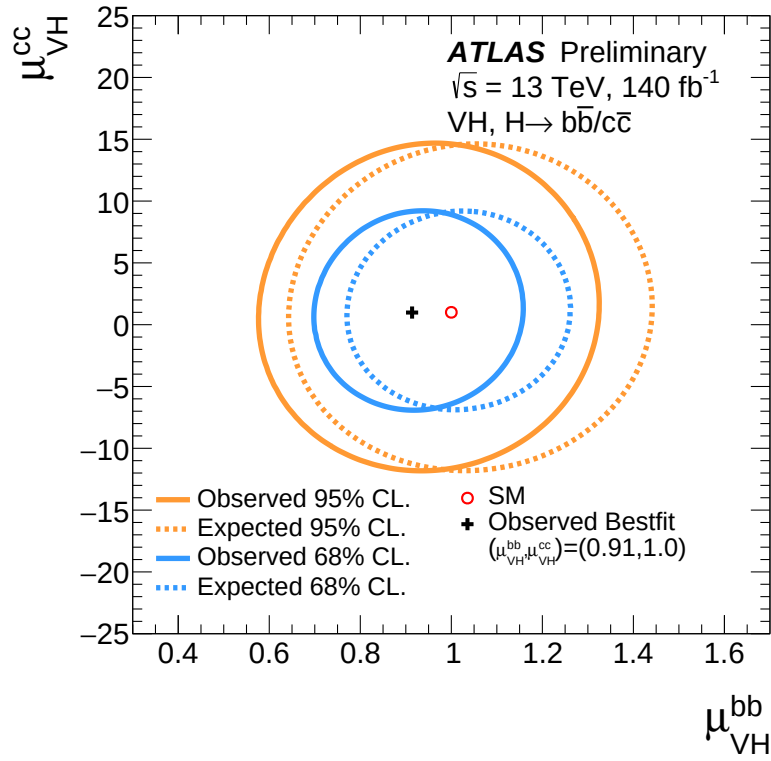


Figure A.3: The observed (solid lines) and expected (dashed lines) 68% and 95% CL contours of the $VH, H \rightarrow b\bar{b}$ and $VH, H \rightarrow c\bar{c}$ signal strengths, along with their best fit values

Bibliography

- [1] G. P. Salam et al., The Higgs boson turns ten, *Nature*, 607(7917):41–47, 2022, doi: 10.1038/s41586-022-04899-4.
- [2] V. Dao et al., Measurements of $W H$ and $Z H$ Higgs production with decays into bottom quarks and direct constraints on the charm Yukawa coupling with 13 TeV collisions in the ATLAS detector., Technical report, CERN, Geneva, 2024, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-010>.
- [3] Wikipedia contributors, Standard model, https://en.wikipedia.org/wiki/Standard_Model.
- [4] S. Weinberg, The Quantum Theory of Fields, Cambridge University Press, 1995.
- [5] J. Schwinger, On quantum-electrodynamics and the magnetic moment of the electron, *Phys. Rev.*, 73:416–417, Feb 1948, doi:10.1103/PhysRev.73.416.
- [6] R. P. Feynman, Space-time approach to quantum electrodynamics, *Phys. Rev.*, 76:769–789, Sep 1949, doi:10.1103/PhysRev.76.769.
- [7] S. Tomonaga, On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*, *Progress of Theoretical Physics*, 1(2):27–42, 08 1946, ISSN 0033-068X, doi:10.1143/PTP.1.27.
- [8] M. Gell-Mann, A Schematic Model of Baryons and Mesons, *Phys. Lett.*, 8:214–215, 1964, doi:10.1016/S0031-9163(64)92001-3.
- [9] D. J. Gross et al., Ultraviolet behavior of non-abelian gauge theories, *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973, doi:10.1103/PhysRevLett.30.1343.
- [10] H. D. Politzer, Reliable perturbative results for strong interactions?, *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973, doi:10.1103/PhysRevLett.30.1346.
- [11] S. L. Glashow, Partial-symmetries of weak interactions, *Nuclear Physics*, 22(4):579–588, 1961, ISSN 0029-5582, doi:[https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).

