
Measurement of the Higgs boson coupling to b -quarks in the associated production with a vector boson with the ATLAS detector

Simona Gargiulo



Measurement of the Higgs boson coupling to b -quarks in the associated production with a vector boson with the ATLAS detector

DISSERTATION

zur Erlangung des Doktorgrades der
Fakultät für Mathematik und Physik der

Albert-Ludwigs-Universität
Freiburg im Breisgau

vorgelegt von
Simona Gargiulo

June 29, 2021

Simona Gargiulo: ***Measurement of the Higgs boson decay to a pair of b -quarks in the associated production with a vector boson with the ATLAS detector***

© June 29, 2021

Datum der mündlichen Prüfung: April 14, 2021

Dekan:	Prof. Dr. Wolfgang Soergel
Betreuer der Arbeit:	Dr. Christian Weiser
Koreferent:	Prof. Dr. Beate Heinemann
Prüfer:	Dr. Christian Weiser
	Prof. Dr. Markus Schumacher
	Prof. Dr. Harald Ita

To my little Alessandro.

Abstract

This thesis presents the measurement of the Higgs boson coupling to a pair of b -quarks. The $H \rightarrow b\bar{b}$ decay is predicted by the Standard Model to have a branching ratio of 58%, the largest among all decay modes. Therefore, accessing this decay is crucial to constrain the overall Higgs boson decay width.

The measurement is performed in the production process with a Z or W boson, whose leptonic decays ensure high rejection of the multi-jet background and efficient triggering of the signal signature. The analysis uses proton-proton collision data collected by the ATLAS experiment during the years 2015-2018, as produced by the Large Hadron Collider at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb^{-1} .

The production of the Higgs boson in association with a Z or W boson is measured with an observed (expected) significance of 5.3 (5.1) and 4.0 (4.1) standard deviations, respectively, constituting the observation of the ZH process and the evidence of the WH one. The measured signal strength of $\mu_{VH}^{b\bar{b}} = 1.02^{+0.12}_{-0.11}(\text{stat.})^{+0.14}_{-0.13}(\text{syst.})$ is highly compatible with the Standard Model expectation. Differential measurements of the $VH, H \rightarrow b\bar{b}$ cross section are performed as a function of the vector boson transverse momentum, yielding good agreement with theoretical predictions.

This thesis also presents performance studies of the impact of b - and c -hadron interactions with the ATLAS pixel detector material on the tagging efficiency of jets originating from b -quarks, known as b -jets. In addition, a new technique to identify b -jets without using tracking information is presented and it is expected to significantly improve the tagging efficiency at very high transverse momenta.

Kurzzusammenfassung

Diese Dissertation präsentiert eine Messung der Kopplung des Higgs-Bosons an ein b -Quark-Paar. Das Standardmodell sagt für den $H \rightarrow b\bar{b}$ Zerfall ein Verzweungsverhältnis von 58% vorher, das größte aller Zerfallskanäle. Daher ist ein Zugang zu diesem Zerfall essenziell um die Gesamtzerfallsbreite des Higgs-Bosons einzuschränken.

Die Messung wird für den Produktionsprozess mit einem Z - oder W -Boson durchgeführt, deren leptonische Zerfälle eine gute Unterdrückung des Multijet-Untergrundes und eine hohe Triggereffizienz für die Signalsignatur gewährleisten. Die Analyse verwendet die Proton-Proton-Kollisionsdaten des ATLAS-Experiments aus den Jahren 2015-2018, die durch den Large-Hadron-Collider bei einer Schwerpunktsenergie von 139 fb^{-1} bereitgestellt wurden.

Die beobachtete (erwartete) Signifikanz für die Produktion des Higgs-Bosons in Assoziation mit einem Z - oder W -Boson beträgt 5.3 (5.1) bzw. 4.0 (4.1) Standardabweichungen, was eine Beobachtung des ZH -Prozesses und einen Nachweis des WH -Prozesses darstellt. Die gemessene Signalstärke von $\mu_{VH}^{b\bar{b}} = 1.02^{+0.12}_{-0.11} \text{ (stat.) } ^{+0.14}_{-0.13} \text{ (syst.)}$ ist sehr gut mit der Erwartung durch das Standardmodell vereinbar. Differenzielle Messungen des $VH, H \rightarrow b\bar{b}$ Wirkungsquerschnittes als Funktion des Vektorboson Transversalimpulses liefern eine gute Übereinstimmung mit den Vorhersagen der Theorie.

Diese Arbeit stellt ebenso Studien des Einflusses von Wechselwirkungen von b - und c -Hadronen mit dem Material des ATLAS-Pixeldetektors auf die Effizienz vor, mit der Jets erkannt werden können, die von b -Quarks stammen. Zusätzlich wird eine neue Technik präsentiert, mit der diese sogenannten b -Jets ohne die Informationen der Spurdetektoren identifiziert werden können. Es ist anzunehmen, dass diese Methode zu einer erheblichen Verbesserung der Erkennung von b -Jets bei sehr hohen Transversalimpulsen führen wird.

Contents

Abstract	i
Introduction	1
1 Theoretical Framework	3
1.1 Standard Model Formalism	3
1.1.1 Fundamental particles	6
1.1.2 Quantum Chromodynamics	7
1.1.3 Electroweak Sector	9
1.2 Mass generation and Higgs mechanism	12
1.2.1 Brout-Englert-Higgs mechanism	12
1.2.2 Yukawa coupling to the fermion sector	14
1.3 Higgs boson searches at the LHC	15
1.3.1 Higgs boson production modes	15
1.3.2 Higgs boson decay channels	17
1.3.3 Higgs boson discovery	18
1.4 Simulation of proton-proton collisions at the LHC	19
1.4.1 Definition of cross-section for pp collisions	19
1.4.2 Event simulation	20
2 The ATLAS Detector at the LHC	25
2.1 The Large Hadron Collider	25
2.2 The ATLAS detector	29
2.2.1 Magnet system	30
2.2.2 Inner detector and tracking	31
2.2.3 Electromagnetic Calorimeter	34
2.2.4 Hadronic Calorimeter	37
2.2.5 Muon Spectrometer	38
2.2.6 Luminosity Detectors	39
2.2.7 Trigger System	40
3 Object Reconstruction	43
3.1 Tracks and primary vertices	43
3.2 Electrons	47
3.2.1 Clusters reconstruction	47
3.2.2 Track-cluster matching and superclusters	48

3.2.3	Identification	48
3.2.4	Isolation	49
3.3	Muons	50
3.3.1	Reconstruction	51
3.3.2	Identification	51
3.3.3	Isolation	52
3.4	Tau leptons	53
3.5	Jets	54
3.5.1	Jet Energy Calibration	55
3.5.2	Jet Energy Resolution	57
3.5.3	Jet selection	57
3.5.4	Truth Flavour Labelling	58
3.6	Missing transverse energy	58
3.7	Overlap removal	59
4	Flavour Tagging in ATLAS	61
4.1	Track selection and track-jet matching	61
4.2	Low-level b -tagging algorithms	62
4.2.1	Impact Parameter based Algorithms	62
4.2.2	Secondary vertex finding algorithm	66
4.2.3	Multi-vertex finding algorithm	66
4.3	High-level tagging algorithms	68
4.3.1	Training samples	68
4.3.2	MV2 discriminant	70
4.3.3	DL1 discriminant	70
4.4	b -jet tagging efficiency and algorithms performance	72
4.5	Truth tagging procedure	74
4.6	Calibration of b -tagging efficiency	76
4.7	Extrapolation to high- p_T jets	79
5	b-hadron interactions with the pixel detector material	81
5.1	The ATLAS Simulation Infrastructure	81
5.2	Interactions of b - and c -hadrons with the pixel detector	82
5.2.1	Impact on the b -tagging related variables	85
5.2.2	Impact on the b -tagging efficiency	93
5.2.3	Extrapolation uncertainty on the b -tagging efficiency	107
5.3	Trackless b -tagging for very energetic jets	109
5.3.1	Effect of b -hadron interactions on trackless b -tagging performance	113
6	Measurement of the $H \rightarrow b\bar{b}$ decay in association with a Vector Boson	117
6.1	Observation of the $H \rightarrow b\bar{b}$ decay and VH production	118

6.2	Data and simulated samples	123
6.2.1	Signal Processes	125
6.2.2	Top-quark Pair Production	125
6.2.3	Vector Boson + Jets Production	126
6.2.4	Single-Top Production	127
6.2.5	Diboson Production	128
6.2.6	Multi-jet Background	128
6.3	Object Reconstruction	129
6.4	Event Selection	132
6.5	Multivariate Algorithms	137
6.5.1	Boosted Decision Tree	139
6.5.2	The multivariate technique in the $VH, H \rightarrow b\bar{b}$ analysis	140
6.6	Statistical Interpretation	144
6.6.1	Significance testing	144
6.6.2	Global profile likelihood procedure in the $VH, H \rightarrow b\bar{b}$ analysis	146
6.6.3	Experimental sensitivity and the Asimov dataset	148
6.6.4	Correlations	150
6.6.5	The fit model of the $VH, H \rightarrow b\bar{b}$ analysis	151
6.7	Modelling Uncertainties	156
6.7.1	Normalisation Uncertainties	156
6.7.2	Acceptance Uncertainties	156
6.7.3	Shape Uncertainties	157
6.8	Signal Modelling	165
6.8.1	Simplified Template Cross Section	165
6.9	Modelling of the W +jets background	167
6.9.1	Normalisation and acceptance uncertainties	169
6.9.2	Flavour composition uncertainties	173
6.9.3	BDT-reweighting shape uncertainties	174
6.10	Modelling of other backgrounds	192
6.11	Experimental Systematic Uncertainties	199
6.12	Smoothing and pruning of systematics	199
6.13	Fit model cross-checks	200
6.14	Fit results	209
6.14.1	Main $VH, H \rightarrow b\bar{b}$ analysis	209
6.14.2	Diboson $VZ, Z \rightarrow b\bar{b}$ cross-check analysis	223
6.14.3	$VH, H \rightarrow b\bar{b}$ di-jet mass analysis	223
	Conclusions	229
	Contributions	231
	Bibliography	233

Appendix	251
A Additional material for the b -hadron interactions with the pixel detector	251
B Additional material for the W +jets modelling systematics	259
C Additional material for the $VH, H \rightarrow b\bar{b}$ fit results	293
Acknowledgements	295

Introduction

The Standard Model (SM) of particle physics describes the fundamental constituents of matter and their interactions. Developed in the 1970s [1–5], it successfully predicted a wide variety of phenomena and showed an astonishing compatibility with almost all measurements performed by different experiments over more than 50 years.

In the theoretical framework of the SM, special relativity and quantum mechanics are elegantly unified by interpreting particles as excitations of *quantum fields* parametrised according to a 4-dimensional space-time coordinate system. The fermions represent the building blocks of matter, and can interact with each other through three fundamental forces: electromagnetic, weak, and strong. They are mediated by gauge vector bosons which emerge to preserve the *local gauge symmetry* imposed in the theory.

The mechanism which allows particles to have mass is not intrinsic to the SM as it would violate the fundamental gauge invariance of the theory. This property can be restored with the introduction of a scalar field, called the *Higgs field*, which spontaneously breaks the electroweak symmetry through the Brout-Englert-Higgs mechanism [4, 5], and gives mass to the gauge vector bosons. In the same way, fermions acquire mass by interacting with the Higgs field through the Yukawa mechanism.

The Higgs boson represents the last particle predicted by the Standard Model and its search has motivated the construction of the Large Hadron Collider (LHC), the world’s largest particle accelerator. Almost fifty years after its prediction, the ATLAS [6] and CMS [7] experiments at the LHC claimed the observation of a scalar particle with a mass of around 125 GeV, compatible with the Higgs boson predicted by the Standard Model. The discovery mainly relied on measurements of the Higgs boson decay to vector bosons ($H \rightarrow ZZ$, $H \rightarrow \gamma\gamma$, and $H \rightarrow W^+W^-$) and the result was found to be in remarkable agreement with the prediction of the Brout-Englert-Higgs mechanism and with the predictions from electroweak precision measurements.

The first direct coupling of the Higgs boson to fermions that could be experimentally confirmed, thus validating the Yukawa mechanism, was to τ -leptons. This was observed in the combination of the ATLAS and CMS measurements of $H \rightarrow \tau\tau$ decays [8] using the full dataset collected in the first Run of the LHC. During Run 2, the increase in the LHC centre-of-mass energy and the larger datasets allowed more fermionic couplings to be probed, with the goal of either confirming or observing deviations from the Yukawa theory. The first significant result came from the observation of the $t\bar{t}H$ production mode made by both ATLAS [9] and CMS [10] which, for the first time, probed the direct Higgs boson coupling to the top-quark.

While the top-quark is too heavy to allow for the decay of an on-shell Higgs boson to a pair of top-quarks, the $H \rightarrow b\bar{b}$ decay is kinematically allowed and represents the most

direct and powerful way to probe the coupling of the Higgs boson to the bottom-quark. It is predicted to have the largest branching ratio ($\sim 58\%$) [11, 12] for $m_H = 125$ GeV, however, the large background from the multi-jet production makes this search very challenging at hadron colliders. The most sensitive production mode to detect the $H \rightarrow b\bar{b}$ decay mode is the associated production of the Higgs boson with a vector boson, Z or W , whose leptonic decays ensure efficient triggering and a significant reduction of the multi-jet background. The observation of the $VH, H \rightarrow b\bar{b}$ process was claimed in 2018 by both the ATLAS [13] and CMS [14] experiments based on the combined datasets collected in the first and second Run of the LHC. The result provided a further confirmation of the validity of the mass generation mechanism and, at the same time, set strong limits on possible beyond the Standard Model physics.

As in the $VH, H \rightarrow b\bar{b}$ analysis, many other analyses utilize b -quarks in the final state. Therefore, the correct identification of the jets originating from b -quarks represents a fundamental task for the ATLAS experiment. In the current ATLAS simulation infrastructure [15], the GEANT4 software [16] does not take into account detector material interactions for the b - and c -hadrons, but only for their decay products. When traversing a pixel layer, the b -(c -)hadron can interact with the material leaving a hit that can be mistakenly associated to the track of one of its decay products. Studies have been performed to evaluate the impact of this effect on the tagging efficiency of b -jets and extrapolation uncertainties have been derived as a function of the b -jet transverse momentum.

To cope with the decrease in the track reconstruction efficiency for very high transverse momentum b -jets, a new tagging algorithm has been developed based on the difference in the number of hits counted in subsequent pixel layers. Preliminary studies on the performance of this new approach are shown in this thesis, but more detailed investigations will be needed to assess if the algorithm can be used to complement the track-based b -tagging algorithms improving their performance at high transverse momentum.

In this thesis the theoretical framework of the SM is described in Chapter 1, as well as the phenomenology of proton-proton interactions. The experimental setup i.e the ATLAS detector is outlined in Chapter 2, while the description of the techniques employed to reconstruct and identify the objects used in the analysis are summarised in Chapter 3. Chapter 4 provides a detailed description of the algorithms used to perform flavour tagging in ATLAS, while Chapter 5 summarises the performance studies of the impact of b - and c -hadron interactions with the ATLAS pixel detector material on the tagging efficiency of b -jets, as well as the studies on the new trackless b -tagging algorithm. The $VH, H \rightarrow b\bar{b}$ analysis is extensively described in Chapter 6, with particular attention to the main contributions of the author, i.e the modelling of the vector boson plus jets background with the implementation of a multivariate reweighting technique, the studies on the fit stability and reliability, as well as the fit results of the main VH analysis, and the cross-check VZ and di-jet mass analyses.

Theoretical Framework

The Standard Model (SM) of elementary particles is the theoretical framework that describes the fundamental constituents of matter and their interactions, and provides a mathematical formulation of the strong, weak and electromagnetic forces¹. The Brout–Englert–Higgs mechanism predicts the existence of a scalar boson field able to explain the mechanisms of mass-generation for all constituents of the theory.

Forty years of precision measurements at collider experiments have proven the remarkable capability of the SM not only in describing experimental observations, but also in providing predictions with a high level of accuracy. The weak force carriers, W and Z , were predicted by the electroweak model and eventually discovered at the Super Proton Synchrotron [17, 18]. In the same way, the top quark was predicted to complete the isospin doublet with the b -quark and its existence was finally confirmed at the Tevatron [19, 20] after 20 years of searches. The latest confirmation to the SM formalism is represented by the discovery of a new scalar particle highly compatible with the predicted Higgs boson [6, 7], the last missing piece of the theory.

Although the SM has provided so far the most accurate description of the fundamental particles and their interactions, several experimental indications are left unexplained by the theory. For example, the matter-antimatter asymmetry in the Universe, the origin of the neutrino masses, and the nature of the dark matter can only be explained by adding additional terms to the SM formalism. Many *Beyond the SM* (BSM) theories have been developed to extend the current theoretical framework, which is believed to be a low-energy limit of a more fundamental theory, but no experimental observation has yet confirmed one particular scenario.

The scope of this chapter is to introduce the theoretical framework used in this thesis, with particular attention to the formulation of the scalar Higgs sector. Furthermore, a discussion on the Higgs boson searches at the LHC will be provided, together with an overview of the physics of proton-proton collisions used to produce the data analysed in this work. The discussion is mainly based on Refs. [21–27].

1.1 Standard Model Formalism

The Standard Model is a renormalizable quantum field theory (QFT)² unifying quantum mechanics and special relativity.

In the theory of relativity, the well known relation between energy and mass $E = mc^2$

¹Fitting gravity in the Standard Model theoretical framework has proven to be very challenging and it is currently an active research area.

²A renormalizable QFT is a theory in which every divergence in the calculation can be regularised and cancelled, yielding finite expressions for measurable quantities [21].

implies that the number of particles of a system can not be defined a priori. In order to make quantum mechanics compatible with special relativity, it is necessary to abandon the concept of a *single particle wave function*. In fact, according to quantum mechanics, each particle of a system is represented by a wave function depending on the time and the three-dimensions space coordinates, $\psi(\mathbf{x}, t)$, whose $|\psi|^2$ gives the probability density of finding the particle at a given point in the space at a certain time. This formulation implies the conservation of the number of particles of the system which is in contrast with the relativistic principle.

In the relativistic formulation of quantum mechanics, the wave function of a single particle is promoted to a field depending on the four-dimensions space-time coordinates and described by a set of operators able to create or destroy particles. This procedure is called *second quantization*³ of the theory.

The dynamic of a quantum field system can be written in terms of a Lagrangian density $\mathcal{L}(x)$ depending on the space-time coordinates x and having dimension four in momentum-energy. Given that the physics of the system and the equation of motion are required to remain the same in every inertial reference system, this Lagrangian density must be invariant under Lorentz transformations, which are applied identically to every space-time point. Furthermore, $\mathcal{L}$ is also required to be invariant under a continuous group of local transformations called *gauge symmetries*, which can act differently at any point in space-time.

The constituents of matter, called *fermions*, are represented by spinors ψ with spin 1/2, whose free dynamic is described by the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \quad (1.1)$$

with γ^μ the Dirac matrices and m the mass of the fermion. Fermions interact with each other through the exchange of integer spin *bosons* mediating the electromagnetic (photon), weak ($W^\pm$ and Z^0) and strong (gluons) forces. These naturally arise in the theory from the requirement of local gauge invariance. In fact, local change of phase of the fermionic field leads to phase differences which would be observable if not compensated for in some way [21]. The introduction of new bosonic fields, whose transformation property cancels out with the one of the fermionic fields, restores the gauge invariance of the Lagrangian.

An example of this can be made considering Quantum Electrodynamics (QED) [28], the theory describing the interaction between charged fermions and the electromagnetic field. The corresponding Lagrangian can be initially written as the sum of the Dirac Lagrangian describing the free propagation of a fermion and the covariant representation of the electromagnetic field described by Maxwell equations [29]:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \quad (1.2)$$

³This term is used to differentiate the quantization in QFT theory with respect to the quantization of classical mechanics which does not respect Lorentz invariance.

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic tensor. This Lagrangian is invariant under global transformations of the U(1) group but, in order to be invariant under local gauge transformations described by the continuous function $\alpha(x)$:

$$\psi \rightarrow e^{i\alpha(x)}\psi, \quad (1.3)$$

the derivative $\partial_\mu\psi$ must be replaced by its *covariant* formulation:

$$D_\mu\psi = (\partial_\mu + ieA_\mu)\psi, \quad (1.4)$$

where e is the fermion electric charge and the field A_μ is a gauge vector field transforming as:

$$A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\alpha(x). \quad (1.5)$$

This way, the Lagrangian in Equation 1.2 can be re-written as:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} = \bar{\psi}(i\partial\!\!\!/ - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - e\bar{\psi}\mathcal{A}\psi, \quad (1.6)$$

with $\partial\!\!\!/ = \gamma^\mu\partial_\mu$ and $\mathcal{A} = \gamma^\mu A_\mu$. The term $-e\bar{\psi}\mathcal{A}\psi$ represents the interaction between the two Dirac spinors and the photon field. This interacting Lagrangian encodes the mathematical expression of the electromagnetic vertex which can be visualized via the Feynman diagrams' prescription [30] and used to calculate the cross section of many EM processes, for example the Compton scattering. The QED fundamental vertex is shown in Figure 1.1.

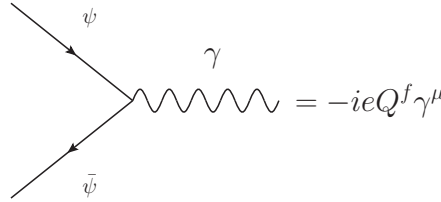


Figure 1.1: Fundamental QED vertex according to Feynman diagrams' prescription. The spinor ψ represents a fermion of charge Q^f .

According to Noether's Theorem [31], if the Lagrangian density is invariant under a continuous group of transformations, the system contains as many conserved currents along the equations of motion as the number of parameters describing the group. The interacting Lagrangian in Equation 1.6, is invariant under transformations of the gauge group U(1) which leads to the conservation of the electric charge.

More generally, the Standard Model is based on the following symmetry group:

$$\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y. \quad (1.7)$$

This group can be divided into separate groups: $\text{SU}(3)_C$ describes the strong sector in which the conserved charge is the *color* (C), while $\text{SU}(2)_L \times \text{U}(1)_Y$ describes the electroweak sector with the *weak isospin* (L stands for “left-handed”) and the *hypercharge* (Y) as

conserved currents. The full Lagrangian describing the SM theory can be written as the sum of the following pieces:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}, \quad (1.8)$$

where $\mathcal{L}_{\text{QCD}}$ is the Lagrangian of the strong interactions, also called Quantum Chromodynamics, $\mathcal{L}_{\text{EW}}$ is the Lagrangian of the electroweak interactions, and $\mathcal{L}_{\text{Higgs}}$, $\mathcal{L}_{\text{Yukawa}}$ are the Lagrangians of the Higgs sector.

1.1.1 Fundamental particles

In the Standard Model, fermions are divided into three generations, each of them containing an up- and a down-type quark, a charged lepton, and a neutral lepton called neutrino. The three generations can be distinguished by the mass of the constituent particles that differs by many orders of magnitude. For each fermion generation there exists an anti-fermion with the exact same properties, but opposite electric charge.

Quarks are color-charged particles divided into up- and down-type quarks depending on the carried electric charge, $2/3$ for up-type quarks and $-1/3$ for down-type quarks. The name “*up*” and “*down*” are taken from the names of the quarks belonging to the first generation. The second generation contains the *charm* and the *strange* quarks, while the third contains the *top* and the *bottom* quarks. Among all fermions, the quarks are the only elementary particles subject to all three fundamental forces. The quark flavour is conserved in the strong and the electromagnetic interactions but not in the weak one. In particular, the relation between the weak eigenstates and the mass eigenstates is encoded in the Cabibbo–Kobayashi–Maskawa (CKM) unitary and complex matrix V_{CKM} [32, 33]. The highest values of the probabilities are for transitions within the same family ($|V_{ud}|$, $|V_{cs}|$, $|V_{tb}|$) while the lowest values occur when the mixing is between the first and the third generation ($|V_{ub}|$ and $|V_{td}|$).

Leptons do not carry color charge, therefore they interact only via the electroweak force. Each generation of leptons is divided into two categories according to the electric charge: the electron (e), the muon (μ) and the tau (τ) carry one unit of electric charge, while the three corresponding neutrinos ν_e, ν_μ, ν_τ are neutral, therefore they can interact only via the weak force.

The electromagnetic, weak and strong forces are all mediated by spin-1 bosons. The electromagnetic interaction between charged particles is mediated by the exchange of a *photon*, which is massless and neutral. The massive $W^\pm$ and Z^0 bosons are the mediators of the weak interaction. The W^+ and W^- bosons carry $+1$ and -1 electric charge, respectively, while the Z^0 boson is neutral. The strong interaction is mediated by eight massless bosons called *gluons*, each carrying color charge, and so are subject themselves to the strong force. The last piece of the SM is the massive and neutral Higgs boson whose associated field gives mass to all elementary particles of the theory except for neutrinos. However, the observation of neutrino oscillations [34] proves that neutrino must be massive.

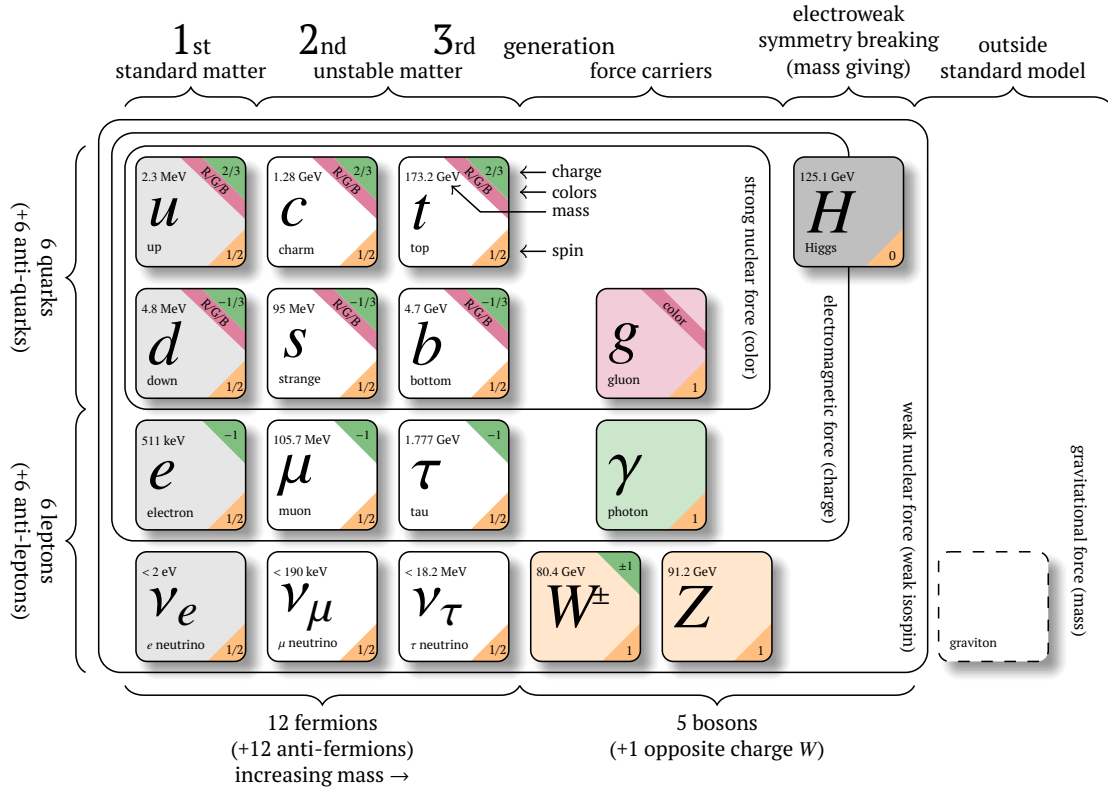


Figure 1.2: Classifications of the Standard Model particles. Fermions are divided into quarks and leptons which are, in turn, split into three generations depending on the mass. Each generation of quarks contains an up-type and a down-type quark, and each lepton family contains a charged and a neutral lepton. The three fundamental forces are carried by spin-1 bosons while the interaction of the elementary particles with the Higgs boson gives them a mass. The contours show which fermions are subject to which force [38].

Furthermore, from studies of the β decay of tritium [35], their masses are expected to be below $O(1\text{eV})$. Similarly to the quarks, the transition between flavour eigenstates and mass eigenstates is described by the Pontecorvo–Maki–Nakagawa–Sakata matrix [36, 37].

A scheme of the classification of the SM particles is shown in Figure 1.2.

1.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory describing the strong force. It is based on a $SU(3)_C$ symmetry group with *color* (red, blue and green) as conserved charged. The strong force is mediated by eight massless and colour-charged gluons which interact also among themselves due to the non-abelian⁴ structure of the $SU(3)_C$ group. The Lagrangian of the QCD can be derived following the same procedure as for the QED Lagrangian. The free QCD Lagrangian is required to be invariant under gauge transformations of the $SU(3)_C$ group:

⁴A non-abelian gauge symmetry group is a set of gauge transformations which do not obey the commutative law.

$$q_i(x) \rightarrow e^{i\vec{a}(x) \cdot \vec{T}} q_i(x), \quad (1.9)$$

where q_i is the Dirac spinor describing a quark carrying color i , $\vec{a}(x) = \alpha_1, \dots, \alpha_8$ is a vector of eight parameters as a function of the Lorentz coordinates x , and the group generators⁵ $\vec{T} = T_1, \dots, T_8$ are linearly independent traceless 3×3 matrices introduced by Gell-Mann to generalize the SU(2) Pauli matrices to a SU(3) theory. The QCD Lagrangian is therefore written as:

$$\mathcal{L}_{\text{QCD}} = \sum_{\text{flavours}} \bar{q}_i (i\gamma_\mu D^\mu - m)_{ij} q_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.10)$$

with i and j color indices of the quarks running from 1 to 8, and a color index of the gluons running from 1 to 8. The covariant derivative D_μ and the field strength tensor $G_{\mu\nu}^a$ have the following forms:

$$D_\mu = \partial_\mu + i\alpha_s T_a A_\mu^a, \quad (1.11)$$

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - \alpha_s f^{abc} A_{\mu,b} A_{\nu,c}, \quad (1.12)$$

with α_s the coupling constant quantifying the strength of the interaction between quarks, A_μ^a the gluon fields, and f^{abc} the structure constants defining the Lie Algebra⁶ of SU(3)_C. The non-abelian term $\alpha_s f^{abc} A_{\mu,b} A_{\nu,c}$ leads to tri-linear and quartic self-interactions of gluons which represents the main difference between QED and QCD. The coupling constant α_s depends on the transferred momentum in the interactions Q^2 : when $Q^2 \gg \Lambda_{\text{QCD}} \sim 200$ MeV, the quarks are considered as *asymptotically free*, while at lower energies the coupling constant increases resulting in the *confinement* of the quarks in colourless bound states called hadrons. The phenomenon of confinement is related to the potential energy of the QCD force, which is dominated by $V(r) \sim r$ at high distances between the quarks. This means that extracting a quark from its bound state is energetically disfavoured such that new quark-antiquark pairs come from the vacuum to form new hadrons with the original quarks. This process is called *hadronization* and it is the reason why quarks and gluons are not observed as free states but, instead, are always confined into colourless bound states. Hadrons can be *mesons* if they contain a quark and an anti-quark, or *baryons* if they contain three quarks.

⁵Any transformation of the symmetry group can be obtained by a linear combination of the generators of the group with some coefficients.

⁶The Lie Algebra defines the commutation rules to which the generators of a symmetry group obey. In SU(3)_C the commutation rules among the generators are encoded in the relations:

$$[T^a, T^b] = if^{abc} T^c \quad (1.13)$$

$$[[T^a, T^b], T^c] + [[T^b, T^c], T^a] + [[T^c, T^a], T^b] = 0 \quad (1.14)$$

1.1.3 Electroweak Sector

The electroweak theory is the unified theory describing the electromagnetic and weak interactions in the framework of the $SU(2)_L \times U(1)_Y$ gauge symmetry, proposed by Glashow [1], Weinberg [2] and Salam [3], independently.

Two types of weak interactions are experimentally observed: the *charged current* and the *neutral current* interactions. The first are mediated by a charged boson, W^+ or W^- , and involve only particles with left-handed *chirality*⁷. A pioneering theory based on a four-fermion interaction was proposed by Fermi in 1934 [40] to explain the phenomenon of β decay, however, the Lagrangian of the theory was only valid at leading order with no predictive power for higher corrections.

The second type of interactions involving neutral current are mediated by the neutral boson Z^0 and involve also particles with right-handed chirality. Predicted by the GSW model, they have been a puzzle of particle physics until their discovery in 1973 at the Gargamelle experiment [41].

After the discovery of parity violation in weak interactions, such as the asymmetry in the angular distribution of the products in the β^- decay by Chien-Shiung Wu [42], it became clear that the Dirac spinors had to be split into left-handed and right-handed chirality components, and that the weak interactions allowed only for couplings between left-handed particles and right-handed anti-particles. Mathematically, the Dirac spinor can be projected onto its left-handed and right-handed chirality components using the γ^5 Dirac matrix:

$$\psi_L = P_L \psi = \frac{1}{2}(1 - \gamma^5)\psi, \quad (1.15)$$

$$\psi_R = P_R \psi = \frac{1}{2}(1 + \gamma^5)\psi. \quad (1.16)$$

In this formulation, left-handed particles are grouped in doublets:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, \quad (1.17)$$

while right-handed particles form singlets:

$$e_R, \mu_R, \tau_R, u_R, d_R, s_R, c_R, b_R, t_R. \quad (1.18)$$

A right-handed neutrino is not introduced since there is still no observation of this particle. For the Lagrangian of the electroweak interactions to be invariant under the gauge symmetry group $SU(2)_L \times U(1)_Y$, the spinors ψ_L and ψ_R need to transform according to a non-abelian $SU(2)$ rotation and a local change of phase of $U(1)$:

⁷The chirality defines whether the particle transforms in a left- or right-handed representation of the Poincaré group [39]. It is also related to the concept of helicity, which is calculated as the sign of the projection of the spin vector on the momentum vector. Left helicity is negative, right helicity is positive. For massless particles the two concepts coincide.

$$\psi_L \rightarrow \underbrace{e^{i\vec{\alpha}(x)\vec{\sigma}/2}}_{\text{SU}(2)_L} \underbrace{e^{i\beta(x)Y}}_{\text{U}(1)} \psi_L \quad \psi_R \rightarrow \underbrace{e^{i\beta(x)Y}}_{\text{U}(1)} \psi_R, \quad (1.19)$$

with $\vec{\alpha}(x)$ a vector of three parameters as a function of the Lorentz coordinates x , $\vec{\sigma}$ the Pauli matrices generating SU(2), $\beta(x)$ a real parameter, and Y the generator of the U(1) group.

According to Noether's theorem, the invariance of the Lagrangian under these transformations leads to the conservation of two quantities: the *isospin* T and the *hypercharge* Y . The latter, is related to the electric charge and to the third component of the weak isospin following the *Gell-Mann-Nishijima* relation:

$$Y = 2(Q - T_3). \quad (1.20)$$

To preserve the invariance of the Lagrangian after the fields transformations (see Equation 1.19), the covariant derivative takes the form:

$$D_\mu = \partial_\mu + ig W_\mu^a T_a + ig' B_\mu \frac{Y}{2}, \quad (1.21)$$

where T^a are the three components of the weak isospin, built via the Pauli matrices $T^a = \sigma^a/2$ ($a = 1, 2, 3$), and Y is the group generator of weak hypercharge. Enforcing a gauge invariance on the Lagrangian introduces four gauge bosons, three W_μ^a ($a = 1, 2, 3$) for the SU(2)_L group and one for the U(1)_Y called B_μ . The coupling constants between the fermions and these new bosons are indicated as g and g' .

Considering only massless particles, the full electroweak Lagrangian becomes:

$$\mathcal{L}_{\text{EW}} = i\bar{\psi}_L \gamma^\mu D_\mu \psi_L + i\bar{\psi}_R \gamma^\mu D_\mu \psi_R - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \quad (1.22)$$

where:

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon_{abc} W_\mu^b W_\nu^c, \quad (1.23)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (1.24)$$

are the tensor fields describing the three SU(2)_L vector fields W_μ^a ($a = 1, 2, 3$) using the structure constants ϵ_{abc} , and the single vector field B_μ of U(1)_Y.

The physical fields corresponding to the experimentally observed vector bosons $W^\pm$, Z^0 , and γ can be obtained as linear combinations of the above four fields:

$$A_\mu = \sin \theta_W W_\mu^3 + \cos \theta_W B_\mu, \quad (1.25)$$

$$Z_\mu = \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu, \quad (1.26)$$

$$W_\mu^\pm = \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}}, \quad (1.27)$$

where A_μ is the physical field describing the photon, while Z_μ and $W_\mu^\pm$ describe the vector

bosons Z^0 and $W^\pm$, respectively. The angle θ_W represents the mixing angle of the weak interactions and is called *weak mixing angle* (sometimes called *Weinberg angle*). It can be written in terms of the coupling constants g and g' and can be determined at leading order from the measurement of the $W^\pm$ and Z^0 masses:

$$\sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \quad \cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}}. \quad (1.28)$$

The fundamental vertices for the weak charged and neutral currents are shown in Figure 1.3.

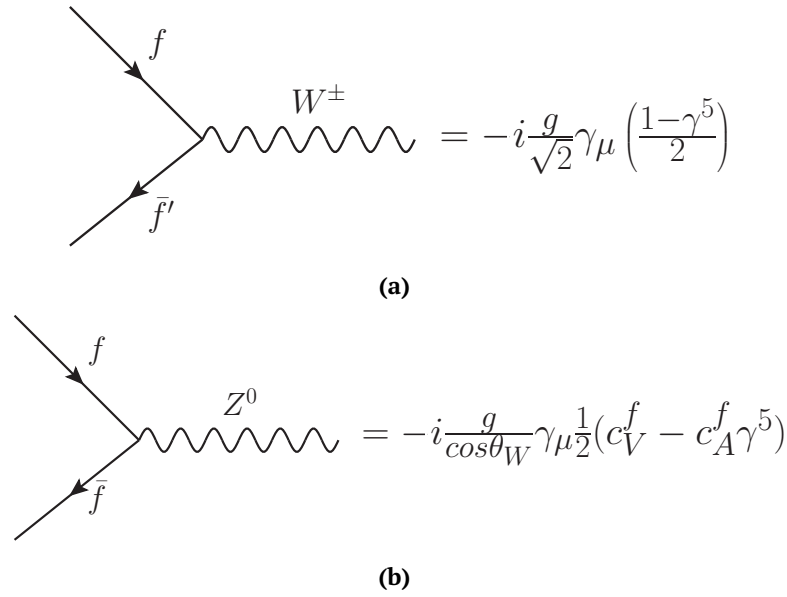


Figure 1.3: Charged and neutral currents vertices according to Feynman diagrams' prescription. The terms c_V^f and c_A^f are the vector and axial couplings of the fermion type f of charge Q^f and can be written as: $c_V^f = T_3^f - 2\sin^2 \theta_W Q^f$ and $c_A^f = T_3^f$.

Any predictive theory should be renormalizable i.e. there must exist a procedure that eliminates the infinities arising in the calculation of physical quantities, such as cross sections and decay rates, at low and high energies (named infrared and ultraviolet divergences, respectively⁸). To cure these divergences, the parameters and fields of the theory are redefined as a function of a *renormalization scale* to absorb the infinities, yielding finite measurable quantities. This procedure is generically named *renormalization*; more details can be found in Refs. [21, 43]. In order to formulate a renormalizable electroweak theory without breaking the symmetry, its Lagrangian must not contain mass terms. This means that both gauge bosons and fermions in the EW theory are massless. The masses are generated, without breaking the renormalizability, through the mechanism of *spontaneous electroweak symmetry breaking*, also known as *Brout–Englert–Higgs mechanism*.

⁸The infrared divergences come from the region of integration where the momentum in the loop tends to zero. The ultraviolet divergences come from the region of integration where the momentum in the loop tends to infinity.

1.2 Mass generation and Higgs mechanism

The electroweak Lagrangian in Equation 1.22 has been written assuming massless particles, as the introduction of a mass term $m\bar{\psi}\psi = m\psi_L\psi_R + m\psi_R\psi_L$ would mix the left-handed and right-handed components breaking the symmetry. Instead, the masses are generated through the mechanism of spontaneous breaking of the electroweak symmetry according to which breaking the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry into $SU(3)_C \times U(1)_Y$ gives mass to the weak bosons of $SU(2)_L$, while photon and gluons remain massless. Fermion masses, instead, are generated by the Yukawa interaction terms with the Higgs field.