-
- [12] S. Weinberg, A model of leptons, *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967, doi:10.1103/PhysRevLett.19.1264.
 - [13] A. Salam, Weak and Electromagnetic Interactions, *Conf. Proc. C*, 680519:367–377, 1968, doi:10.1142/9789812795915_0034.
 - [14] F. Englert et al., Broken symmetry and the mass of gauge vector mesons, *Phys. Rev. Lett.*, 13:321–323, Aug 1964, doi:10.1103/PhysRevLett.13.321.
 - [15] P. W. Higgs, Broken symmetries and the masses of gauge bosons, *Phys. Rev. Lett.*, 13:508–509, Oct 1964, doi:10.1103/PhysRevLett.13.508.
 - [16] O. S. Brüning et al., LHC Design Report, CERN Yellow Reports: Monographs, CERN, Geneva, 2004, doi:10.5170/CERN-2004-003-V-1.
 - [17] G. Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider, *JINST*, 3:S08003, 2008, doi:10.1088/1748-0221/3/08/S08003.
 - [18] S. Chatrchyan et al., The CMS Experiment at the CERN LHC, *JINST*, 3:S08004, 2008, doi:10.1088/1748-0221/3/08/S08004.
 - [19] A. A. Alves, Jr. et al., The LHCb Detector at the LHC, *JINST*, 3:S08005, 2008, doi:10.1088/1748-0221/3/08/S08005.
 - [20] K. Aamodt et al., The ALICE experiment at the CERN LHC, *JINST*, 3:S08002, 2008, doi:10.1088/1748-0221/3/08/S08002.
 - [21] ATLAS Collaboration, Online luminosity summary plots for run-3, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
 - [22] J. Pequeno, Computer generated image of the whole ATLAS detector, 2008, <https://cds.cern.ch/record/1095924>.
 - [23] K. Potamianos, The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider, *PoS*, EPS-HEP2015:261, 2015.
 - [24] A. Collaboration, Impact parameter resolution, 2015, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/>.
 - [25] ATLAS Collaboration, Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties, *JINST*, 12:P05002, 2017, doi:10.1088/1748-0221/12/05/P05002.
 - [26] ATLAS Collaboration, ATLAS Liquid Argon Calorimeter: Technical Design Report, 1996.

-
- [27] A. Collaboration, Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter, *Nucl. Instrum. Methods Phys. Res., A*, 558:388–418, 2006, doi:10.1016/j.nima.2005.11.212.
- [28] D. M. G. al, Construction, assembly and testing of the atlas hadronic end-cap calorimeter, *Journal of Instrumentation*, 2(05):P05005, may 2007, doi:10.1088/1748-0221/2/05/P05005.
- [29] ATLAS tile calorimeter: Technical Design Report, Technical design report. ATLAS, CERN, Geneva, 1996, doi:10.17181/CERN.JRBJ.7O28.
- [30] J. Pequenaio, Computer Generated image of the ATLAS calorimeter, 2008, <https://cds.cern.ch/record/1095927>.
- [31] D. M. G. al, Construction, assembly and testing of the atlas hadronic end-cap calorimeter, *Journal of Instrumentation*, 2(05):P05005, may 2007, doi:10.1088/1748-0221/2/05/P05005.
- [32] F. Bauer et al., Construction and Test of MDT Chambers for the ATLAS Muon Spectrometer, *Nucl. Instrum. Meth. A*, 461:17–20, 2001, doi:10.1016/S0168-9002(00)01156-6.
- [33] K. Nikolopoulos et al., Cathode strip chambers in ATLAS : Installation, commissioning and in situ performance, in 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference and 16th International Workshop on Room-Temperature Semiconductor X-Ray and Gamma-Ray Detectors, pages 2819–2824, 2008, doi:10.1109/NSSMIC.2008.4774958.
- [34] G. Aielli et al., The rpc first level muon trigger in the barrel of the atlas experiment, *Nuclear Physics B - Proceedings Supplements*, 158:11–15, 2006, ISSN 0920-5632, doi:<https://doi.org/10.1016/j.nuclphysbps.2006.07.031>, proceedings of the 8th International Workshop on Resistive Plate Chambers and Related Detectors.
- [35] S. Majewski et al., A thin multiwire chamber operating in the high multiplication mode, *Nuclear Instruments and Methods in Physics Research*, 217(1):265–271, 1983, ISSN 0167-5087, doi:[https://doi.org/10.1016/0167-5087\(83\)90146-1](https://doi.org/10.1016/0167-5087(83)90146-1).
- [36] ATLAS Collaboration, Performance of the ATLAS trigger system in 2015, *Eur. Phys. J. C*, 77:317, 2017, doi:10.1140/epjc/s10052-017-4852-3.
- [37] J. Pequenaio et al., How ATLAS detects particles: diagram of particle paths in the detector, 2013, <https://cds.cern.ch/record/1505342>.
- [38] A. Collaboration, A neural network clustering algorithm for the atlas silicon pixel detector, *JINST*, 9:P09009, 2014, doi:10.1088/1748-0221/9/09/P09009.