1.2.1 Brout-Englert-Higgs mechanism

The Brout-Englert-Higgs mechanism [4, 5] starts with the introduction of a doublet of $SU(2)_L \times U(1)_Y$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \quad (1.29)$$

The Lagrangian associated to this field can be written as:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \quad (1.30)$$

where the covariant derivative is the same as in Equation 1.21 and the potential $V(\phi)$ has the following form:

$$V(\phi) = -\mu^2(\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2 = -\mu^2|\phi|^2 + \lambda|\phi|^4. \quad (1.31)$$

This potential contains a mass term with constant μ^2 , and a self-interaction term with coupling constant $\lambda > 0$ which allows the energy to be bounded from below. Depending on the value of μ , the potential $V(\phi)$ assumes different shapes:

1. $\mu^2 > 0$: the potential has a unique minimum at $\phi = 0$ corresponding to the ground state, i.e the vacuum. This means that the ground state has zero vacuum expectation value and that the $\mathcal{L}_{\text{Higgs}}$ describes a scalar particle with mass μ^2 and massless gauge bosons;
2. $\mu^2 < 0$: the ground state has a non-zero vacuum expectation value.

Since the vector bosons of $SU(2)_L$ are observed to be massive, μ^2 is assumed to be < 0 and the potential then shows the shape of a “*Mexican hat*”, as shown in Figure 1.4. In this case, the expectation value of the vacuum (VEV) can be derived by minimizing the derivative of the potential:

$$\frac{dV(\phi)}{d\phi} = 0 \quad \Rightarrow \quad v = \sqrt{\frac{-\mu^2}{2\lambda}}, \quad (1.32)$$

and it is measured to be $v = 246.22$ GeV [45] from its relation with the W boson mass (see later).

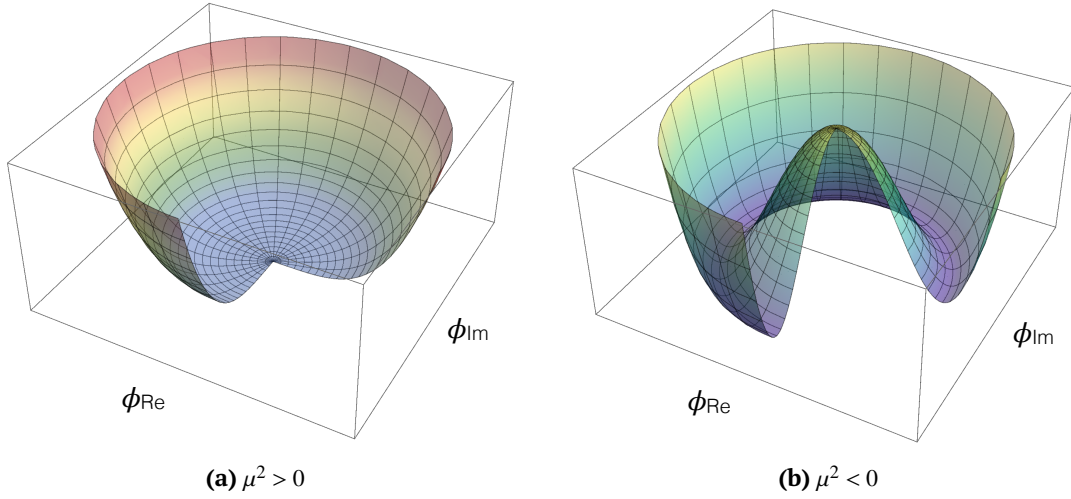


Figure 1.4: Effective potential for different values of μ^2 [44].

The $SU(2)_L \times U(1)_Y$ is spontaneously broken when the system falls into one of the multiple ground states, for example:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.33)$$

Expanding around the chosen ground state, the field can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (1.34)$$

with $H(x)$ a real scalar field. Evaluating the kinetic term $(D_\mu \phi)^\dagger (D^\mu \phi)$ of the Lagrangian in Equation 1.30 in $\phi(x)$:

$$(D_\mu \phi)^\dagger (D^\mu \phi) = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) + \phi^\dagger (g \mathbf{W}_\mu \cdot \mathbf{T}) + \frac{g'}{2} (B_\mu \cdot Y)^2 \phi = \quad (1.35)$$

$$= (\partial_\mu \phi)^\dagger (\partial^\mu \phi) + \left(\frac{g(v+H)}{2} \right)^2 W_\mu^+ W^{\mu-} + \frac{1}{2} \left(\sqrt{g^2 + g'^2} \frac{(v+H)}{2} \right)^2 Z_\mu Z^\mu, \quad (1.36)$$

the observable gauge bosons $W_\mu^\pm$ and Z_μ defined in Equation 1.25 appear and their mass is related to the VEV⁹ with coupling constants g and g' :

$$m_W = \frac{v}{2} g \quad \text{and} \quad m_Z = \frac{v}{2} \sqrt{g^2 + g'^2} \quad \text{with} \quad \frac{m_W^2}{m_Z^2 \cos^2 \theta_W} = 1. \quad (1.37)$$

As expected, there is no mass term for the A_μ field therefore the photon remains massless. The other terms in Equation 1.36 of the covariate derivative account for the interaction

⁹From the relation between the W boson mass and its coupling g with the Higgs boson, the vacuum expectation value can be derived as:

$$v = \frac{1}{\sqrt{\sqrt{2} G_F}} = 2 \frac{m_W}{g},$$

with G_F the Fermi constant measured from the weak muon decay.

between a single or two Higgs bosons with a pair of massive gauge bosons, as well as trilinear and quartic self-interactions of the Higgs boson. The different types of Higgs boson interaction vertices are represented in Figure 1.5 according to the Feynman diagrams' prescriptions.

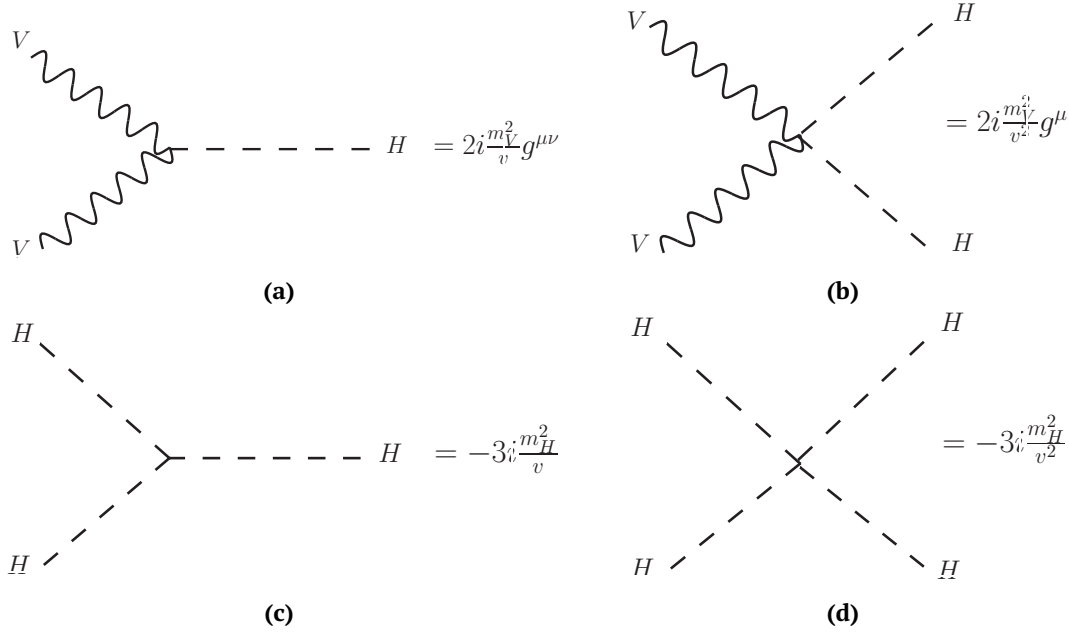


Figure 1.5: Higgs boson interaction vertices according to Feynman diagrams' prescription. Figures (a) and (b) describe the Higgs boson interaction with the vector bosons $W^\pm$ or Z^0 , while (c) and (d) represent the self-couplings of the Higgs boson.

The mass of the Higgs boson field can be derived by evaluating the Lagrangian in Equation 1.30 in the field $\phi(x)$ defined in Equation 1.34 and is given by:

$$m_H = \sqrt{2} |\mu| = \sqrt{2}\lambda v. \quad (1.38)$$

1.2.2 Yukawa coupling to the fermion sector

The fermions acquire mass by interacting with the Higgs boson field. As the introduction of ad-hoc mass terms would break the gauge invariance of the electroweak Lagrangian, the interaction is expressed in terms of an invariant Yukawa Lagrangian of the form:

$$\mathcal{L}_{\text{Yukawa}} = -g_{ffH} \bar{L}_L \phi l_R + h.c., \quad (1.39)$$

where L_L is the left-handed doublet for leptons or quarks, l_R is the correspondent right-handed singlet and ϕ is the Higgs boson field. Expanding the Higgs field around the ground state, the Lagrangian becomes:

$$\mathcal{L}_{\text{Yukawa}} = -g_{ffH} \frac{v}{\sqrt{2}} \bar{L}_L l_R - g_{ffH} \frac{H(x)}{\sqrt{2}} \bar{L}_L l_R + h.c., \quad (1.40)$$

where it can be observed that the Higgs boson coupling to fermions is proportional to the fermions mass:

$$m_f = -g_{ffH} \frac{v}{\sqrt{2}}. \quad (1.41)$$

The Higgs-to-fermion interaction vertex is illustrated in Figure 1.6.

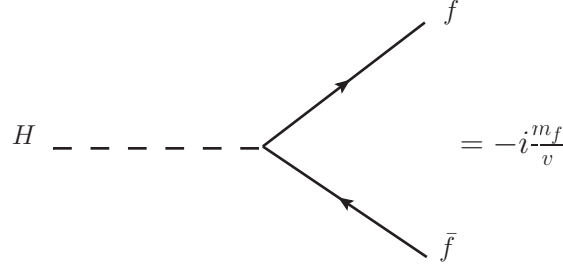


Figure 1.6: Interaction vertex of the Higgs boson to fermions as defined by the Yukawa theory.

1.3 Higgs boson searches at the LHC

The discovery of the Higgs boson by the ATLAS and CMS experiments at the LHC was a crucial step in the history of particle physics. It finally confirmed the mechanism of electroweak symmetry breaking and marked the beginning of the Higgs precision era. This section introduces a phenomenological approach to study the Higgs boson, focusing on detectable production and decay modes.

1.3.1 Higgs boson production modes

At the LHC, there are several Higgs boson production modes coming from quark or gluon initiated collisions. Given the high density of low momentum gluons inside the proton, *gluon-gluon fusion* (ggF) is the dominant process with a cross section of ~ 49 pb for the production of a Higgs boson with mass $m_H = 125$ GeV at 13 TeV [12]. As seen in Section 1.2.2, the coupling of the Higgs boson to the fermions is proportional to their mass, therefore, the production is mainly mediated by a loop of heavy top quarks. As the only particle produced in the interaction is the Higgs boson, as shown in Figure 1.7 (a), this mode is characterized by a small transverse momentum.

The cross section of the second largest production mode, *vector boson fusion* (VBF), is an order of magnitude smaller than the ggF one. The Higgs boson is produced within the rapidity gap¹⁰ of two high energy jets produced in the forward and backward regions. The correspondent Feynman diagram of this production mode is shown in Figure 1.7 (b).

The third production mode is the *Higgs-Strahlung* (VH) shown in Figure 1.7 (c), where the Higgs boson is produced in association with a vector boson ($W^\pm$ or Z^0). Its cross

¹⁰As explained in Chapter 2, the *rapidity* is defined as $y = \frac{1}{2} \ln\left(\frac{E+p_L}{E-p_L}\right)$ where p_L is the projection of the particle momentum along the beam axis

section is very small, but the leptonic decays of the vector boson help to suppress the QCD background, enhancing the signal sensitivity.

The production mode having the smallest cross section is the associated production with top quarks ($t\bar{t}H$) and is particularly important for the measurement of the Higgs boson coupling to the top quark.

The cross sections of the four production modes at a centre-of-mass energy of $\sqrt{s} = 13$ TeV are shown in Figure 1.8.

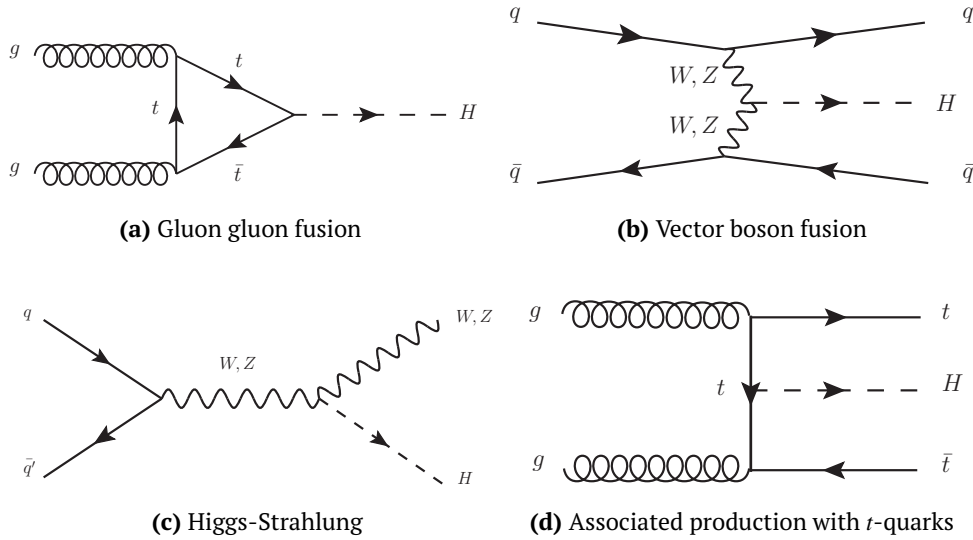


Figure 1.7: Feynman diagrams of the four Higgs production modes at the LHC.

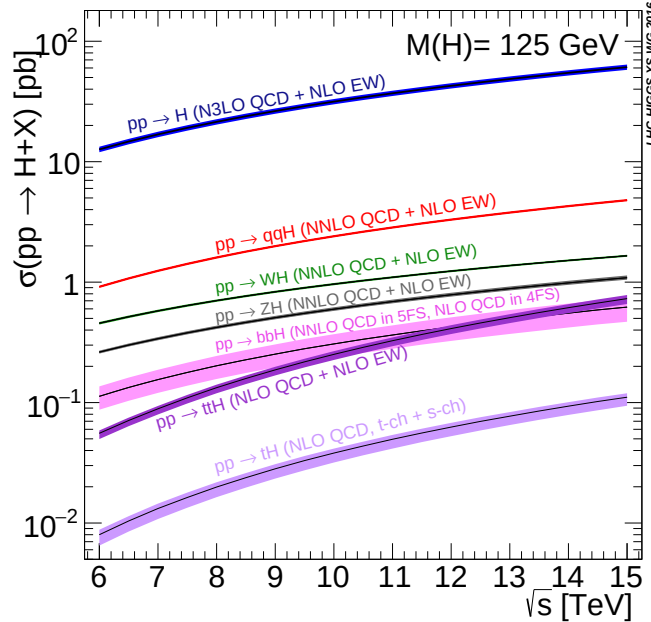


Figure 1.8: Cross section as a function of $\sqrt{s}$ for a Higgs boson of mass $m_H = 125$ GeV. The blue line is the cross section for the ggF production mode, the green and the black ones are the cross sections for WH and ZH , respectively, and the violet line is the $t\bar{t}H$ cross section [46].

1.3.2 Higgs boson decay channels

The probability for a particle to decay into a specific final state X is given by the so-called Branching Ratio (BR), defined as the ratio of the partial decay width¹¹ of the specific final state over the total decay width of the particle, which is calculated as the sum over the partial decay widths of each final state X_i :

$$\text{BR}(H \rightarrow X) = \frac{\Gamma(H \rightarrow X)}{\sum_i \Gamma(H \rightarrow X_i)}. \quad (1.42)$$

Due to its mass, the Higgs boson can decay only into particles kinematically allowed. For instance, the top quark is too heavy to allow for the decay of an on-shell Higgs boson to a pair of t -quarks, therefore, the Higgs coupling with the top quark can only be directly probed via the $t\bar{t}H$ production mode. In the same way, in case of the decay to two W or Z bosons, at least one of them has to be produced off-shell, reducing the probability for this decay mode.

Figure 1.9 shows the Standard Model Higgs boson decay branching ratios in the 120-130 GeV Higgs mass range, but in the following the $\text{BR}(m_H = 125 \text{ GeV})$ will be considered. The $b\bar{b}$ decay mode is dominant with a BR of $\sim 58\%$, however, it suffers from a large contamination of QCD processes. It represents the decay mode studied in this thesis and more details are given in Chapter 6. The decay into $c\bar{c}$ has a low branching ratio and, as the $b\bar{b}$ decay mode, it is largely contaminated by QCD processes like $gg \rightarrow c\bar{c}$. For this reason, this channel has not been observed yet. The couplings to leptons are opened thanks to the decay into $\tau\bar{\tau}$. However, this is also contaminated by the QCD background yielding to a poor signal over background ratio. The decays to lighter leptons like muons are also allowed but they are largely suppressed due to the very small mass of these particles.

Among the bosonic decay channels, the decay to WW has the second largest BR at $m_H = 125 \text{ GeV}$. The W bosons can subsequently decay into a quark/anti-quark pair or into a lepton and a neutrino. In the first case, this decay is overwhelmed by the QCD background, while the presence of the neutrino results in a low mass resolution. A cleaner channel, at the price of a much smaller BR, is given by the decay to ZZ , each of them decaying into a pair of charged leptons (e or μ).

Decays into massless gauge bosons are also possible requiring loops of virtual heavy quarks (t or b) or heavy gauge bosons. The decay into a pair of gluons has the third largest BR among the bosonic decay modes, but it is also indistinguishable from the QCD background. On the contrary, the decay to two photons has a much clearer signature and a very good mass resolution. For these reasons, $H \rightarrow \gamma\gamma$ together with $H \rightarrow ZZ^* \rightarrow 4l$ have been chosen as “golden channels” in the search for the Higgs boson at the LHC. A similar decay $H \rightarrow Z\gamma$ is also possible but it has a much smaller BR when considering $Z \rightarrow e^+e^-$ and it is contaminated by the Z radiative processes.

¹¹The decay width of a process Γ is the probability per unit time that a given particle will decay.

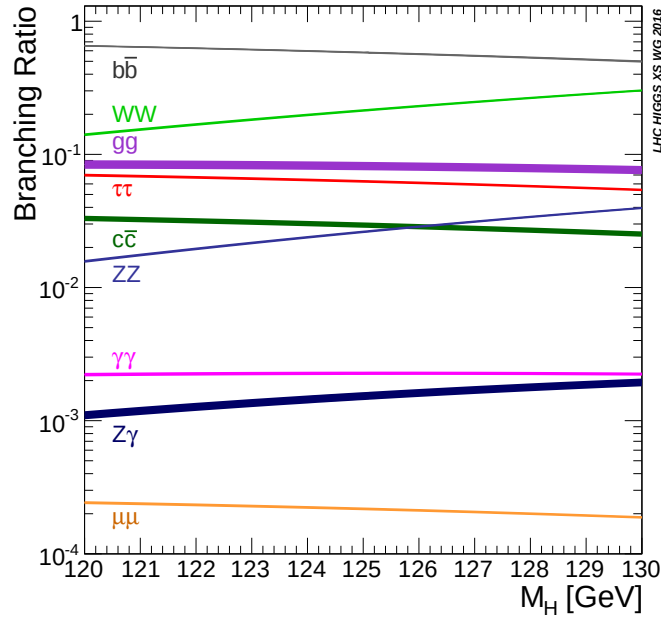


Figure 1.9: Higgs boson decay branching ratios in the 120-130 GeV mass range [46].

1.3.3 Higgs boson discovery

The first direct observation of a new boson compatible with the Higgs boson was announced in July 2012 by both ATLAS and CMS Collaborations [6, 7]. The dataset used by both experiments was collected at 7 TeV in 2011 and 8 TeV in 2012. An excess was found around a mass of 125 GeV, based on a combination of all analyses available at the time, with the main contributions coming from the $H \rightarrow ZZ^* \rightarrow 4l$, $H \rightarrow \gamma\gamma$ and $H \rightarrow W^+W^- \rightarrow l^+\nu_l l^-\bar{\nu}_l$ channels.

Figure 1.10 shows the di-photon and four-lepton invariant mass distributions for the ATLAS experiment. An excess of events is visible at the same mass in both channels.