-
- [39] T. G. Cornelissen et al., Updates of the ATLAS Tracking Event Data Model (Release 13), 6 2007.
- [40] T. Cornelissen et al., The new atlas track reconstruction (newt), *Journal of Physics: Conference Series*, 119(3):032014, jul 2008, doi:10.1088/1742-6596/119/3/032014.
- [41] ATLAS Collaboration, Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC, *Eur. Phys. J. C*, 77:332, 2017, doi:10.1140/epjc/s10052-017-4887-5.
- [42] ATLAS Collaboration, Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-026, 2015.
- [43] ATLAS Collaboration, Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, *Eur. Phys. J. C*, 77:490, 2017, doi:10.1140/epjc/s10052-017-5004-5.
- [44] M. Cacciari et al., The anti- k_t jet clustering algorithm, *JHEP*, 04:063, 2008, doi:10.1088/1126-6708/2008/04/063.
- [45] M. Cacciari et al., FastJet user manual, *Eur. Phys. J. C*, 72:1896, 2012, doi:10.1140/epjc/s10052-012-1896-2.
- [46] M. Cacciari et al., The catchment area of jets, *Journal of High Energy Physics*, 2008(04):005, apr 2008, doi:10.1088/1126-6708/2008/04/005.
- [47] ATLAS Collaboration, Jet reconstruction and performance using particle flow with the ATLAS Detector, *Eur. Phys. J. C*, 77:466, 2017, doi:10.1140/epjc/s10052-017-5031-2.
- [48] D. Krohn et al., Jets with variable r , *Journal of High Energy Physics*, 2009(06):059, jun 2009, doi:10.1088/1126-6708/2009/06/059.
- [49] ATLAS Collaboration, Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data, *JINST*, 14:P03017, 2019, doi:10.1088/1748-0221/14/03/P03017.
- [50] ATLAS Collaboration, Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C*, 79:639, 2019, doi:10.1140/epjc/s10052-019-7140-6.
- [51] ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C*, 76:292, 2016, doi:10.1140/epjc/s10052-016-4120-y.
- [52] ATLAS Collaboration, Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-027, 2015.

-
- [53] ATLAS Collaboration, ATLAS data quality operations and performance for 2015–2018 data-taking, *JINST*, 15:P04003, 2020, doi:10.1088/1748-0221/15/04/P04003.
- [54] J. C. Collins et al., Factorization of Hard Processes in QCD, *Adv. Ser. Direct. High Energy Phys.*, 5:1–91, 1989, doi:10.1142/9789814503266_0001.
- [55] ATLAS Collaboration, The ATLAS Simulation Infrastructure, *Eur. Phys. J. C*, 70:823, 2010, doi:10.1140/epjc/s10052-010-1429-9.
- [56] A. Collaboration, Athena, zenodo, 2019, doi:https://doi.org/10.5281/zenodo.2641997.
- [57] GEANT4 Collaboration et al., GEANT4 – a simulation toolkit, *Nucl. Instrum. Meth. A*, 506:250, 2003, doi:10.1016/S0168-9002(03)01368-8.
- [58] ATLAS Collaboration, Measurement of the $t\bar{t}$ cross-section and $t\bar{t}/Z$ cross-section ratio using LHC Run 3 pp collision data at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV, ATLAS-CONF-2022-070, 2022.
- [59] ATLAS Collaboration, Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *JHEP*, 03:145, 2020, doi:10.1007/JHEP03(2020)145.
- [60] ATLAS Collaboration, Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector, *Phys. Lett. B*, 786:59, 2018, doi:10.1016/j.physletb.2018.09.013.
- [61] R. L. Workman et al., Review of Particle Physics, *PTEP*, 2022:083C01, 2022, doi:10.1093/ptep/ptac097.
- [62] ATLAS Collaboration, Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets, ATL-PHYS-PUB-2014-008, 2014.
- [63] G. Brandenburg et al., Charged track multiplicity in b meson decay, *Phys. Rev. D*, 61, 03 2000, doi:10.1103/PhysRevD.61.072002.
- [64] ATLAS Collaboration, Measurements of b -jet moments sensitive to b -quark fragmentation in $t\bar{t}$ events at the LHC with the ATLAS detector, ATLAS-CONF-2020-050, 2020.
- [65] ATLAS Collaboration, Jet flavour tagging with gn1 and dll1d. generator dependence, run 2 and run 3 data agreement studies, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/>.
- [66] ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset, 11 2022.
- [67] ATLAS Collaboration, Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV, ATL-PHYS-PUB-2015-051, 2015.

-
- [68] M. Aaboud et al., Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC, *Eur. Phys. J. C*, 77(5):332, 2017, doi:10.1140/epjc/s10052-017-4887-5.
 - [69] ATLAS Collaboration, Tagging and suppression of pileup jets with the ATLAS detector, ATLAS-CONF-2014-018, 2014.
 - [70] ATLAS Collaboration, Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run, ATL-PHYS-PUB-2017-013, 2017.
 - [71] ATLAS Collaboration, Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment, ATL-PHYS-PUB-2017-003, 2017.
 - [72] ATLAS Collaboration, Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS, ATL-PHYS-PUB-2020-014, 2020.
 - [73] ATLAS Collaboration, Secondary vertex finding for jet flavour identification with the ATLAS detector, ATL-PHYS-PUB-2017-011, 2017.
 - [74] ATLAS Collaboration, Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC, ATL-PHYS-PUB-2018-025, 2018.
 - [75] R. Frühwirth, Application of kalman filtering to track and vertex fitting, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987, ISSN 0168-9002, doi:[https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
 - [76] ATLAS Collaboration, Expected performance of the ATLAS b -tagging algorithms in Run-2, ATL-PHYS-PUB-2015-022, 2015.
 - [77] ATLAS Collaboration, Graph Neural Network Jet Flavour Tagging with the ATLAS Detector, ATL-PHYS-PUB-2022-027, 2022.
 - [78] S. Brody et al., How attentive are graph attention networks?, 2022.
 - [79] A. Vaswani et al., Attention is all you need, 2023.
 - [80] Calibration of the light-flavour jet mistagging efficiency of the b -tagging algorithms with Z +jets events using 139 fb^{-1} of ATLAS proton-proton collision data at $\sqrt{s} = 13\text{ TeV}$, *Eur. Phys. J. C*, 83:728, 2023, doi:10.1140/epjc/s10052-023-11736-z.
 - [81] ATLAS Collaboration, ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\text{ TeV}$, *Eur. Phys. J. C*, 79:970, 2019, doi:10.1140/epjc/s10052-019-7450-8.

-
- [82] ATLAS Collaboration, Measurement of the c -jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector, *Eur. Phys. J. C*, 82:95, 2021, doi:10.1140/epjc/s10052-021-09843-w.
- [83] ATLAS Collaboration, Simulation-based extrapolation of b -tagging calibrations towards high transverse momenta in the ATLAS experiment, ATL-PHYS-PUB-2021-003, 2021.
- [84] ATLAS Collaboration, Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2, *Eur. Phys. J. C*, 77:673, 2017, doi:10.1140/epjc/s10052-017-5225-7.
- [85] ATLAS Collaboration, Simulated flavour composition of tracks processed by the atlas b -tagging algorithms applied to particle flow jets, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-003/>.
- [86] ATLAS Collaboration, Impact of atlas quasi-stable particle simulation on quantities related to flavour tagging and tau reconstruction, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-003/>.
- [87] B. T. Huffman et al., Tagging b quarks at extreme energies without tracks, *J. Phys. G*, 43(8):085001, 2016, doi:10.1088/0954-3899/43/8/085001.
- [88] C. Peterson et al., Scaling violations in inclusive e^+e^- annihilation spectra, *Phys. Rev. D*, 27:105–111, Jan 1983, doi:10.1103/PhysRevD.27.105.
- [89] M. Tanasini, Developing novel jet flavour tagging techniques at high transverse momentum for the ATLAS experiment at the LHC, *Nuovo Cim. C*, 45(4):69, 2022, doi:10.1393/ncc/i2022-22069-8.
- [90] M. Tanasini, Trackless flavour-tagging at high jet transverse momentum and its impact on the physics reach of the ATLAS experiment at LHC, 2020, presented 29 Oct 2020.
- [91] D. P. Kingma et al., Adam: A method for stochastic optimization, 2017.
- [92] ATLAS Collaboration, Calibration of light-flavour b -jet mistagging rates using ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV, ATLAS-CONF-2018-006, 2018.
- [93] F. A. Di Bello et al., Efficiency Parameterization with Neural Networks, *Comput. Softw. Big Sci.*, 5(1):14, 2021, doi:10.1007/s41781-021-00059-x.
- [94] ATLAS Collaboration, Flavor Tagging Efficiency Parametrisations with Graph Neural Networks, ATL-PHYS-PUB-2022-041, 2022.
- [95] ATLAS Collaboration, Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector, 2022.