The latest available combination of the ATLAS and CMS analyses in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$ dates back to the end of Run 1 with a measured mass of [47]:

$$m_H = 125.09 \pm 0.24(\text{stat}) \pm 0.11(\text{syst}) \text{ GeV}. \quad (1.43)$$

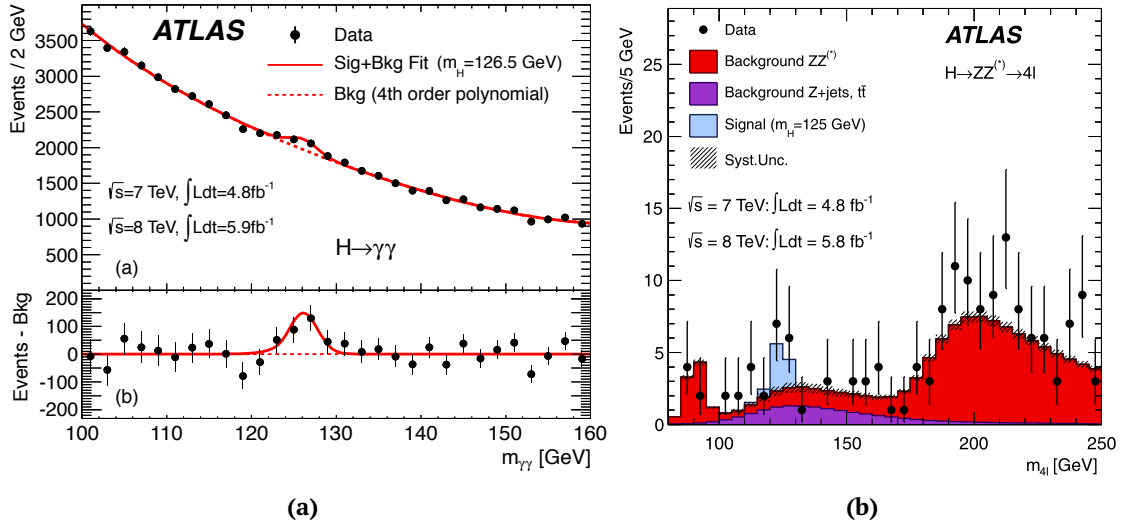


Figure 1.10: Di-photon (a) and four-lepton (b) invariant mass distributions for the 4.8 fb^{-1} and 5.8 fb^{-1} datasets collected by ATLAS detector at 7 TeV and 8 TeV, respectively [6].

1.4 Simulation of proton-proton collisions at the LHC

The Lagrangian formalism discussed so far provides the necessary tools to build measurable observables of the dynamics of elementary particles such as the *cross section* of a process. This represents the probability for two initial particles a and b to interact and produce the final state c . It can be written using Fermi's *golden rule* [48, 49] stating that the reaction rate is proportional to the product of the phase space and the absolute square of the amplitude term:

$$\hat{\sigma}_{ab \rightarrow c} \propto \int \frac{1}{F} |\mathcal{M}_{ab \rightarrow c}|^2(\Phi) d\Phi, \quad (1.44)$$

with F the incident flux, $|\mathcal{M}_{ab \rightarrow c}|^2(\Phi)$ the *transition amplitude* squared proportional to the so-called *matrix element*, and $d\Phi$ the phase space element of the final state.

1.4.1 Definition of cross-section for pp collisions

The formalism introduced in Equation 1.44 for a proton-proton collision is complicated by the composite nature of the protons which are made of three *valence quarks* (uud) and virtual pairs of quarks, referred to as *sea quarks*, and gluons. All components are referred to as *partons*. When the energy of the proton is much higher than the binding force between its constituents, the interaction happens between two asymptotically free quarks or gluons carrying only a fraction of the total momentum of the incoming protons. At the same time, multiple low energy interactions take place between the other partons, which are usually a source of QCD background for many analysis.

The general formula of the cross section for a collision between proton A and proton B

at a centre-of-mass energy s is given by the *factorization theorem*:

$$\sigma_{AB \rightarrow c}(s, \mu_R, \mu_F) = \sum_{a,b} \int_0^1 \int_0^1 dx_a dx_b f_{a/A}(x_a, \mu_F) f_{b/B}(x_b, \mu_F) \hat{\sigma}_{ab \rightarrow c}(\hat{s}, \mu_R, \mu_F), \quad (1.45)$$

where the partonic cross section $\hat{\sigma}_{ab \rightarrow c}$ occurs at a reduced centre-of-mass energy $\hat{s} = x_a x_b s$, with x_a (x_b) the momentum fraction carried by parton a (b). It can be written as a perturbative sum of matrix elements weighted by the strong coupling α_s :

$$\hat{\sigma}_{ab \rightarrow c} = \alpha_s^k \sum_{n=0}^{N_{\text{QCD}}} \alpha_s^n c^{(n)}(\hat{s}, \mu_R, \mu_F), \quad (1.46)$$

where the coefficients $c^{(n)}$ encode the kinematic of the process, k is the lowest possible order of the strong coupling for such process, and the terms in the sum define the order of the calculation. If only the first term is present ($n = 0$), the calculation is said to be at *Leading-Order* (LO) and represents the *tree-level* process. The term with $n = 1$ defines the *Next-To-Leading-Order* (NLO) calculation while the expansion up to $n = 2$ is the *Next-To-Next-To-Leading-Order* (NNLO) and so on. Most of the current event generators (see Section 1.4.2 for more details on the event simulation) employ LO or NLO matrix elements. In order to take into account higher order corrections, the so-called *k-factor* is used to reweight the LO (NLO) cross section and it is obtained by the ratio of the total NLO (NNLO) cross section to the LO (NLO) one.

The partonic cross section also depends on the renormalization scale μ_R through the coupling constant α_s in order to absorb singularities arising from virtual loop contributions at each order of the calculation.

The functions $f_{a/A}(x_a, \mu_F)$ and $f_{b/B}(x_b, \mu_F)$ in Equation 1.45 are the *parton density functions* (PDFs) representing the distribution of quarks/gluons carrying a fraction x of the total momentum of the proton. The dependency of the PDFs on the *factorization scale* (μ_F) is introduced to renormalise the singularities arising from collinear gluon emission and gluon splitting. The PDFs are typically derived in Deep Inelastic Scattering (DIS) experiments (e.g. HERA) [50, 51] at a certain energy scale Q_0^2 (which is typically the momentum transferred in the hard scattering) and extrapolated to any energy scale using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [52]. An example is shown in Figure 1.11 for both valence and sea quarks.

At high energies, as those reached by the LHC, sea quarks and gluons are dominant with respect to the valence quarks, and each of them carries a small fraction x of the total proton momentum. Given that the gluon density is the largest one, many processes can be gluon-initiated, and the LHC is often referred to as a “gluon collider”.

1.4.2 Event simulation

At the LHC, a single high energy pp collision produces hundreds of particles with a momentum range spanning many orders of magnitude. Thanks to the factorization

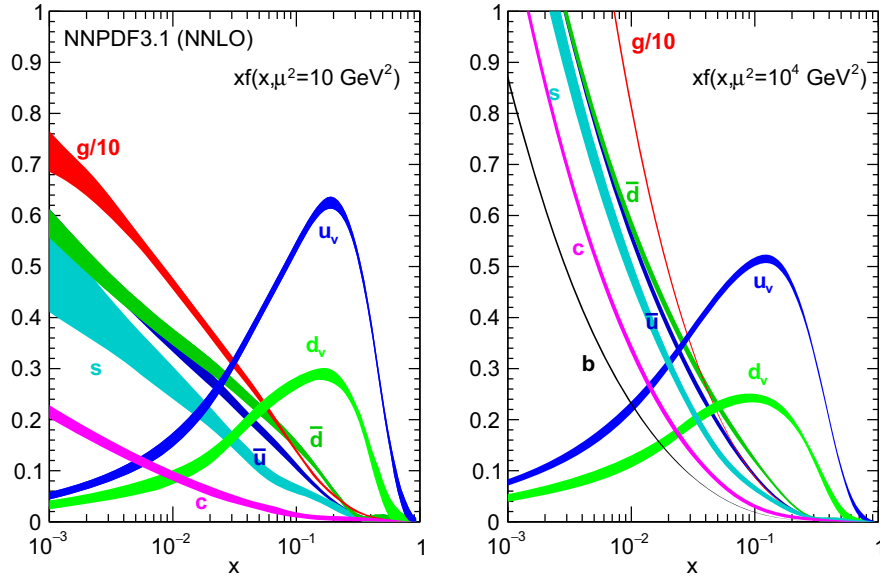


Figure 1.11: Parton distribution functions evaluated at ~ 3 GeV on the left and 10 GeV on the right [53].

theorem, the description of such interactions can be divided into different regimes according to the transferred momentum. At high energy scales, the hard scattering between the constituents of the protons can be perturbatively computed in terms of the matrix element of the process. At lower energies of the order of 1 GeV, many other physics processes happen whose non-perturbative nature leads to the use of simulation with tunable parameters to describe the data. The simulation relies on several available programs based on Monte Carlo methods [25, 27] able to describe the different phases of the proton-proton scattering. As shown in Figure 1.12, a $pp \rightarrow X$ interaction can be divided into different steps: the hard scattering, the parton shower, the hadronization, the underlying event, and the decay of unstable particles in the final state.

Parton Density Functions (PDF) The parton density functions are obtained from a global fit to the input datasets from DIS, fixed-target Drell–Yan, and collider experiments from the Tevatron and the LHC, and are extrapolated over a wide range of energies. Several PDF sets are available on the market (i.e CT14 [55], NNPDF3.0 [56] or MMHT14 [57], PDF4LHC [58]), which differ for the value of the factorization scale, the input datasets, and the order of perturbation at which they are calculated.

Matrix Element (ME) The matrix element represents the hard scattering part of the process and it is often calculated using Feynman diagrams. In the case of NLO or NNLO simulations, many diagrams describing virtual loops are added to the computation making the simulation very slow. The ME is calculated at the renormalization scale μ_R which is typically identified as the scale of the process. For example, if a resonance of mass M is produced in the hard scattering, μ_R is often chosen as $\mu_R = M^2$.

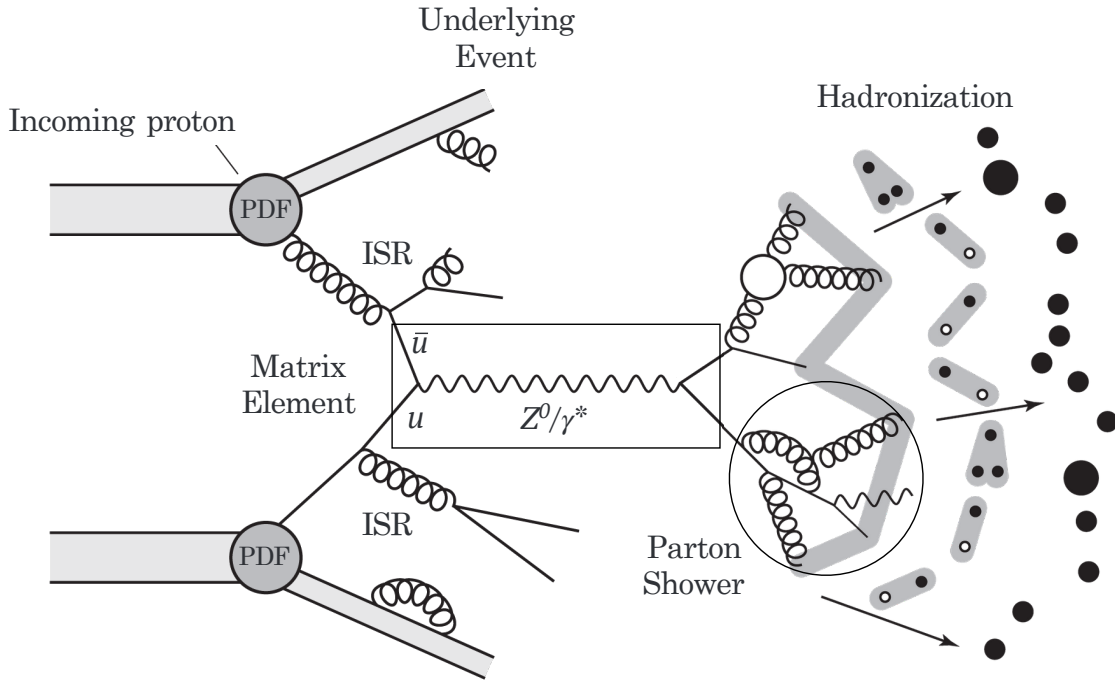


Figure 1.12: Schematic representation of a proton-proton collision as simulated by the event generator. Two partons from the colliding protons, described by the PDFs, undergo a hard interaction (Matrix Element). Additional hard QCD radiation is produced (Parton Shower), before the final state partons start hadronizing (Hadronization). Additional activity in the event originates from the Underlying Event [54].

Parton Shower (PS) The partons involved in the hard scattering and those produced in the final state can radiate gluons which can, in turn, split into quark-antiquark pairs or gluons generating a shower of outgoing particles. This is simulated using parton shower algorithms which are typically built as an evolution from the energy scale of the hard process down to low scales, of order 1 GeV, associated to the confinement of the partons inside the hadrons. The radiation from incoming particles is typically referred to as *initial state radiation* (ISR), while emissions from the final state partons is named *final state radiation* (FSR). In both cases, the probability to radiate a parton is controlled by the so-called *Sudakov form factor* [59], however, in the case of ISR, the ratio between the PDFs of the colliding partons has to be included. The Sudakov form factor describes the probability for a parton q *not to* radiate in the time interval Δt . This depends on the scale of the hard-scattering, the parton type and the initial x of the parton. The radiation emission is controlled by a “time-dependent” variable t ensuring that the subsequent emission happens at lower energy with respect to the previous one. The choice of the evolution variable can differ across generators: the opening angle (HERWIG [60]) or the transverse momentum (PYTHIA8 [61]) are two possible choices.

When the ME and the PS are simulated with two different generators, the two have to be carefully matched to remove the overlapping diagrams and cover the entire phase space. The ME is typically used to describe the production of hard and well separated particles, while the PS is employed for soft and collinear particles. The scale at which

the two simulations are combined is called *matching scale*.

Hadronization When the final state partons reach the energy scale of ~ 1 GeV, partons end up in colourless hadrons. The models describing the process of hadronization depends on many free parameters that have to be tuned to match the data. The two main hadronization classes in use are the *Lund string model* [62] used by PYTHIA and the *cluster model* [63] used by HERWIG and SHERPA [64] generators. The main difference between the two is that the former transforms the parton system directly into a hadron, while the latter introduces an intermediate stage in which clusters of partons with mass scale of ~ 1 GeV are formed.

Underlying Event (UE) In a pp collision the underlying event refers to the additional activity which is not related to the hard process. This includes *multi-parton interactions* of the other constituents of the protons not involved in the hard scattering (spectators) as well as the beam-beam remnants. The UE is simulated using phenomenological models tuned to the data.

Beside these events, also *pile-up* needs to be accounted for in the simulation (more details in Chapter 2). This refers to additional interactions occurring between protons in the same bunch (in-time pile-up) or in different bunches (out-of-time pile-up).

Decays of unstable particles Following the hadronization step, several unstable particles are produced which are then decayed into stable particles that can interact with the detector (see Chapter 5). When using the PYTHIA generator, the decays of heavy hadrons and τ leptons are often simulated using the EvtGen [65] and the TAUOLA [66] external packages, respectively. HERWIG++ and SHERPA include an internal simulation of both hadronic (except for heavy flavour decays) and τ decays.

The ATLAS Detector at the LHC

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [67, 68] at the European Centre for Nuclear Research (CERN) is the world's largest particle accelerator ever built to explore the high-energy frontier of particle physics. It is a circular proton-proton collider totalling 26.7 km long in circumference and about 100 meters underneath the border between France and Switzerland, near the city of Geneva.

The primary goal of the LHC was to finally claim the discovery or the absence of the Higgs boson, as well as searching for physics beyond the Standard Model. The large increase in the centre-of-mass energy $\sqrt{s}$, with respect to the previous generation of colliders, was decided to have a Higgs boson production cross section above a few picobarns for a large mass spectrum. Such a high energy could be reached only accelerating relatively heavy particles, such as protons. In fact, given that the radiated energy per turn from an accelerated charged particle follows the relativistic limit of Larmor's formula, $\frac{\delta E}{\text{turn}} \propto \left(\frac{E}{m}\right)^4$ [69], the energy loss for protons is significantly smaller than for electrons. Furthermore, the high beam intensity required to have a large rate of collisions at the interaction point excluded the possibility of building a proton-antiproton collider, due to the difficulties in maintaining and operating a high intensity anti-proton beam. As a consequence of choosing protons as colliding particles, the LHC is able to probe multiple energy ranges thanks to the broad momentum distribution of the partons inside the protons.

The LHC is designed to reach a centre-of-mass energy of $\sqrt{s} = 14$ TeV (5.4 TeV per nucleon pair) and an *instantaneous luminosity* of $\mathcal{L} = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ ($\mathcal{L} = 10^{27} \text{ cm}^{-2}\text{s}^{-1}$) for proton-proton (heavy ions) collisions. The instantaneous luminosity parametrises the rate of collisions at the interaction point and directly relates the event rate of a physics process to its cross section:

$$\frac{dN_{pp \rightarrow X}}{dt} = \mathcal{L} \times \sigma_{pp \rightarrow X}. \quad (2.1)$$

It depends on several parameters of the machine; for two Gaussian distributed proton bunches, it can be defined as:

$$\mathcal{L} = f n_b F \gamma \frac{N_1 N_2}{4\pi \sigma_x \sigma_y \epsilon \beta^*}, \quad (2.2)$$

where f is the revolutionary frequency of the bunches, n_b is the number of circulating bunches, N_1 (N_2) is the number of protons in each bunch of the first (second) beam, σ_x and σ_y are the geometrical dimensions of the beams, γ is a relativistic term and F accounts

for the geometrical correction due to the crossing angle used to make the beams collide at the interaction point. Finally, the parameters ε and β^* corresponds to the *emittance* and the *amplitude function* at the interaction point. The emittance ε is defined as the smallest width the beam can be squeezed to; in a colliding beam accelerator such as the LHC, the lower the emittance, the higher the probability for the particles to interact, resulting in a higher luminosity. The amplitude function β is defined as the distance from the interaction point at which the beam width is twice as wide as at the interaction point. Therefore, one would like the amplitude function at the interaction point β^* to be as small as possible.

To reach the nominal beam energy, bunches of protons are gradually accelerated in several steps before entering the LHC ring. Figure 2.1 shows the complex of accelerators used to inject protons into the LHC. Protons are produced via ionization of hydrogen gas, and are accelerated up to 50 MeV by the LINAC 2. The BOOSTER accelerates them up to 1.4 GeV and then injects them into the Proton Synchrotron (PS) which brings their energy to 26 GeV. The last acceleration step before entering the LHC ring is given by the Super Proton Synchrotron (SPS); here protons are accelerated up to 450 GeV and the bunches are given the design shape and geometrical dimensions utilised by the LHC collider.

The LHC tunnel has eight straight sections and eight arcs arranged on a plane inclined at 1.4% towards the city of Geneva. Given the internal diameter of the tunnel of only 3.7 m, there is not enough room for two separate rings of magnets, therefore, two sets of coils and two identical beam pipes are placed within the same mechanical structure and cryostat. The magnets of the LHC are characterized by superconducting technology. There are 1232 dipoles of 14.3 m in length which bend the beams, and 858 quadrupole magnets, each 5–7 meters long, which focus the beams. The coils' niobium-titanium (NbTi) wires are maintained at 1.9 K by a closed liquid-helium circuit and produce a magnetic field of 8.33 T. To accelerate the beams eight radiofrequency (RF) cavities per beam are housed in cylindrical refrigerators, called cryomodules, which keep the RFs working in a superconducting state, without losing energy to electrical resistance. Each cavity works at 400 MHz and is able to increase the proton energy by 485 keV each turn. A summary of the key LHC design parameters can be found in Table 2.1.

In 2010 the LHC was operating at $\sqrt{s} = 7$ TeV marking the beginning of the Run 1. In 2012, the energy of the beams was increased to 4 TeV, reaching a centre-of-mass energy of 8 TeV. After a technical stop of two years, from 2013 to 2015, the LHC started the Run 2 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV and was able to reach twice the design luminosity thanks to the very small transverse beam emittances and β^* of 30 cm [71].