-
- [96] ATLAS Collaboration, Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector, *Eur. Phys. J. C*, 81:178, 2021, doi:10.1140/epjc/s10052-020-08677-2.
- [97] T. S. M. Sugiyama et al., Density Ratio Estimation in Machine Learning, 2012, doi:10.1017/CBO9781139035613.
- [98] G. Aad et al., Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector, *JHEP*, 01:069, 2015, doi:10.1007/JHEP01(2015)069.
- [99] ATLAS and CMS Collaborations, Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV, *JHEP*, 08:045, 2016, doi:10.1007/JHEP08(2016)045.
- [100] ATLAS Collaboration, Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector, *Phys. Lett. B*, 784:173, 2018, doi:10.1016/j.physletb.2018.07.035.
- [101] A. M. Sirunyan et al., Observation of $t\bar{t}H$ production, *Phys. Rev. Lett.*, 120(23):231801, 2018, doi:10.1103/PhysRevLett.120.231801.
- [102] A. M. Sirunyan et al., Observation of Higgs boson decay to bottom quarks, *Phys. Rev. Lett.*, 121(12):121801, 2018, doi:10.1103/PhysRevLett.121.121801.
- [103] ATLAS Collaboration, A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery, *Nature*, 607:52–59, 2022, doi:10.1038/s41586-022-04893-w.
- [104] Standard Model Summary Plots October 2023, Technical report, CERN, Geneva, 2023, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-039>.
- [105] ATLAS Collaboration, Measurement of the associated production of a Higgs boson decaying into b -quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Phys. Lett. B*, 816:136204, 2021, doi:10.1016/j.physletb.2021.136204.
- [106] ATLAS Collaboration, Combination of measurements of Higgs boson production in association with a W or Z boson in the $b\bar{b}$ decay channel with the ATLAS experiment at $\sqrt{s} = 13$ TeV, ATLAS-CONF-2021-051, 2021.
- [107] F. J. Botella et al., What if the masses of the first two quark families are not generated by the standard model Higgs boson?, *Phys. Rev. D*, 94(11):115031, 2016, doi:10.1103/PhysRevD.94.115031.

-
- [108] D. Ghosh et al., Is the Higgs Mechanism of Fermion Mass Generation a Fact? A Yukawa-less First-Two-Generation Model, *Phys. Lett. B*, 755:504–508, 2016, doi:10.1016/j.physletb.2016.02.059.
- [109] D. Egana-Ugrinovic et al., Higgs bosons with large couplings to light quarks, *Phys. Rev. D*, 100(11):115041, 2019, doi:10.1103/PhysRevD.100.115041.
- [110] C. Delaunay et al., Enhanced Higgs boson coupling to charm pairs, *Phys. Rev. D*, 89(3):033014, 2014, doi:10.1103/PhysRevD.89.033014.
- [111] K. Becker et al., Precise predictions for boosted Higgs production, *SciPost Phys. Core*, 7:001, 2024, doi:10.21468/SciPostPhysCore.7.1.001.
- [112] A. M. Sirunyan et al., Evidence for Higgs boson decay to a pair of muons, *JHEP*, 01:148, 2021, doi:10.1007/JHEP01(2021)148.
- [113] ATLAS Collaboration, A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector, *Phys. Lett. B*, 812:135980, 2021, doi:10.1016/j.physletb.2020.135980.
- [114] A. Djouadi et al., HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension, *Comput. Phys. Commun.*, 108:56–74, 1998, doi:10.1016/S0010-4655(97)00123-9.
- [115] A. Tumasyan et al., Search for Higgs Boson Decay to a Charm Quark-Antiquark Pair in Proton-Proton Collisions at $\sqrt{s}=13$ TeV, *Phys. Rev. Lett.*, 131(6):061801, 2023, doi:10.1103/PhysRevLett.131.061801.
- [116] K. Cranmer et al., Approximating Likelihood Ratios with Calibrated Discriminative Classifiers, 6 2015.
- [117] Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS, Technical report, CERN, Geneva, 2023, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-021>.
- [118] ATLAS Collaboration, Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment, ATL-PHYS-PUB-2019-033, 2019.
- [119] S. Frixione et al., Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, *JHEP*, 11:070, 2007, doi:10.1088/1126-6708/2007/11/070.
- [120] S. Alioli et al., A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, *JHEP*, 06:043, 2010, doi:10.1007/JHEP06(2010)043.

-
- [121] NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II, *JHEP*, 04:040, 2015, doi:10.1007/JHEP04(2015)040.
- [122] T. Sjöstrand et al., An introduction to PYTHIA 8.2, *Comput. Phys. Commun.*, 191:159–177, 2015, doi:10.1016/j.cpc.2015.01.024.
- [123] ATLAS Collaboration, Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, *JHEP*, 09:145, 2014, doi:10.1007/JHEP09(2014)145.
- [124] M. L. Ciccolini et al., Electroweak radiative corrections to associated WH and ZH production at hadron colliders, *Phys. Rev. D*, 68:073003, 2003, doi:10.1103/PhysRevD.68.073003.
- [125] O. Brein et al., NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders, *Phys. Lett. B*, 579:149–156, 2004, doi:10.1016/j.physletb.2003.10.112.
- [126] O. Brein et al., Top-quark mediated effects in hadronic Higgs-Strahlung, *Eur. Phys. J. C*, 72:1868, 2012, doi:10.1140/epjc/s10052-012-1868-6.
- [127] A. Denner et al., 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, doi:10.1016/j.cpc.2015.04.021.
- [128] O. Brein et al., $vh@nnlo$ – Higgs Strahlung at hadron colliders, *Comput. Phys. Commun.*, 184:998–1003, 2013, doi:10.1016/j.cpc.2012.11.002.
- [129] G. Cullen et al., Automated one-loop calculations with GOSAM, *Eur. Phys. J. C*, 72:1889, 2012, doi:10.1140/epjc/s10052-012-1889-1.
- [130] K. Hamilton et al., MINLO: Multi-Scale Improved NLO, *JHEP*, 10:155, 2012, doi:10.1007/JHEP10(2012)155.
- [131] G. Luisoni et al., $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MinLO, *JHEP*, 10:083, 2013, doi:10.1007/JHEP10(2013)083.
- [132] L. Altenkamp et al., Gluon-induced Higgs-strahlung at next-to-leading order QCD, *JHEP*, 02:078, 2013, doi:10.1007/JHEP02(2013)078.
- [133] R. V. Harlander et al., Soft gluon resummation for gluon-induced Higgs Strahlung, *JHEP*, 11:082, 2014, doi:10.1007/JHEP11(2014)082.
- [134] S. Frixione et al., A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction, *JHEP*, 09:126, 2007, doi:10.1088/1126-6708/2007/09/126.
- [135] ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021, 2014.