In the LHC ring the beams collide in four interaction points, where four detectors with different purposes are located:

- **ATLAS (A Toroidal LHC ApparatuS)** [72] and **CMS (Compact Muon Solenoid)** [73] are general-purpose detectors. They investigate a wide range of physics, from the search for the Higgs boson (whose discovery was announced in July 2012 by both

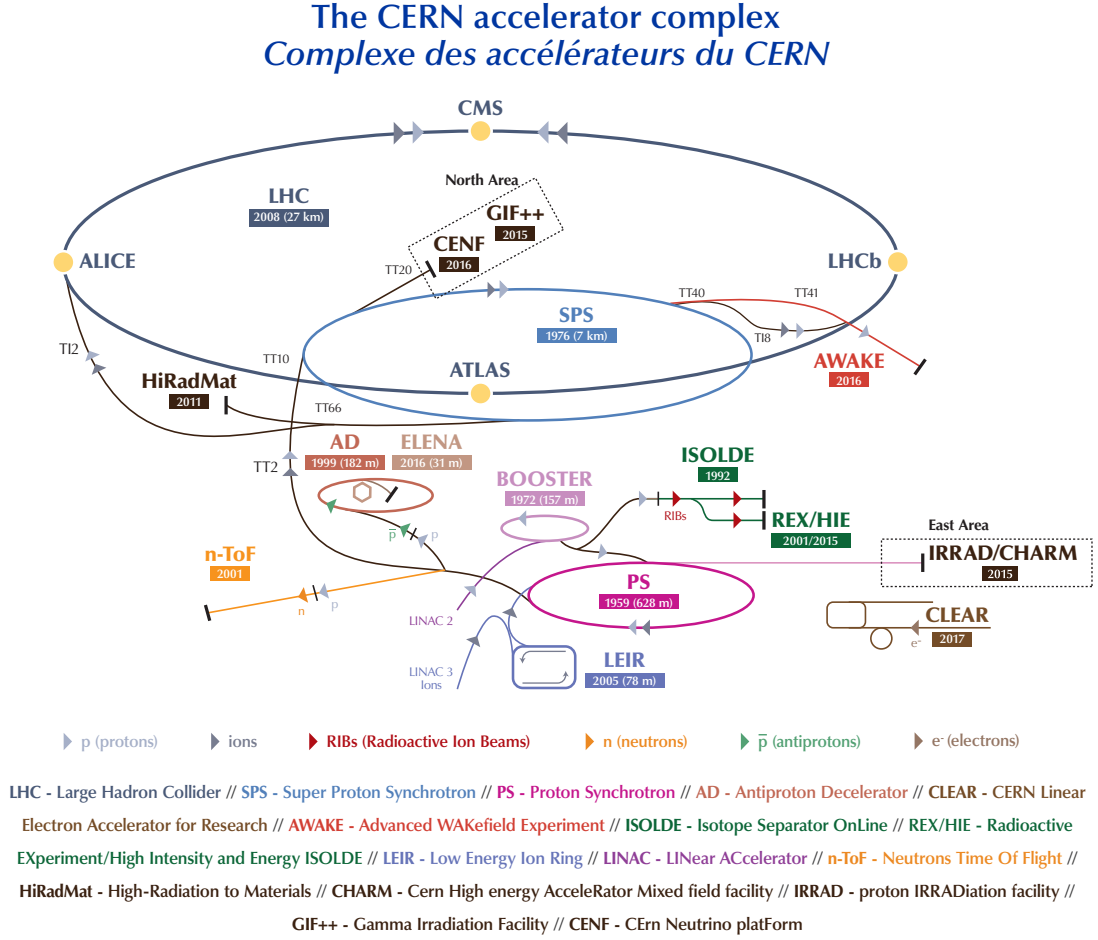


Figure 2.1: LHC injection chain: the LINAC 2 accelerates the protons up to 50 MeV and injects them into the BOOSTER that increase their energy up to 1.4 GeV. The protons are then injected to the PS which accelerates them up to 26 GeV before injecting them into the SPS in which they reach an energy of 450 GeV. Finally, the LHC accelerates each beam of protons at 6.5 TeV. The injection chain of the heavy ions and the Large Electron Positron collider (LEP) chain for the acceleration of positrons and electrons are also shown [70].

Table 2.1: Summary of the key LHC design parameters.

Circumference	27 km	Bunch spacing	7.48 m
Max beam energy	7 TeV	Bunch time separation	25 ns
Revolution frequency	11.1 kHz	Bunch length	7.55 cm
Peak luminosity	$10^{34} \text{cm}^{-2} \text{s}^{-1}$	Bunch transverse size	$16 \mu\text{m}$
RF frequency	400 MHz	Crossing angle	$285 \mu\text{rad}$
Dipole magnetic field	8.33 T	Beam lifetime	13.9 h
Magnet temperature	1.9 K	Beam energy loss per turn	7 keV
Number of bunches	2808	Horizontal beam size	$18 \mu\text{m}$
Number of magnet dipoles	1232	Vertical beam size	$71 \mu\text{m}$
Protons per bunch	$1.15 \cdot 10^{11}$	Amplitude function β^*	80 cm

experiments) to precision measurements of SM parameters, B-physics, search for Supersymmetry (SUSY) and for dark matter particles. Although they have the same scientific goals, they use different technical solutions and a different magnet-system design.

- **ALICE (A Large Ion Collider Experiment)** [74] is a heavy-ion detector. It is designed to study lead-lead collisions but data-taking with lighter ions, pp and proton-nucleus collisions is equally part of the physics programme. The aim is to study a particular state of matter, called quark-gluon plasma, which is believed to represent the conditions of the Universe a fraction of second after the Big Bang.
- **LHCb (Large Hadron Collider beauty)** [75] is a forward detector developed only on one side of the interaction point. It is focused in performing precise measurements of rare decays of the b - and c -hadrons, as well as investigating the asymmetry between matter and antimatter in the Universe.

Given that the inelastic cross-section for a pp interaction is $\sigma \sim 80$ mb at 13 TeV and each bunch has $\sim 10^{11}$ protons, there is a high probability of having more than one proton-pair collision during the same bunch crossing; such effect is called *pile-up*. It is characterised by the average number of interactions per bunch crossing defined as:

$$\mu = \frac{\mathcal{L}\sigma}{n_b f}, \quad (2.3)$$

with $\mathcal{L}$ the instantaneous luminosity, σ the pp inelastic cross section, n_b the number of bunches, and f the revolutionary frequency of the bunches in the LHC ring. High values of $\langle \mu \rangle$ imply high occupancy in the detectors, distortion of the tracks during the reconstruction, or mis-measurement of the energy in the calorimeters. Therefore, a strong effort has been made to develop high granularity detectors (especially in the region close to the interaction point) capable of disentangling multiple pp collisions.

As shown in Figure 2.2 for the ATLAS detector, each data-taking period has its own pile-up profile for an average value of 33.7 interactions per bunch crossing for the whole Run 2 data taking period.

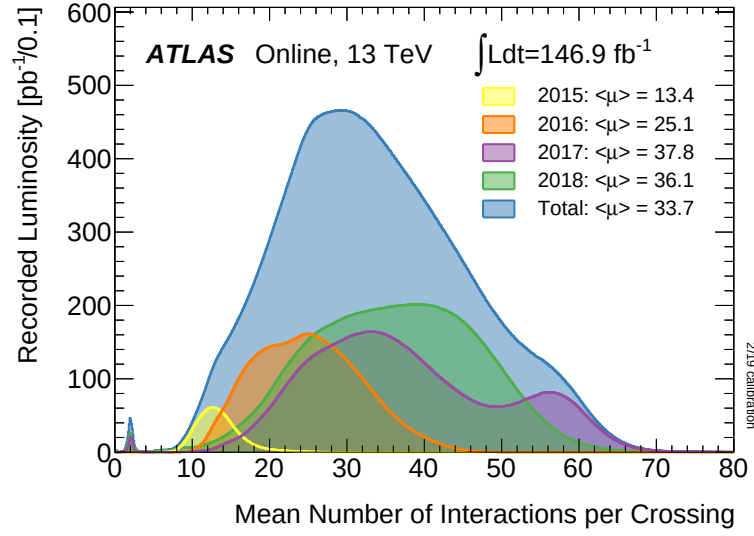


Figure 2.2: Recorded luminosity-weighted distribution of the mean number of interactions per bunch crossing for the 2015-2018 datasets recorded by the ATLAS detector at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The mean number of interactions per crossing corresponds to the mean of the Poisson distribution of the number of interactions per crossing calculated for each bunch as in Equation 2.3 [76].

2.2 The ATLAS detector

The ATLAS detector is the largest experiment along the LHC ring, with a diameter of 25 m, a length of 44 m, and a weight of 7000 tons. It has a forward-backward symmetry around the interaction point and it is composed of an inner detector (ID), an electromagnetic and a hadronic calorimeter, and a muon spectrometer (MS). A schematic view of the ATLAS detector is shown in Figure 2.3.

The nominal interaction point defines the origin of the coordinate system; the beam direction defines the z -axis, the positive x -axis goes from the interaction point to the centre of the LHC, and the positive y -axis points towards the surface. This way, the $x - y$ plane is transverse to the beam direction. On this plane the coordinates (r, ϕ) are used, where ϕ is the azimuthal angle around the beam axis and r is the radial distance from the interaction point outwards.

The polar angle θ that the particle forms with the beam axis is usually replaced with the *pseudorapidity*:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \quad (2.4)$$

In the limit of ultrarelativistic particles, this is an approximation of the *rapidity*:

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

where p_z is the projection of the particle momentum along the z -axis. Differences in rapidity are invariant under Lorentz transformations and the density of emitted particles

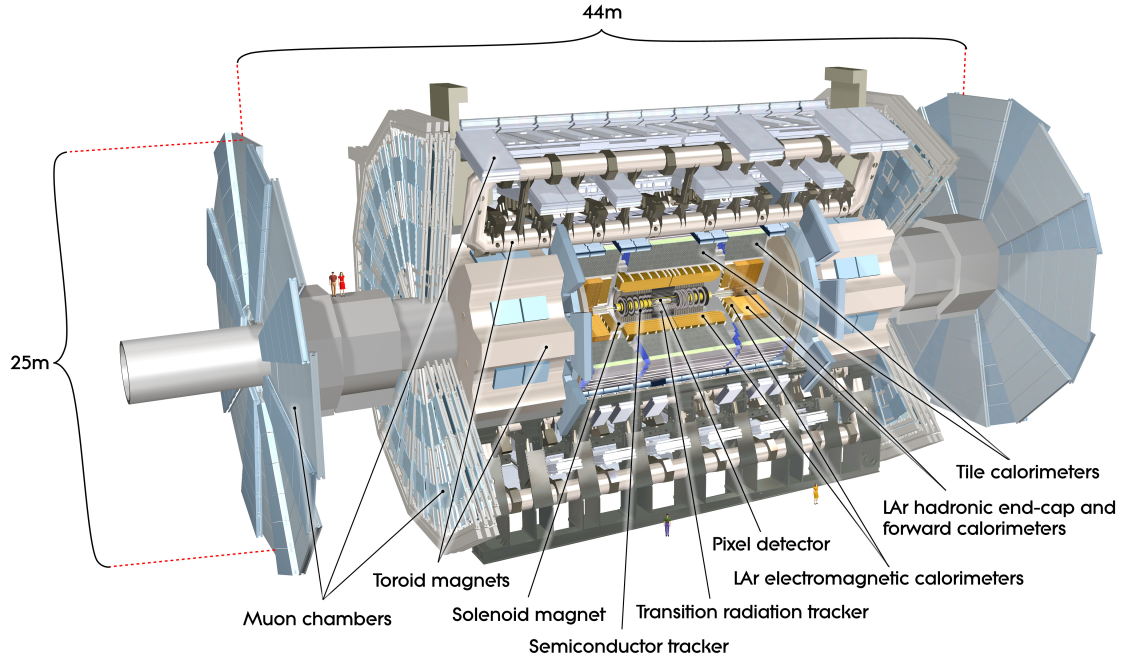


Figure 2.3: Schematic view of the ATLAS experiment with its main components [72].

$\frac{dN}{dy}$ as a function of the rapidity is almost constant.

Given that the initial longitudinal momentum of the two colliding partons is not known, some quantities are expressed in the transverse plane, such as the transverse momentum p_T and the transverse energy E_T , defined as:

$$p_T = p \sin \theta \quad \text{and} \quad E_T = E \sin \theta. \quad (2.6)$$

Finally, since longitudinal boosts do not affect the azimuthal angle ϕ , a useful quantity to measure is the angular distance in the transverse plane:

$$\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}. \quad (2.7)$$

2.2.1 Magnet system

The measurement of charged particles' momenta is based on their deflection within a magnetic field. In the ATLAS detector this is produced by four large superconductive magnets [77] whose spatial arrangement is shown in Figure 2.4.

A central *solenoid* is placed inside the calorimeter and produces a 2 T axial magnetic field for the inner detector, bending the particles in the transverse plane. It has an inner and outer diameters of 2.46 m and 2.56 m, respectively, and a length of 5.8 m. In order to achieve the required performances for the calorimeters, in terms of electron and photon identification and accurate jet and missing E_T measurement, the material thickness of the solenoid has been maintained as small as possible. The magnetic field flux is returned

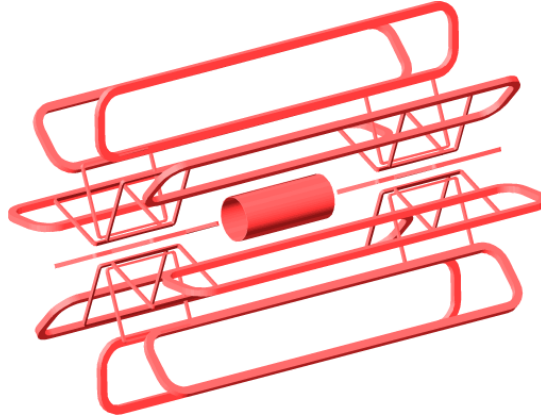


Figure 2.4: Spatial arrangement of the magnetic system with the eight barrel and end-cap toroid coils, and the central solenoid [78].

by the steel of the hadronic calorimeter.

The *toroid system* produces the magnetic field for the muon spectrometer bending the particles in the $z-y$ plane and covering the region for $|\eta| < 2.7$. It is composed of a barrel and two end-cap toroids, centrally located around the calorimeters, which produce a magnetic field of approximately 0.5 T and 1 T, respectively, and are maintained to a temperature of 4.5 K by a helium-cooling system in order to reach the superconductivity state. The end-cap toroids are rotated by 22.5° with respect to the barrel in order to increase the overlap of the two magnetic fields and achieve a higher uniformity. Each toroid is composed of eight coils assembled symmetrically around the beam axis. The choice of the “open-air” toroid configuration is one of the strengths of the ATLAS detector, as it improves the muon reconstruction performance without relying on the inner detector. In fact, the toroids generate the magnetic field in a large volume with a small amount of material. This minimises the amount of multiple scattering¹, which represents one of the limiting factors in the momentum resolution measurement.

2.2.2 Inner detector and tracking

Every proton-proton collision produces approximately 1000 tracks every 25 ns within $|\eta| < 2.5$. To achieve high charged particle momentum resolution and reconstruction efficiency, the Inner Detector [79] must have fine granularity and a limited amount of material. In fact, large amounts of material can create an issue of multiple scattering which limits the momentum resolution of the ID as well as degrading the resolution of the calorimeters.

The ID is immersed in a 2 T magnetic field generated by the central solenoid and contained in a cylindrical envelope of length 3.5 m and diameter 1.15 m. Three different technologies are utilised, with decreasing granularity at increasing distances from the centre of the detector. The innermost component is the Pixel Detector, followed by the

¹It is defined as the electromagnetic interaction of a particle with a large number of atomic nuclei in the material which results in a random smearing of the original momentum of the particle.

SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The first two cover the region $|\eta| < 2.5$, while the TRT allows for track reconstruction up to $|\eta| = 2.0$. The layout of the ID is illustrated in Figure 2.5.

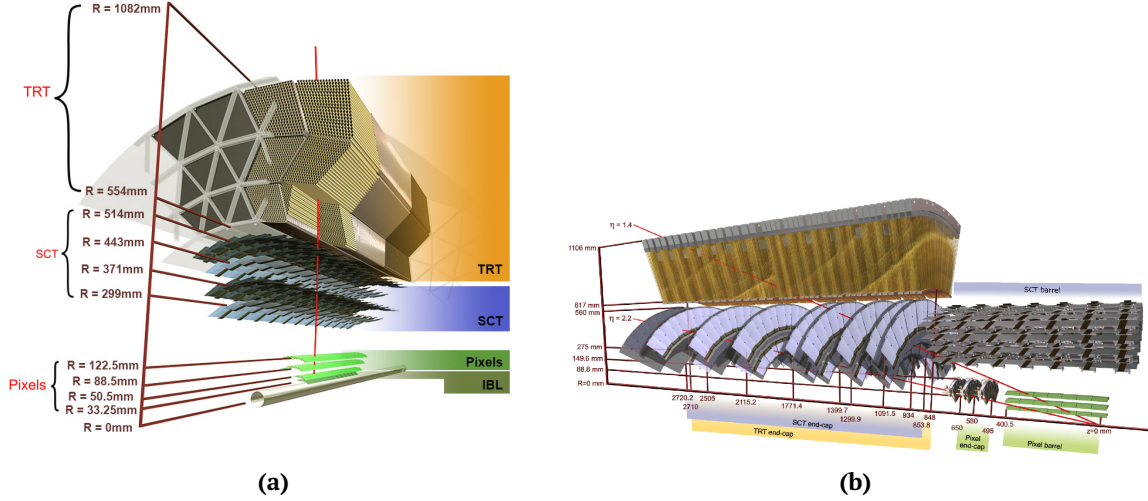


Figure 2.5: Schematic view of the front (a) and side (b) layout of the Inner Detector showing the three main components: the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker [72, 80].

The three sub-detectors combined provide full tracking reconstruction up to $|\eta| = 2.0$ with a designed momentum resolution of [72]:

$$\sigma_{p_T} = 0.05\% \cdot p_T \text{ (GeV)} \oplus 1\%. \quad (2.8)$$

Pixel detector The pixel detector is the innermost element of the ID and provides a robust pattern recognition for charged particle tracks having $p_T > 0.5$ GeV and $|\eta| < 2.5$. It is the most important detector for the reconstruction of the primary vertex, even in presence of pile-up, and for the reconstruction and identification of secondary vertices from the decay of b - or c -hadrons. Furthermore, it is crucial in the measurement of the track impact parameter d_0 , defined as the minimum distance between the track and the primary vertex. These are the main ingredients for the identification of jets originating from b -, c -quarks, and τ -leptons (see Chapter 4).

In the barrel, four pixel layers of 80 cm length, conventionally called Insertable B-Layer (IBL), B-layer (BL), Layer 1 and Layer 2, are inserted at radii of 33.25 mm, 50.5 mm, 88.5 mm and 122.5 mm, respectively. The IBL [81] was installed in 2014 to cope with the high luminosity expected to be delivered by the LHC during Run 2. In fact, the increase in the event pile-up and in the local occupancy rate was expected to deteriorate the tracking resolution, directly affecting the b -tagging performances. The IBL not only improves the vertexing and tagging precision, but is able to restore the full b -tagging efficiency even with a complete failure of the B-layer.

Each end-cap contains three disks placed at distances of $|z| = 49.5$ cm, 58.0 cm and 65.0 cm from the centre of the detector, with a radius of 34 cm.

All layers are made with silicon sensors segmented in $\phi \times z$ to allow for the simultaneous measurement of both coordinates. In all layers, the modules are overlapped on the support structure to ensure hermetic coverage (see Figure 2.6 (a) for the layout of the IBL, as an example). The IBL sensors utilise two technologies, planar sensors [81] and 3D sensors [82], both with a size of $50 \times 200 \mu\text{m}^2$ and a thickness of $200 \mu\text{m}$ and $230 \mu\text{m}$. The sensors mounted on the outer layers have a size of $50 \times 400 \mu\text{m}^2$ and thickness of $250 \mu\text{m}$. The IBL reaches a hit spacial resolution of $10.0 \mu\text{m}$ and $66.6 \mu\text{m}$ in the ϕ and z directions, respectively. This has been measured in the region of overlap between adjacent IBL modules (see Figure 2.6 (b)) in order to minimize the dependence from the track extrapolation [83]. The accuracy reached by the outer layers in the barrel (end-caps) is $10 \mu\text{m}$ in ϕ and $115 \mu\text{m}$ in z (r).

The pixel detector provides at least four hits to the track reconstruction. A single hit is reconstructed from clusters of pixel cells having a charge deposit above a given discriminator threshold.