-
- [136] M. Beneke et al., Hadronic top-quark pair production with NNLL threshold resummation, *Nucl. Phys. B*, 855:695–741, 2012, doi:10.1016/j.nuclphysb.2011.10.021.
- [137] M. Cacciari et al., Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation, *Phys. Lett. B*, 710:612–622, 2012, doi:10.1016/j.physletb.2012.03.013.
- [138] P. Bärnreuther et al., Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$, *Phys. Rev. Lett.*, 109:132001, 2012, doi:10.1103/PhysRevLett.109.132001.
- [139] M. Czakon et al., NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels, *JHEP*, 12:054, 2012, doi:10.1007/JHEP12(2012)054.
- [140] M. Czakon et al., NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction, *JHEP*, 01:080, 2013, doi:10.1007/JHEP01(2013)080.
- [141] M. Czakon et al., Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$, *Phys. Rev. Lett.*, 110:252004, 2013, doi:10.1103/PhysRevLett.110.252004.
- [142] M. Czakon et al., Top++: A program for the calculation of the top-pair cross-section at hadron colliders, *Comput. Phys. Commun.*, 185:2930, 2014, doi:10.1016/j.cpc.2014.06.021.
- [143] S. Alioli et al., NLO single-top production matched with shower in POWHEG: s- and t-channel contributions, *JHEP*, 09:111, 2009, doi:10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111, [Erratum: JHEP02,011(2010)].
- [144] M. Aliev et al., HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR, *Comput. Phys. Commun.*, 182:1034–1046, 2011, doi:10.1016/j.cpc.2010.12.040.
- [145] P. Kant et al., HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions, *Comput. Phys. Commun.*, 191:74–89, 2015, doi:10.1016/j.cpc.2015.02.001.
- [146] E. Re, Single-top Wt-channel production matched with parton showers using the POWHEG method, *Eur. Phys. J. C*, 71:1547, 2011, doi:10.1140/epjc/s10052-011-1547-z.
- [147] N. Kidonakis, Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^- , *Phys. Rev. D*, 82:054018, 2010, doi:10.1103/PhysRevD.82.054018.
- [148] N. Kidonakis, Top Quark Production, in Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013), pages 139–168, doi:10.3204/DESY-PROC-2013-03/Kidonakis.

-
- [149] E. Bothmann et al., Event Generation with Sherpa 2.2, 2019.
- [150] C. Anastasiou et al., High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO, *Phys. Rev. D*, 69:094008, 2004, doi:10.1103/PhysRevD.69.094008.
- [151] M. Proissl, Dijet Invariant Mass Studies in the Higgs boson $H \rightarrow b\bar{b}$ resonance search in association with a W/Z boson using the ATLAS detector, 2014, presented 2014.
- [152] D. Gonçalves et al., Role of the z polarization in the $h \rightarrow b\bar{b}$ measurement, *Phys. Rev. D*, 98:093005, Nov 2018, doi:10.1103/PhysRevD.98.093005.
- [153] A. Buckley et al., An optimal observable for color singlet identification, *SciPost Phys.*, 9:026, 2020, doi:10.21468/SciPostPhys.9.2.026.
- [154] G. Cowan et al., Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C*, 71:1554, 2011, doi:10.1140/epjc/s10052-011-1554-0.
- [155] ATLAS Collaboration, Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with 80 fb^{-1} of $\sqrt{s} = 13\text{ TeV}$ pp data, ATLAS-CONF-2020-022, 2020.
- [156] ATLAS Collaboration, Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$, *Eur. Phys. J. C*, 78:903, 2018, doi:10.1140/epjc/s10052-018-6288-9.
- [157] K. A. L., Sulla determinazione empirica di una legge di distribuzione, *G. Ist. Ital. Attuari*, 4:83–91, 1933.
- [158] R. D. Ball et al., Parton distributions for the LHC Run II, *JHEP*, 04:040, 2015, doi:10.1007/JHEP04(2015)040.
- [159] S. Carrazza et al., An Unbiased Hessian Representation for Monte Carlo PDFs, *Eur. Phys. J. C*, 75(8):369, 2015, doi:10.1140/epjc/s10052-015-3590-7.
- [160] T. Gleisberg et al., Comix, a new matrix element generator, *JHEP*, 12:039, 2008, doi:10.1088/1126-6708/2008/12/039.
- [161] F. Cascioli et al., Scattering Amplitudes with Open Loops, *Phys. Rev. Lett.*, 108:111601, 2012, doi:10.1103/PhysRevLett.108.111601.
- [162] S. Schumann et al., A Parton shower algorithm based on Catani-Seymour dipole factorisation, *JHEP*, 03:038, 2008, doi:10.1088/1126-6708/2008/03/038.
- [163] J.-C. Winter et al., A modified cluster-hadronization model, *Eur. Phys. J. C*, 36:381–395, 2004, doi:10.1140/epjc/s2004-01960-8.
- [164] S. Höche et al., A critical appraisal of NLO+PS matching methods, *JHEP*, 09:049, 2012, doi:10.1007/JHEP09(2012)049.