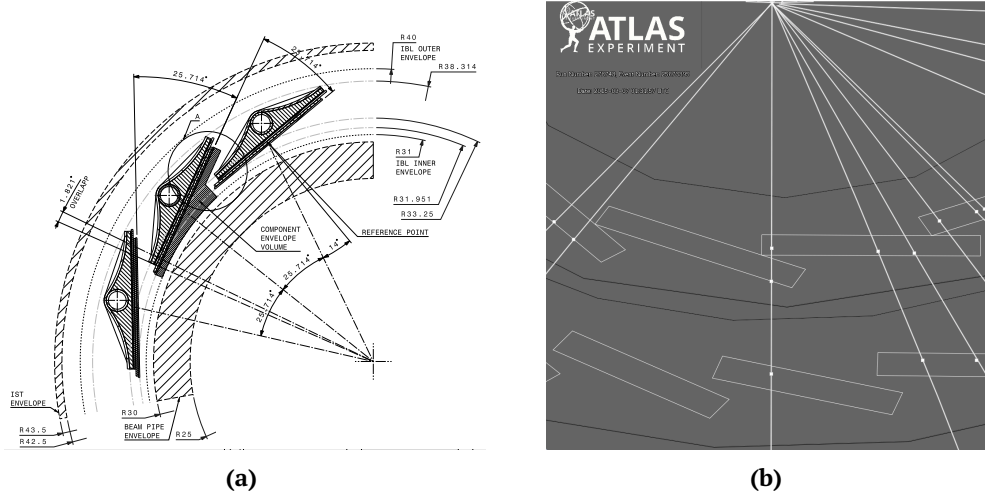


Figure 2.6: The pixel modules mounted on each layer have a region of overlap to ensure a more hermetic coverage. Figure (a) shows the layout of the IBL in the $r-\phi$ plane as an example. In this case, the pixel modules overlap in a slice of 1.821° , therefore, if a particle crosses the material in this region, it can leave not one but two hits, as shown in (b) [83].

SemiConductor Tracker The SCT system [84] is designed to provide at least four hits in the track reconstruction. Four layers of silicon microstrip detectors are placed in concentric cylinders with radii of 300 mm, 373 mm, 447 mm and 520 mm around the beam axis. The end-cap modules are arranged onto nine wheels from 83.9 cm to 273.5 cm along $|z|$, each of which has up to three rings of modules.

Each module in the barrel has a size of $6.36 \times 6.40 \text{ cm}^2$, a thickness of $300 \mu\text{m}$ and a pitch of $80 \mu\text{m}$. In the end-caps, the exact shape and size of the modules depend on the radial position they occupy in ATLAS; those closer to the beam pipe are smaller than the modules mounted in the outermost radial position. Both barrel and end-cap modules consist of two single-sided silicon microstrip sensors mounted back-to-back, and the

coordinate perpendicular to the strip direction is measured exploiting the stereoscopic effect: the sensors are rotated by an angle of 40 mrad and the crossing point of the strips of the two sensors is used to determine the space position. The spatial resolutions reached by the modules are $17\ \mu\text{m}$ in $r - \phi$ and $580\ \mu\text{m}$ in z .

Transition Radiation Tracker The TRT system [85] is a straw-tube tracker which provides ~ 36 hits per track and useful information for the electron identification, due to the measurement of the transition radiation.

The barrel contains 50000 straw tubes and covers the radial range from 56 cm to 107 cm. It is composed of individual modules with between 329 and 793 axial straws 150 cm long, covering the pseudorapidity range of $|\eta| < 1$.

The two end-caps consist of 18 wheels each, with a total of 32000 straws arranged in the radial direction and covering the radial range from 54 to 103 cm.

Each straw tube has a diameter of 4 mm and was filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2 during the Run 1 of the LHC. In Run 2, all straws belonging to modules with large gas leaks were filled with a gas mixture of 70% Ar, 27% CO_2 and 3% O_2 . Argon has a much lower efficiency to absorb the transition radiation photons in this energy range, but has similar tracking capabilities as Xe [86].

A charged particle traversing the TRT ionises the gas inside the straws and the resulting free electrons migrate towards a gold-tungsten wire of $31\ \mu\text{m}$ in diameter placed at the centre of the tube at a $-1.5\ \text{kV}$ voltage.

The space between the tubes is filled with plastic material (polyethylene) in order to produce transition radiation at the material boundary. This effect is particularly useful for the identification of electrons as it depends on the Lorentz boost $\gamma = E/m$, which is stronger for electrons. The electrons and photon conversion reconstruction is particularly challenging for the inner detector due to the 20%-50% of energy lost by the electrons in the pixel and SCT due to bremsstrahlung and ionization. The TRT provides the identification of electrons with energies below 25 GeV, complementing the measurements from the calorimeter, and contributes to the reconstruction of electron track segments from photon conversions down to 1 GeV.

The signal induced by a particle passing through the TRT straws is collected and compared to two separate thresholds: the higher is used to identify the photons from the transition radiation of the electrons, the lower is used to distinguish other kinds of particles, such as muons, pions or kaons.

2.2.3 Electromagnetic Calorimeter

For high energy electrons and photons, the combination of the bremsstrahlung and the pair production phenomena creates an electromagnetic cascade which spreads out laterally due to low energy γ produced in Compton's scattering and photoelectric effect,

and to the Coulomb scattering². The longitudinal spread is determined by the *radiation length* X_0 in the medium, defined as the mean length (in cm) needed to reduce the energy of an electron of $1/e$. The angular deflection can be quantified by the *Molière radius*, defined as the radius which contains 90% of the shower, $R_m = 21(X_0/E_c)$ with E_c the critical energy defined as the energy at which the losses from ionization and from bremsstrahlung are equal. For this reason, electromagnetic-shower detectors are built from high- Z materials of small X_0 in order to contain the shower in a small volume.

The ATLAS Electromagnetic CALorimeter (ECAL) [87] is designed to trigger on and provide precise measurement of electrons, photons, jets and missing transverse energy. It uses liquid argon as active material and 1.8 mm lead plates as absorbers. The liquid argon has been chosen for its intrinsic linear behaviour, its response stability and radiation resistance. The calorimeter is composed of a barrel and two end-caps components, each housed in their own cryostat. The barrel consists of two half-barrels separated by a 4 mm gap at $z = 0$ and covering the range $|\eta| < 1.475$. It is made of 16 separate modules, each covering $\Delta\phi = 22.5^\circ$ and with a total thickness of $22 X_0$, increasing from 22 to $30 X_0$ between $|\eta| = 0$ and $|\eta| = 0.8$ and from 24 to $33 X_0$ between $|\eta| = 0.8$ and $|\eta| = 1.3$. A *presampler* layer of 11 mm is placed in front of the first layer of the calorimeter to recover the energy lost by electrons and photons in the passive material of the solenoid.

The two end-caps are divided into two coaxial wheels: an outer wheel covering the range $1.375 < |\eta| < 2.5$, and an inner wheel covering the region $2.5 < |\eta| < 3.2$. The total active thickness goes from 24 to $38 X_0$ for $1.475 < |\eta| < 2.5$, from 26 to $36 X_0$ for $2.5 < |\eta| < 3.2$, and smaller than $24 X_0$ for $|\eta| < 1.475$. In order to improve the energy resolution, a 5 mm liquid argon presampler is placed in front of both end-caps for $1.5 < |\eta| < 1.8$.

The lead absorbers and the electrodes of both the barrel and the two end-caps have an accordion geometry which provides full coverage in ϕ without any gap. In both cases, the folding angle of the waves varies with the distance from the centre of the detector to maintain the LAr gap constant.

Each module of both the barrel and end-cap calorimeters is composed of three layers with decreasing granularity to follow the evolution of the electromagnetic shower: the first layer is finely segmented in η to achieve an accurate measurement of the shower starting point, the second layer collects the largest fraction of the shower energy, and the third layer is mainly used to estimate the amount of energy leaking beyond the ECAL. The layout of the calorimeter system can be found in Figure 2.7 together with a detailed view of a barrel module.

The resolution of the electromagnetic calorimeter can be parametrized with the following expression:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E(\text{GeV})}} \oplus \frac{b}{E(\text{GeV})} \oplus c, \quad (2.9)$$

²A particle traversing a material undergoes elastic interactions with the nuclei that deflect the particle direction by an angle θ . The mean value of θ follows a Gaussian distribution with mean value of zero.

where a takes into account the stochastic behaviour of the shower development, b accounts for the electronic and pile-up noise, and c is a constant term that quantifies the non-uniform response of the calorimeter, the ageing and the radiation damage. The design value for these terms are $a = 10\%$, $b = 170$ MeV (without pile-up), and $c = 0.7\%$ [89].

2.2.4 Hadronic Calorimeter

When a hadron traverses the material, it can undergo an inelastic collision with the nuclei producing secondary hadrons which, in turn, interact inelastically to produce a third set of hadrons, etc. The scale for the hadron shower is the *nuclear absorption length* λ , which is bigger than X_0 resulting in a larger longitudinal and radial development with respect to the electromagnetic shower.

The ATLAS Hadronic CALorimeter (HCAL) is made of three independent sampling calorimeters at different values of η , each using its own technology and material to face the different radiation fluxes, and satisfy the required performances.

The Tile Calorimeter (TileCal) [90] provides the energy reconstruction of the jets and, together with the ECAL, a good measurement of the missing E_T . It is located in the region of $|\eta| < 1.7$, after the ECAL, and it is composed of a central barrel of 5.64 m in length covering the region $|\eta| < 1$ and two extended barrels of 2.91 m covering the range $0.8 < |\eta| < 1.7$. In the central barrel there are three radial layers approximately 1.5, 4.1 and 1.8 λ thick, while in the extended barrel the layers have a thickness of 1.5, 2.6, and 3.3 λ . An Intermediate Tile Calorimeter (ITC) has been installed to improve the performance in the region between the central and the extended barrels, which is occupied by cables and electronics from the inner tracker and the ECAL.

Both the central and the extended barrels are made of 64 independent towers in the z and ϕ directions, each made of scintillating tiles as active material and laminated steel plates as absorber. The scintillating tiles are placed perpendicular to the colliding beams and staggered along r . The signal produced in the scintillating tiles is read-out by wavelength shifting fibres, as shown in Figure 2.7 (c).

The Hadronic End-cap Calorimeter (HEC) [91] is a liquid argon calorimeter with copper as absorber, integrated in the end-cap cryostat and covering the region $1.5 < |\eta| < 3.2$. It consists of two independent wheels per end-cap, each with an outer diameter of 4 m and a length of 0.82 m and 0.96 m. Each wheel is built from 32 identical wedge-shaped modules with 25 mm parallel copper plates for those closer to the interaction point, and 50 mm copper plates for those further away.

The liquid argon hadronic Forward Calorimeter (FCal) [92] is also integrated into the end-cap cryostat and covers the region $3.2 < |\eta| < 4.9$ where the energy and the density of the particles are very high. It provides good measurement of the forward jet, improves the reconstruction of the missing E_T , and helps in reducing the background level in the muon spectrometer. It consists of three modules in each end-cap, the first is optimized for electromagnetic measurements and it is made of copper, while the other two measure the hadronic shower and are made of tungsten.

The design resolutions for the ATLAS hadronic calorimeters [72] are:

$$\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E(\text{GeV})}} \oplus 3\% \quad \text{for TileCal and HEC,} \quad (2.10)$$

$$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E(\text{GeV})}} \oplus 10\% \quad \text{for FCal.} \quad (2.11)$$

2.2.5 Muon Spectrometer

The Muon Spectrometer (MS) [93] is designed to provide standalone precise measurements of muons up to 3 TeV and in the pseudorapidity range $|\eta| < 2.7$. The capability of reconstructing high- p_T muons without relying on the inner detector is crucial for performing fast muon triggering in the region $|\eta| < 2.4$. The design momentum resolution is 4% for muons with p_T up to 100 GeV and 10% at p_T of 1 TeV. At small energy, the performances are recovered by combining tracks reconstructed by the muon spectrometer with those in the inner detector.

The layout of the spectrometer is based on the magnetic deflection of muon tracks inside the detector. For $|\eta| < 1.4$, the deflection is provided by the 0.5 T magnetic field of the barrel toroid, for $1.6 < |\eta| < 2.7$, muon tracks are bent by two end-cap toroids. Over $1.4 < |\eta| < 1.6$ the deflection is provided by a combination of barrel and end-cap magnetic fields. Figure 2.8 shows the transversal view of the ATLAS muon spectrometer.

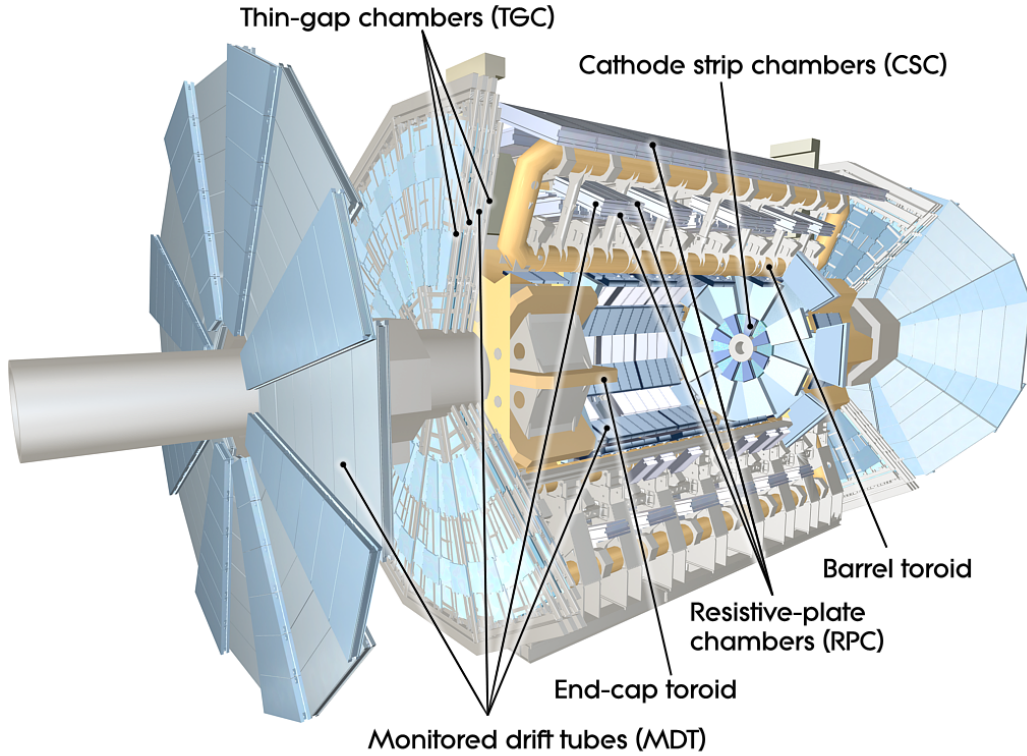


Figure 2.8: Cross-section of the muon system [72].

In the barrel region, tracks are measured in chambers arranged into three cylindrical layers at radii of approximately 5 m, 7.5 m, and 10 m. In the end-caps, the chambers are installed perpendicular to the beam line at distances of $|z| \approx 7.4$ m, 10.8 m, 14 m, and 21.5 m. Precise measurement of the position and momentum of the muons is provided by the *tracking chambers*, namely the Monitored Drift Tubes (MDTs) chambers and the Cathode Strip Chambers (CSCs). The trigger system covers the range $|\eta| \leq 2.4$ and is provided by the *triggering chambers* divided in Resistive Plate Chambers (RPCs) in the barrel and Thin Gap Chambers (TGCs) in the end-caps.

The MDTs chambers' size increase with their distance from the interaction point. They are made of three to eight layers of drift tubes filled with a mixture of Ar-CO₂ (93%-7%) and operating at a pressure of 3 bar. A muon traversing the detector ionizes the atoms of the material and the resulting free electrons are collected at a 50 μm tungsten-rhenium wire located at the centre of the tube. The average resolution for a single tube is 80 μm and for a chamber is 35 μm .

The CSCs are multiwire proportional chambers with wires oriented in the radial direction and cathods segmented in strips with high granularity to sustain the high background rate and the radiation ageing, typical of the forward regions. They cover the region $2 < |\eta| < 2.7$, and are arranged in two disks of different sizes having eight chambers each. Each chamber contains four independent CSC planes providing four different measurements in η and ϕ . The resolution of a chamber reaches 40 μm in the bending plane and ~ 5 mm in the transverse plane.

The capability of the Muon Spectrometer to trigger on high- p_T muons is granted by the RPCs in the region $|\eta| < 1.05$ and by the TGCs in the $1.05 < |\eta| < 2.4$.

The RPCs are made of parallel resistive plates and a gas-filled gap of 2 μm . A strong electric field allows the production of $e^\pm$ /ions avalanches along the ionizing track towards the anode, ensuring a rate capability of ~ 1 kHz cm^{-2} . The resolution per chamber is 10 mm in both the z and ϕ directions.

The TGCs are multi-wires chambers whose main difference with the RPCs' technology is that the anode wire pitch is larger than the cathode-anode distance, which ensures a fast collecting time. The resolution per chamber is 2-6 mm in z and 3-7 mm in ϕ .

Both RPCs and TGCs deliver the signals with a spread of 15-25 ns providing the ability to tag the beam crossing with an efficiency $> 99\%$.

2.2.6 Luminosity Detectors

The instantaneous luminosity and the beam conditions are constantly monitored by two forward detectors, LUCID and *Beam Monitor Conditions* (BCM). LUCID [72] is the only detector dedicated to the online monitoring of the luminosity. Two identical modules are installed, one in each end-cap region of ATLAS, at a distance of ~ 17 m from the interaction point. Each module consists of 20 aluminium tubes filled with C₄F₁₀ and coupled to photomultipliers.

The BCM [94] consists of two stations with four 1 cm² diamond sensors placed around

the beam pipe at a distance of $|z| = \pm 1.84$ m from the interaction point. Its fast readout and the good time resolution of 0.7 ps provide measurement of the bunch-by-bunch luminosity as well as triggering on beam losses to prevent damages to the silicon pixel sensors of the ID, dumping the beam if necessary.

The total integrated luminosity delivered by the LHC and recorded by the ATLAS detector for the full Run 2 data-taking is shown in Figure 2.9. The small differences between the luminosity delivered by the LHC and the one recorded by ATLAS are due to the time needed for ramping-up the high-voltages of the tracking detectors and the preamplifiers of the pixel detector after the beams are declared to be stable. In addition, the data retained for physics analyses is a subset of that recorded due to the requirements of “good-quality” data, meaning that all sub-detectors were functioning within a nominal operation threshold for voltages, temperature, humidity etc., and that the reconstruction algorithms of physics objects were performing as expected.

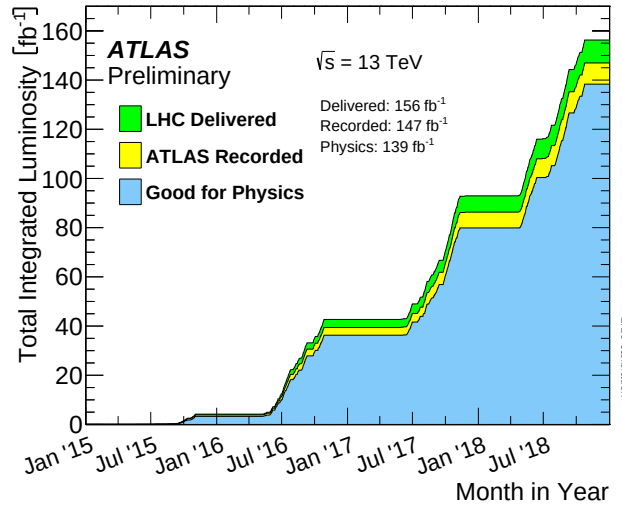


Figure 2.9: Cumulative luminosity versus time delivered to ATLAS (green) and recorded by ATLAS (yellow) and of good quality for physics analysis (blue) during stable beams for pp collisions at $\sqrt{s} = 13$ TeV in 2015–2018. The complete pp data sample in 2018 is shown and the Good Data Quality criteria require all physics objects to be of good data quality [76].

2.2.7 Trigger System

The ATLAS detector is designed to observe up to 1.7 billions pp collisions per second for a combined data volume of 60 millions MB per second. However, only some of these events contains interesting characteristics for physics analyses. These events are selected by a complex trigger system which reduces the data flow from 40 MHz to ~ 1 kHz for permanent storage.

The trigger system was originally designed to have three levels of event selection: the Level-1 (LVL1) hardware trigger, the software-based Level-2 (LVL2) and the event filter.

The LVL1 trigger [95] uses reduced-granularity information from the RPCs and the TGCs searching for patterns of hits consistent with high- p_T muons, and from the calorimeters

for the identification of electromagnetic clusters, jets and τ -leptons. LVL1 can accept a maximum rate of 100 kHz per bunch crossing and uses custom-built hardware processors to unambiguously identify the bunch crossing of interest making a decision in a maximum of $2.5 \mu\text{s}$. Upon the event being accepted by LVL1, the information about the geometrical location of the trigger objects is organized in *Regions-of-Interest* (RoI's) which are sent to the LVL2 trigger.

The LVL2 trigger can handle a maximum rate of 75 - 100 kHz and uses the full granularity of the sub-detectors but only in the RoI's provided by the LVL1 trigger. It defines selection criteria analysing the properties and topology of the event and making a decision in approximately 40 ms. The accepted event rate, approximately 3.5 kHz, is passed to the event filter which runs a complete and offline reconstruction of the events making a final decision in about one second.

In Run 2, the system has been modified to just two levels, LVL1 and the High-Level Trigger (HLT) unifying LVL2 and the event filter. The HLT is implemented in software and operates on large farms of commercial computer processors [96].

Object Reconstruction

Each recorded event undergoes an *offline* event reconstruction. The responses of all ATLAS sub-detectors are combined to form objects like tracks and calorimeter clusters which are then used to reconstruct final physics objects. This chapter describes the techniques used by the ATLAS Collaboration to reconstruct and identify the objects of interest for the studies presented in this thesis such as leptons, jets and missing transverse energy.

3.1 Tracks and primary vertices

Track reconstruction It is usually divided into two steps, the track finding or pattern recognition, and the fit used to extract the track parameters and their uncertainties. The clustering algorithm groups the neighbouring pixels and the strips in a given sensor if the charge deposit is higher than a given threshold. The charge in a pixel sensor is typically collected on multiple pixels, and the total charge depends on the incident angle of the particle with respect to the sensor. Clusters of pixels originating from a single particle are called *single particle* clusters, while clusters originating from charge deposits of multiple particles are called *merged* clusters. Figure 3.1 shows the different topology of the two types of clusters.

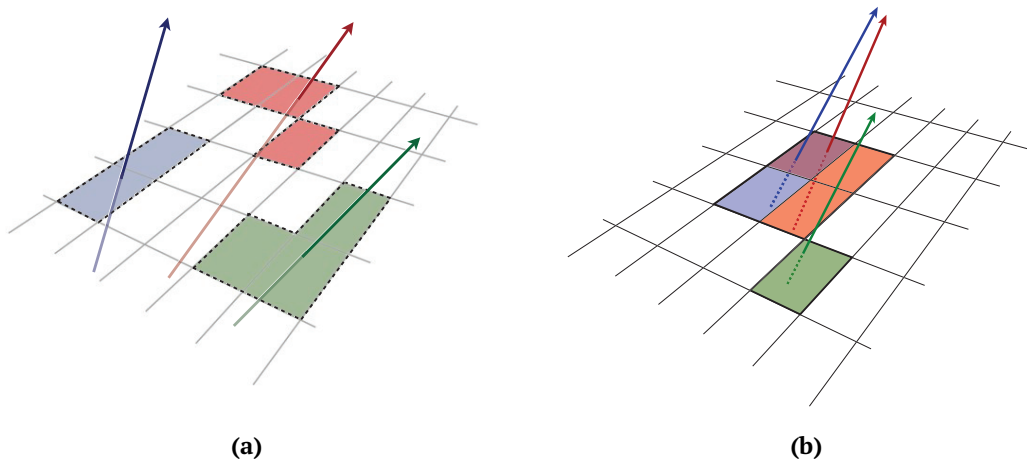


Figure 3.1: Illustration of single particle pixel clusters on a pixel sensor (a) and a merged pixel cluster (b) due to very collimated charged particles. Different colours represent energy deposits from different charged particles traversing the sensor and the particles trajectories are shown as arrows [97].

The track finding defines a *seed* of three-dimensional measurements, called *space-points*, from a combination of pixel and SCT clusters which is used to create track can-

didates. Criteria on which sub-detectors recorded the space points are used to define the purity of the track candidates; tracks using SCT-only seeds are placed first, followed by pixel-only, and by mixed-detector seeds. Purity is further enhanced by requiring an additional space-point in a subsequent layer compatible with the particle trajectory estimated from the seed¹. A combinatorial Kalman filter [98] is then used to extend the seeds by incorporating additional space-points from the remaining layers of the pixel and SCT sub-detectors.

An ambiguity solver then processes all track candidates in descending order of track score. This is based on the number of clusters assigned to the track and if it reflects the expected cluster multiplicities in the different sub-detectors. The χ^2 of the track fit penalizes the track score as well as the presence of holes². Furthermore, high energy tracks are favoured with respect to low p_T tracks which are likely to have mis-associated clusters. Track candidates are rejected by the ambiguity solver if they do not fulfill quality criteria based on their kinematic properties, the number of silicon hits/holes along the particle trajectory and the number of *shared* clusters. These are clusters used in the reconstruction of multiple tracks representing a strong indicator of incorrect assignment of a cluster to a track. An artificial neural network (NN) using information on the charge deposit, the position of the pixels in the clusters, and the incident angle of the particle, is employed to identify merged clusters against shared clusters, as they can be a source of degradation in the cluster assignment efficiency and, as a consequence, in the track reconstruction efficiency. Figure 3.2 shows the cluster assignment efficiency for a simulated sample of B^0 mesons up to 1 TeV decaying before the IBL as a function of the separation between two particles in the x and y directions at the sensor's surface for the IBL and the B-layer. The cluster assignment efficiency is defined as the fraction of clusters created by a particle that are used on the reconstructed track of that particle. It can be observed that when the particles are well separated, the efficiency reaches $\sim 98\%$, but for smaller separations, single clusters start to merge and the efficiency drops, especially for the IBL where the tracks are much closer to each other.

All tracks passing the ambiguity solver stage are passed to the combinatorial Kalman filter that extends them outwards to the TRT. A high resolution fit is then performed and a set of five parameters describing the track is extracted, together with their uncertainties:

$$\mathbf{x} = (d_0, z_0, \phi, \theta, q/p) \quad (3.1)$$

where d_0 and z_0 are the Impact Parameters (IP) representing the minimum distances to the centre of the detector in the $x-y$ plane and z direction, respectively, ϕ and θ define the azimuthal and polar angular directions, and q/p is the charge over momentum which quantifies the curvature of the track in the transverse plane³. All five parameters are

¹This is assumed to be a perfect helical trajectory in a uniform magnetic field.

²A hole is defined as a hit expected to be associated with a track but not present [99].

³The curvature of a track in the transverse plane can be derived from the formula $p_T \propto qBr$, with r the radius of the curvature, for particles in a magnetic field B perpendicular to p_T .

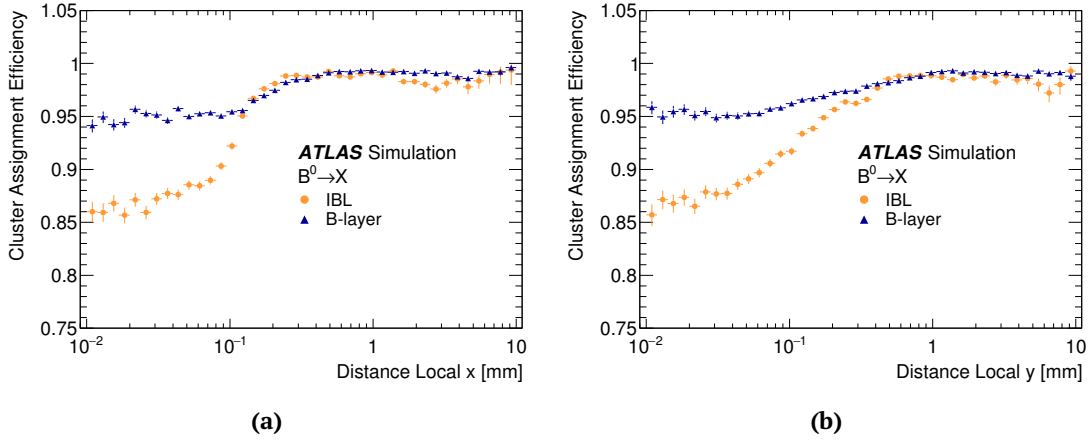


Figure 3.2: Efficiency with which reconstructed clusters are assigned to the tracks of the B^0 decay products is shown for the two innermost pixel layers (IBL and B-layer) as a function of the minimum truth-particle separation in local x (a) and y (b), corresponding to the pixel dimensions transverse and longitudinal to the beam axis. It is required that the stable charged particles are created before the IBL [97].

expressed with respect to the point of the closest approach to the primary vertex in the transverse plane. The resolution of a track parameter can be expressed as a function of the track p_T :

$$\sigma_X = \sigma_{X(\infty)} \left(1 \oplus \frac{p_X}{p_T} \right), \quad (3.2)$$

where $\sigma_{X(\infty)}$ is the asymptotic resolution expected for an infinite transverse momentum, and p_X is a constant representing the value of p_T at which the multiple-scattering contribution equals that from the detector resolution. The design resolutions for the five track parameters are reported in Table 3.1 at infinite transverse momentum $\sigma_{X(\infty)}$, and at momentum p_X .

Table 3.1: Expected track parameters resolutions (RMS) at infinite transverse momentum $\sigma_{X(\infty)}$, and at p_X representing the value of p_T at which the multiple-scattering contribution equals that from the detector resolution. The momentum and angular resolutions are shown for muons, whereas the impact-parameter resolutions are shown for pions. The values are shown for two η regions, one in the barrel ID where the amount of material is close to its minimum and one in the endcap where the amount of material is close to its maximum. These values have been estimated before the IBL was inserted in the inner detector. With the IBL, the resolution of all track parameters, but especially the one of the impact parameters, improves [72].

Track parameter	$0.25 < \eta < 0.50$		$1.50 < \eta < 1.75$	
	$\sigma_{X(\infty)}$	p_X (GeV)	$\sigma_{X(\infty)}$	p_X (GeV)
Inverse transverse momentum ($1/p_T$)	0.34 TeV^{-1}	44	0.41 TeV^{-1}	80
Azimuthal angle (ϕ)	$70 \mu\text{rad}$	39	$92 \mu\text{rad}$	49
Polar angle ($\cot \theta$)	0.7×10^{-3}	5.0	1.2×10^{-3}	10
Transverse impact parameter (d_0)	$10 \mu\text{m}$	14	$12 \mu\text{m}$	20
Longitudinal impact parameter ($z_0 \times \sin \theta$)	$91 \mu\text{m}$	2.3	$71 \mu\text{m}$	3.7

The final track reconstruction efficiency is shown for charged primary particles as a function of the jet p_T in Figure 3.3 (a) and the production radius of the charged particles in (b) for jets in the $|\eta| < 1.2$ and $|\eta| > 1.2$ ranges.

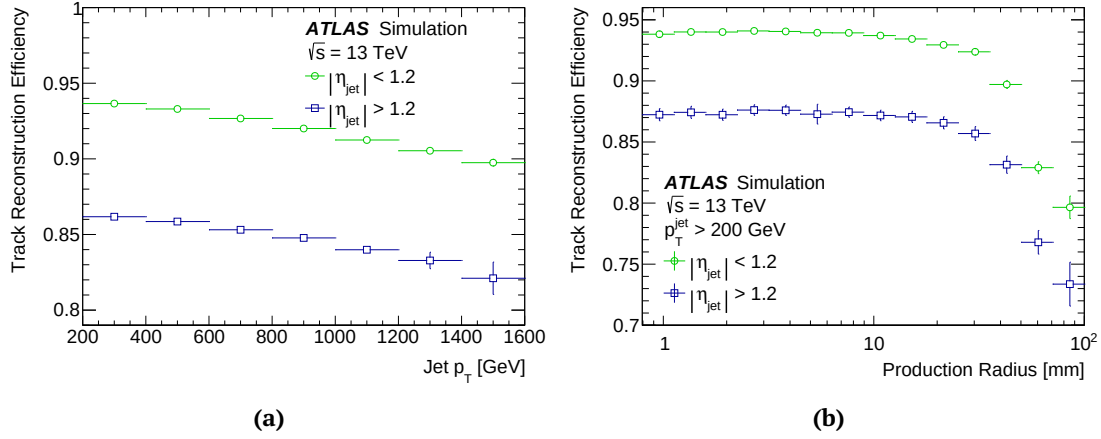


Figure 3.3: Track reconstruction efficiency for charged primary particles as a function of the jet p_T (a) and the production radius of the particle (b) for jets in the $|\eta| < 1.2$ and $|\eta| > 1.2$ ranges [97].

The track reconstruction algorithm described in this section works well for all charged particles but, in the case of electrons, it does not take into account the energy loss due to bremsstrahlung. Therefore, a Gaussian Sum Filter (GSF) model [100] is adopted, which consists of refitting the tracks loosely matched to an ECAL seed cluster to better estimate the track parameters. The GSF assumes that the energy losses can be approximated by a weighted sum of n Gaussian functions, each one representing the quantised components of the radiative loss of electrons as they traverse the pixel and SCT trackers. In the TRT, the GSF does not give a significant improvement in the electron track reconstruction due to the more homogeneous distribution of the detector material and the lower precision of the measurements.

Primary vertex reconstruction The primary vertex reconstruction procedure can be divided into a vertex finding and a vertex fitting stage [101]. The first one deals with the association of the reconstructed tracks to vertex candidates, the second reconstructs the position of the primary vertex. The tracks entering the vertex finding stage must fulfill more stringent criteria with respect to the track reconstruction step, in order to reduce tracks originating from pile-up. Only tracks with $p_T > 400$ MeV, $|\eta| < 2.5$, and satisfying requirements on the number of silicon and pixel hits enter the vertex reconstruction algorithms [102].

During the vertex finding stage, the seed positions of candidate vertices are found based on the reconstructed position and width of the beam spot. This is derived through a three-dimensional Gaussian measurement centred around the beam spot centre and having the beam spot size as width. Tracks outside the beam spot do not enter the vertex fitting stage. At this stage, the seed is given as input to an adaptive vertex fitting

algorithm [103] used to derive the optimal vertex position. Each input track is assigned a weight depending on its compatibility with the initial vertex position estimate; the higher the weight the larger its influence in the vertex position calculation. After the last iteration, the final weight of each track is calculated and tracks found incompatible with the best fit vertex position by more than 7σ are removed and allowed to be used in the reconstruction of other vertices.

The primary vertex corresponding to a hard-scattering physics process is identified as the primary vertex among all candidates with the highest sum of squared transverse momenta of all contributing tracks, $\sum p_T^2$. This criterion is based on the assumption that the charged particles produced in hard-scattering interactions have a harder p_T spectrum than those from pile-up collisions.

3.2 Electrons

Electrons are reconstructed from localised clusters of energy in the electromagnetic calorimeter, matched to charged-particle tracks in the inner detector.

3.2.1 Clusters reconstruction

The electromagnetic (EM) energy cluster candidate is reconstructed using a *topological cell clustering*-based approach [104], developed to recover the energy lost by the electron due to bremsstrahlung interactions with the ID material. This approach uses *superclusters* which associate the primary cluster formed by the electron shower (*seed-cluster*) to the secondary cluster produced by the irradiated photon (*satellite-cluster*). Figure 3.4 shows the diagram of a supercluster with a seed electron cluster and a satellite photon cluster.

The main advantage of this methodology is the ability of following the natural development of the EM cluster, both laterally and longitudinally, by varying the size of the cluster to capture most of the shower energy.

The seed and growth of topological clusters is controlled by the *cell significance*, defined as:

$$\zeta_{\text{cell}}^{\text{EM}} = \left| \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise, cell}}^{\text{EM}}} \right|, \quad (3.3)$$

where $|E_{\text{cell}}^{\text{EM}}|$ is the absolute cell energy and $\sigma_{\text{noise, cell}}^{\text{EM}}$ is the expected electronic plus pile-up cell noise. The seed topo-cluster candidates (also called *proto-clusters*) are selected using only cells of the second and third layer of the ECAL to suppress noise, and by requiring $\zeta_{\text{cell}}^{\text{EM}} \geq 4$. The neighbouring cells are iteratively added to the proto-cluster if $\zeta_{\text{cell}}^{\text{EM}} \geq 2$. Once all cells with absolute energies above 2σ are added, a final set of nearby cells with $\zeta_{\text{cell}}^{\text{EM}} \geq 0$ are added to the cluster in order to control the shower development at the edges. If two proto-clusters share a cell having $\zeta_{\text{cell}}^{\text{EM}} \geq 2$, they are merged. On the other hand, in case two or more local maxima with $E_{\text{cell}}^{\text{EM}} > 400$ MeV are found, the proto-cluster is split into separate clusters, if the cell where the local maximum is located has at least

four neighbours, none having a larger signal than the local maximum.

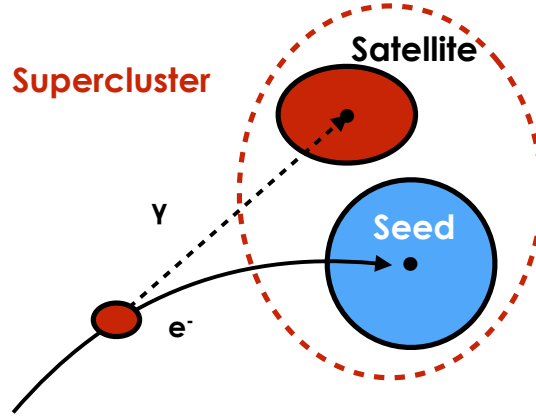


Figure 3.4: Diagram of a supercluster formed by a seed electron cluster and a satellite photon cluster [104].

3.2.2 Track-cluster matching and superclusters

Once tracks and clusters are built, they are matched together to build the final electron candidate. Electron tracks having at least four silicon hits are extrapolated to the middle layer of the EM calorimeter, followed by the application of a series of geometrical requirements in order to match the track to the corresponding EM cluster:

- the distance in η between the EM cluster and the track has to be within 0.05;
- the distance between the track and the EM cluster has to be within 0.1 in ϕ in the bending direction and 0.05 on the opposite direction;
- tracks reconstructed only in the TRT must be within $|\Delta\eta| < 0.35$ ($|\Delta\eta| < 0.02$) in the barrel (endcaps), $|\Delta\phi| < 0.03$ in the bending direction, and $|\Delta\phi| < 0.02$ for the opposite direction in both barrel and endcaps.

Superclusters are built after the track-cluster matching starting from the highest p_T proto-cluster and iterating over all satellite clusters. The supercluster seed is required to have $E_T > 1$ GeV ($E_T > 1.5$ GeV) for electrons (photons) and must be matched to a track with at least four hits in the silicon detector. For both electrons and photons a satellite cluster is required to fall within a certain $\Delta\eta \times \Delta\phi$ window around the seed barycentre. Once all satellite clusters have been associated to a given seed cluster, the algorithm iterates to the next highest p_T seed cluster that has not been used already as a satellite cluster, until no more seed candidates are found.

3.2.3 Identification

The electron identification criteria aim at discriminating between prompt isolated electrons from non-isolated electrons in hadronic jets, converted photons, electrons pro-

duced in the decay of heavy-flavours hadrons, or “hadronic fakes”, in which electromagnetic deposits are combined with random hadronic tracks. A multivariate analysis based on a likelihood (LH) discriminant is built from tracking, calorimeter and combined variables, each being described by a probability distribution function (PDF). This is created by smoothing the histogram of the discriminating variable with an adaptive kernel density estimator [105]. The PDFs for both signal and background, and for each discriminating variable, are combined into a final discriminant d_L on which three operating points are defined: *Loose*, *Medium* and *Tight*. These are optimized for identifying prompt electrons with $E_T = 40$ GeV with efficiencies of 93%, 88% and 80%, respectively, gradually increasing with higher values of E_T . The identification efficiency is evaluated using $Z \rightarrow ee$ and $J/\Psi \rightarrow ee$ events and is reported in Figure 3.5 for the three working points. An overall systematic uncertainty on the efficiency is estimated as $\pm 7\%$ at $E_T = 4.5$ GeV, and decreases to $\pm 1\%$ for $30 < E_T < 250$ GeV. For larger values of E_T , additional uncertainties of $\pm 0.5\%$, $\pm 1.0\%$ and $\pm 1.5\%$ are assigned due to the variations in the efficiency for LH-Loose, LH-Medium and LH-Tight selected electrons. The efficiencies derived from simulation are matched to those in data by applying scale factors derived on pure $Z \rightarrow ee$ events with the Tag and Probe method [106].