-
- [165] S. Catani et al., QCD Matrix Elements + Parton Showers, *JHEP*, 11:063, 2001, doi:10.1088/1126-6708/2001/11/063.
- [166] S. Höche et al., QCD matrix elements and truncated showers, *JHEP*, 05:053, 2009, doi:10.1088/1126-6708/2009/05/053.
- [167] S. Höche et al., QCD matrix elements + parton showers. The NLO case, *JHEP*, 04:027, 2013, doi:10.1007/JHEP04(2013)027.
- [168] G. Aad et al., Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment, *JHEP*, 08:089, 2022, doi:10.1007/JHEP08(2022)089.
- [169] S. Catani et al., Vector boson production at hadron colliders: A fully exclusive qcd calculation at next-to-next-to-leading order, *Phys. Rev. Lett.*, 103:082001, Aug 2009, doi:10.1103/PhysRevLett.103.082001.
- [170] J. Alwall et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *JHEP*, 07:079, 2014, doi:10.1007/JHEP07(2014)079.
- [171] V. Bertone et al., Illuminating the photon content of the proton within a global PDF analysis, *SciPost Phys.*, 5(1):008, 2018, doi:10.21468/SciPostPhys.5.1.008.
- [172] T. Sjöstrand et al., A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.*, 178:852–867, 2008, doi:10.1016/j.cpc.2008.01.036.
- [173] R. D. Ball et al., Parton distributions with LHC data, *Nucl. Phys. B*, 867:244–289, 2013, doi:10.1016/j.nuclphysb.2012.10.003.
- [174] R. Frederix et al., Merging meets matching in MC@NLO, *JHEP*, 12:061, 2012, doi:10.1007/JHEP12(2012)061.
- [175] E. Bothmann et al., Reweighting QCD matrix-element and parton-shower calculations, *Eur. Phys. J. C*, 76(11):590, 2016, doi:10.1140/epjc/s10052-016-4430-0.
- [176] T.-J. Hou et al., Progress in the CTEQ-TEA NNLO global QCD analysis, 8 2019.
- [177] S. Bailey et al., Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs, *Eur. Phys. J. C*, 81(4):341, 2021, doi:10.1140/epjc/s10052-021-09057-0.
- [178] A. Buckley et al., LHAPDF6: parton density access in the LHC precision era, *Eur. Phys. J. C*, 75:132, 2015, doi:10.1140/epjc/s10052-015-3318-8.
- [179] ATLAS Collaboration, Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment, 2021.

-
- [180] ATLAS Collaboration, Measurement of the transverse momentum distribution of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Eur. Phys. J. C*, 80:616, 2020, doi:10.1140/epjc/s10052-020-8001-z.
- [181] K. Cranmer et al., HistFactory: A tool for creating statistical models for use with RooFit and RooStats, Technical report, New York U., New York, 2012, doi:10.17181/CERN-OPEN-2012-016.
- [182] W. Verkerke et al., The roofit toolkit for data modeling, 2003.
- [183] L. Moneta et al., The roostats project, 2011.
- [184] R. Barlow et al., Fitting using finite monte carlo samples, *Computer Physics Communications*, 77(2):219–228, 1993, ISSN 0010-4655, doi:https://doi.org/10.1016/0010-4655(93)90005-W.
- [185] F. James et al., Minuit - a system for function minimization and analysis of the parameter errors and correlations, *Computer Physics Communications*, 10(6):343–367, 1975, ISSN 0010-4655, doi:https://doi.org/10.1016/0010-4655(75)90039-9.
- [186] A. L. Read, Presentation of search results: the CL_S technique, *J. Phys. G*, 28:2693, 2002, doi:10.1088/0954-3899/28/10/313.
- [187] A. Roodman, Blind analysis in particle physics, *eConf*, C030908:TUIT001, 2003.
- [188] A. Tumasyan et al., Measurement of simplified template cross sections of the Higgs boson produced in association with W or Z bosons in the $H \rightarrow b\bar{b}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV, 12 2023.
- [189] H. Qu et al., Jet tagging via particle clouds, *Phys. Rev. D*, 101:056019, Mar 2020, doi:10.1103/PhysRevD.101.056019.
- [190] Planck Collaboration et al., Planck 2018 results - i. overview and the cosmological legacy of planck, *A and A*, 641:A1, 2020, doi:10.1051/0004-6361/201833880.
- [191] U. G. Assembly, Universal declaration of human rights, December 1948, <http://www.unhchr.ch/udhr/lang/eng.pdf>.