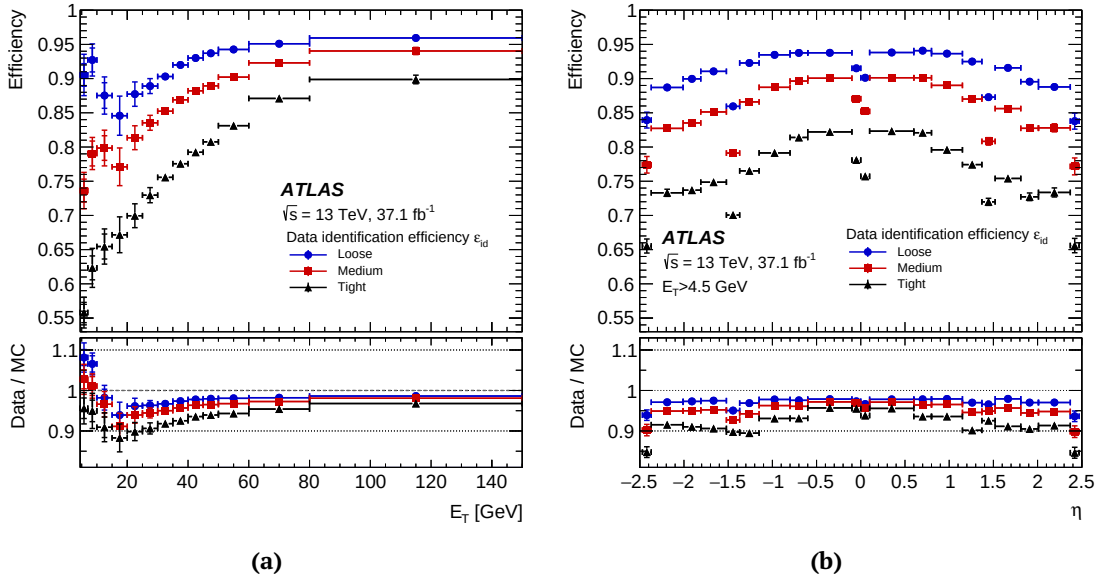


Figure 3.5: Measured LH electron identification efficiencies in $Z \rightarrow ee$ events in data for the Loose (blue circle), Medium (red square), and Tight (black triangle) operating points as a function of the electron E_T (a) and η (b). The vertical bars are very small and represent the statistical (inner bars) and total (outer bars) uncertainties. The efficiencies are obtained by applying data-to-simulation efficiency ratios measured in $J/\Psi \rightarrow ee$ and $Z \rightarrow ee$ events to $Z \rightarrow ee$ simulation. In the bottom panel, the data-to-simulation ratio is shown [106].

3.2.4 Isolation

In order to further reject background electrons from hadron fakes or semileptonic heavy-flavours decays, track-based and calorimeter-based isolation requirements are applied

to the electron candidates.

The energies of all positive-energy topo-clusters are summed together in a raw isolation energy variable, $E_{T,raw}^{isol}$, if their barycentre falls within a cone of fixed radius $\Delta R = 0.2$ around the electron candidate. The core energy from the electron candidate is subtracted from $E_{T,raw}^{isol}$ by removing the cells contained in a rectangle of $\Delta\eta \times \Delta\phi = 0.125 \times 0.175$ around the candidate's direction.

While this methodology results in a very stable and simple subtraction scheme of fake or non-prompt objects, the electron candidate energy might not be fully contained inside the fixed rectangular area. Therefore a leakage correction to the subtracted core energy is applied. Additionally, a correction for the pile-up and underlying event contribution is estimated from the ambient energy density [107].

The track-based isolation variable, $p_T^{varconeXX}$, corresponds to the sum of the transverse momenta of all selected tracks falling into a cone centred around the electron direction, excluding tracks matched to the electron candidate or converted photons. The size of the cone varies depending on the p_T of the electron: the higher the p_T , the smaller the cone size. Typically, the maximum radius XX of the track-based isolation variable is 0.2. Only the tracks used in the primary vertex reconstruction, having $p_T > 1$ GeV and $|\eta| < 2.5$ and a given number of hits in the silicon (pixel and SCT) detectors, are considered in the calculation of $p_T^{varconeXX}$.

The isolation operating points relevant for the electron energy calibration and the $VH, H \rightarrow b\bar{b}$ analysis described in Chapter 6 are:

- *HighPtCaloOnly*, optimised on fixed radius giving an efficiency of 98% for $p_T^{cone0.2} < \max(0.015 \times p_T, 3.5)$ GeV;
- *Fixed-Cut Loose*, using both a calorimeter-based isolation requirement of $E_T^{cone0.2}/p_T < 0.20$ and a track-based isolation requirement of $p_T^{varcone0.2}/p_T < 0.15$.

The isolation efficiencies are evaluated using $J/\Psi \rightarrow ee$ and $Z \rightarrow ee$ events up to 150 GeV and as a function of E_T and η . The efficiency derived in simulation are matched to the data by applying scale factors very close to unity. The relative uncertainties vary from 5% for electrons with $E_T < 7$ GeV to less than 0.5% with increasing values of E_T .

The electron energy scale and resolution are calibrated using $Z \rightarrow ee$ events by fitting the mass line shape of the resonance in both simulation and data in different regions of η . The ratio between the two fits for the energy scale and the resolution is taken as a correction factor applied to the data or the simulation, respectively.

3.3 Muons

Muons leave a tiny fraction of their energy inside the inner detector and the calorimeters and are the only detectable particles able to reach the outer layers of the ATLAS detector. Their reconstruction and identification is extremely important both for precision measurements and new physics searches.

3.3.1 Reconstruction

Muon candidates are reconstructed independently from measurements in the inner detector (ID) and the muon spectrometer (MS) [108]. Four muon types are defined based on the information provided by the ID, the MS and the calorimeters: the *standalone* or *extrapolated* muons, the *combined* muons, the *segment-tagged* muons, and the *calorimeter-tagged* muons.

The *standalone* or *extrapolated* muons (ME) are reconstructed starting from the middle layers of the MS, where more hits are expected. Using a Hough transformation [109], a pattern of hits aligned with a preliminary hypothesised trajectory is created, forming a track candidate. The track is then extrapolated back to the point of closest approach to the interaction point. Standalone muons are used to recover the muon reconstruction acceptance in the region $2.5 < |\eta| < 2.7$, not covered by the ID.

The *combined* (CB) muons result from a global fit of the ID and MS hits which follows an outside-in pattern recognition where muons reconstructed in the MS are extrapolated to the ID, taking into account the energy losses in the calorimeters.

When a muon has a low p_T or falls into the region of reduced acceptance of the MS, it traverses only one layer of the MS chambers. In this case, if the track in the ID is matched to at least one track segment in the MDT or CSC chambers, the muon is reconstructed as *segment-tagged* (ST).

In the region where the MS is only partially instrumented, due to the ID cables and the calorimeter services ($|\eta| < 0.1$), the track in the ID is identified as a muon if it is matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle (MIP). This type of muon is called *calorimeter-tagged* (CT) muon.

When two muon types share the same track in the ID, the overlap is resolved by preferring the CB category first, followed by ST and finally CT. The track quality and the largest number of hits are taken as discriminant to resolve the overlap with the ME category.

The overall reconstruction efficiency is calculated using $J/\Psi \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu$ events with a Tag and Probe method and measured to be close to 99% over most of the covered space $|\eta| < 2.5$ and $5 < p_T < 100$ GeV range [108]. As for the electrons, efficiency scale factors are calculated to correct the simulation to the data. On the same set of events, the calibration of the muon momentum resolution and scale is obtained with a simultaneous fit to the mass line shape of the resonance and the p_T imbalance between the measurements in the ID and the MS. By combining the two results, correction factors are derived in different p_T and η bins.

3.3.2 Identification

Muon candidates originating from charged hadrons decays in the ID are characterized by a deflection in their flight path due to the displaced topology of the decay vertex. Therefore, these tracks are expected to have a worse fit quality and a bad compatibility

between the ID and MS tracks. The following criteria are applied to discriminate these fake muons against prompt muons [108]:

- q/p significance, defined as the ratio between the difference in charge and transverse momentum measured in the ID and MS and normalised by the corresponding uncertainties⁴;
- ρ' , defined as the ratio between the absolute value of the difference of the transverse momenta measured in the ID and the MS divided by p_T of the combined track⁵;
- at least one pixel hit and five SCT hits with less than three holes in both silicon detectors;
- normalized χ^2 value of the combined track fit.

Four identification selections are defined using different cuts on these variables: *Loose*, *Medium*, *Tight*, *High- p_T* . The Loose identification gives the best muon reconstruction efficiency whilst the Tight criteria offer the best muon purity. The High- p_T selection, instead, offers the best muon resolution for muons with $p_T > 100$ GeV.

The muon reconstruction efficiency measured using $Z \rightarrow \mu\mu$ events as a function of the muon η and for muons with $p_T > 15$ GeV is shown in Figure 3.6 (a) for the three identification selections. Loose and Medium selections have very similar efficiencies, with the exception of the $|\eta| < 0.1$ region, where the Loose selection fills the MS acceptance gap using both segment-tagged and calorimeter-tagged muons. The muon reconstruction efficiency as a function of the muon p_T and for muons passing the medium selection is shown in Figure 3.6 (b).

3.3.3 Isolation

Two isolation variables are defined, a track-based isolation variable using only the ID information, and a calorimeter-based isolation variable when only the calorimeter measurement is employed.

The track-based isolation variable can be defined either with a fixed radius of $\Delta R = 0.2$, in which case is referred to as $p_T^{\text{cone0.2}}$, or a p_T -dependent size of $\Delta R = \min(10 \text{ GeV}/p_T^\mu, 0.3)$, in which case it is called $p_T^{\text{varcone0.3}}$, and it is defined as the sum of transverse momenta of all tracks around the muon candidate excluding its track p_T . The variable radius is useful to improve the performances for muons produced from the decays of particles with high transverse momentum.

The calorimeter-based isolation variable $E_T^{\text{topcone0.2}}$ is defined as the sum of transverse energy of topological clusters in a cone of size of $\Delta R = 0.2$ around the muon candidate, with the exception of its energy deposit and correcting for pile-up effects [107].

⁴This variable is defined as: $\sigma_{q/p} = \frac{(q/p^{\text{MS}}) - (q/p^{\text{ID}})}{\sqrt{[\sigma(q/p^{\text{MS}})]^2 + [\sigma(q/p^{\text{ID}})]^2}}$

⁵This variable can be written as: $\rho' = \frac{p_{T,\text{ID}} - p_{T,\text{MS}}}{p_{T,\text{CB}}}$

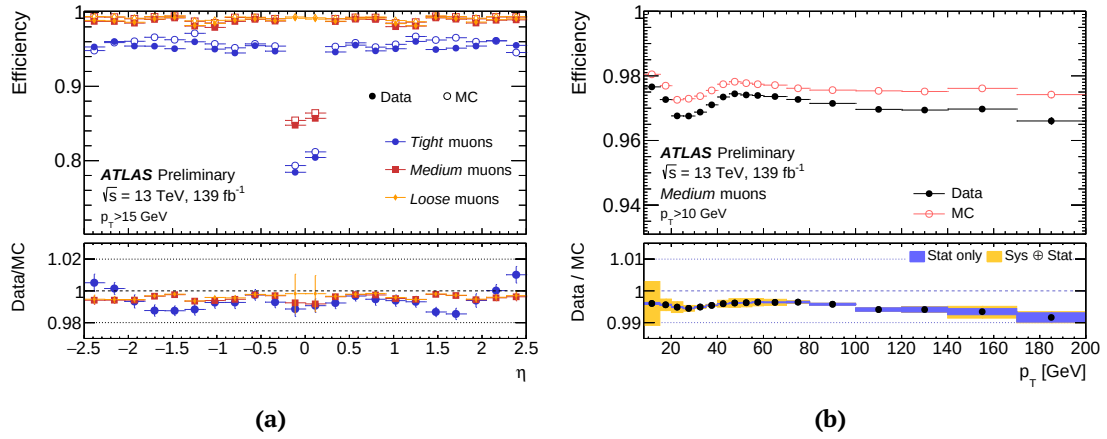


Figure 3.6: Figure (a) shows the muon reconstruction efficiency measured using $Z \rightarrow \mu\mu$ events as a function of the muon η and for muons with $p_T > 15$ GeV for the Loose (yellow rhombus), Medium (red square) and Tight (blue circle) identification selection using the full Run 2 dataset. Figure (b) shows the muon reconstruction efficiency for the medium selection as a function of the muon p_T . The expected efficiency as predicted by the simulation is shown as open symbols, while filled dots indicate the observed efficiency measured using the full Run 2 dataset with statistical uncertainties. In both figures, the bottom panel shows the scale factors as the ratio between the observed and expected efficiencies [110].

Several working points are defined for different analysis purposes. The $VH, H \rightarrow b\bar{b}$ analysis described in Chapter 6 uses the *HighPtTrackOnly* optimized using a fixed radius cone giving an efficiency greater than 95% for prompt muons (η and p_T dependent), and the *Fixed-Cut Loose* with a variable cone size for both the track-based and the calorimeter-based variables yielding a 99% isolation efficiency for prompt muons across the full p_T spectrum. The muon isolation efficiency for prompt muons is measured using $Z \rightarrow \mu\mu$ events in the momentum range between 10 and 120 GeV with the simulation being matched to the data using scale factors compatible with unity [108].

3.4 Tau leptons

With a mass of $m_\tau = 1.777$ GeV [45], the τ is the only lepton heavy enough to decay both into leptons and hadrons, decaying leptonically in 35% of the cases ($\tau \rightarrow e^- \bar{\nu}_e \nu_\tau$ and $\tau \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$), and hadronically (typically pions or kaons) in the remaining 65%. Having a proper decay length of $87\mu\text{m}$ [45], τ -leptons typically decay before the active region of the ATLAS detector, therefore they can only be identified from their decay products.

Hadronic τ decays are characterized by a clear signature in the detector represented by a “spray” of very collimated particles, called a τ -jet, containing one or three charged pions in 72% and 22% of the cases, respectively. Charged kaons are present in the majority of the remaining hadronic decays. Jets from τ decays are reconstructed using the anti- k_t algorithm (see Section 3.5) with a radius of $R = 0.4$ and calorimeter topo-clusters as input. Kinematic requirements are imposed on τ candidates of $p_T > 20$ GeV and $|\eta| < 2.5$. A BDT-based identification is then performed to reject the main background contribution

given by jets of energetic hadrons produced via quarks or gluons fragmentation [111].

Leptonic τ decays are reconstructed using the electrons and muons selection criteria defined in Sections 3.2 and 3.3.

3.5 Jets

The hadronization of quarks and gluons produces a spray of collimated hadrons called jet. The primary aim of the jet reconstruction algorithm is to re-combine all particles from the hadronization to obtain a physics object whose characteristics are as close as possible to those of the initial parton. The jet definition changes according to the type of inputs used to reconstruct the jet. Jets reconstructed using truth stable particles from Monte Carlo simulation are called *particle jets*, jets built from a combination of reconstructed tracks in the detector are referred to as *track jets* and jets reconstructed from the energy deposits in the calorimeter topo-clusters (see Section 3.2.1) are called *topo-cluster jets*. This last definition is used by many ATLAS analyses, including the $VH, H \rightarrow b\bar{b}$ one described in Chapter 6.

An alternative approach was developed exploiting the measurements from both the tracker and the calorimeter to build an object called *particle-flow jet* [112]. The advantage of this new jet definition is the combination of two complementary measurements. The calorimeter has a better momentum resolution at high- p_T and provides energy measurement in the region not covered by the tracker. On the other hand, the tracker has a better momentum resolution at low- p_T , a better angular resolution and the ability to disentangle multi-parton interactions to suppress pile-up vertices.

In the following, only topo-cluster jets will be discussed as they represent the default jet definition used in the $VH, H \rightarrow b\bar{b}$ analysis. The algorithm used to re-combine the topo-clusters to form a jet is the *anti- k_t* algorithm [113]. Topo-clusters are sequentially added one at a time according to the following distance measures:

$$d_{ij} = \min\left(\frac{1}{k_{ii}^2}, \frac{1}{k_{ij}^2}\right) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.4)$$

$$d_{iB} = \frac{1}{k_{ii}^2}, \quad (3.5)$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and $k_{i(j)}$, $y_{i(j)}$, $\phi_{i(j)}$ are the transverse momentum, the rapidity and the azimuthal angle of cluster $i(j)$, respectively, and R is a fixed parameter of the algorithm related to the radius of the jet. The algorithm starts from building all possible d_{ij} and d_{iB} , finding the smallest among them. If it is a d_{ij} , then clusters i and j are recombined, if it is a d_{iB} , the cluster i is defined as a jet and removed from the list of available topo-clusters. This can be rephrased saying that the hard particles are clustered first with respect to the soft ones. The distances are then recalculated and the procedure repeated until no more clusters are found. A nice feature of the anti- k_t algorithm is that the jets built from hard particles have a circular shape in the $\eta - \phi$ plane around the jet

axis, as shown in Figure 3.7.

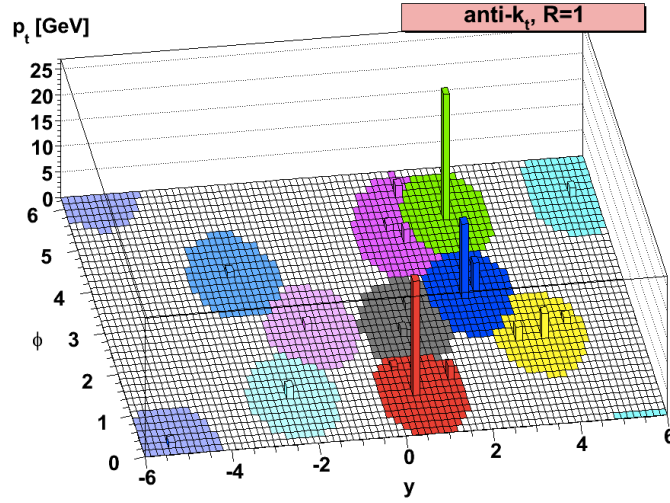


Figure 3.7: A sample parton-level event (generated with HERWIG [114]), clustered with the anti- k_t jet algorithm with radius $R = 1$, illustrating the circular areas of the resulting hard jets [113].

3.5.1 Jet Energy Calibration

The jet energy scale (JES) calibration aims at restoring the energy scale of the jets reconstructed in the detector to the one of the truth jets. Five different corrections are applied to account for pile-up contaminations, energy lost in dead material, showers not fully contained in the jet cone (out-of-cone effects) and non-uniform response of the electromagnetic and hadronic calorimeters.

Origin correction The reconstructed jet is expected to originate from the hard scatter vertex, therefore the direction of its four momentum is corrected so that it points to the primary vertex of the interaction.

Pile-up correction The pile-up contribution inside a jet is estimated from the median p_T density of the reconstructed jets in the $y - \phi$ plane, divided by the jet area. Once subtracted from the total jet energy, an additional correction is estimated at Monte Carlo level as a function of the mean number of interactions per bunch crossings $\langle \mu \rangle$ (see Equation 2.3) and the number of reconstructed primary vertices in the event.

Absolute JES calibration This correction is applied to the reconstructed jet four-momentum in order to match it to the particle-level energy scale. First, the reconstructed jets are matched to the truth jets within a cone of $\Delta R = 0.3$, then the jet energy response is calculated from a fit to the $E^{\text{reco}}/E^{\text{truth}}$ ratio, which is assumed to be gaussian-distributed, in bins of E^{truth} and η .

Absolute η -intercalibration The jet energy response can vary as a function of η , especially at boundaries between different calorimeter technologies and granularities. An

additional correction is derived from the ratio between the reconstructed and truth jet η , parametrized as a function of the reconstructed jet η and truth jet energy E^{truth} .

Global sequential calibration The jet energy response can be particularly sensitive to the jet flavour, the energy distribution of the constituent particles, and the initiating particle. In fact, the hadrons contained in a quark-initiated jet carry a large fraction of the jet p_T , whereas a gluon-initiated jet typically includes much softer particles. These fluctuations in the jet response are reduced by a series of multiplicative corrections that improve the jet energy resolution without changing the average jet response.

In-situ corrections This final step aims to recover the residual differences between the jet energy response measured in simulation and data. First, an η -intercalibration is performed exploiting the p_T -balance between jets in different η regions, where the energy scale of the forward jets ($0.8 \leq \eta \leq 4.5$) is calibrated against the one of the central jets ($|\eta| < 0.8$). In a second step, the Z/γ +jets analysis uses a well calibrated Z boson decaying into an electron or muon pair, or a γ to constrain the p_T of the recoiling jet in the $20 < p_T < 500$ GeV and $36 < p_T < 950$ GeV ranges, respectively. Finally, a multi-jet method is used to extend the p_T range of the calibration up to 2 TeV using a system of well calibrated lower p_T jets balanced against a single very high- p_T jet. A convolution of all these corrections provides a smooth single calibration which can be applied to the full p_T spectrum of the data, as shown in Figure 3.8 (a).

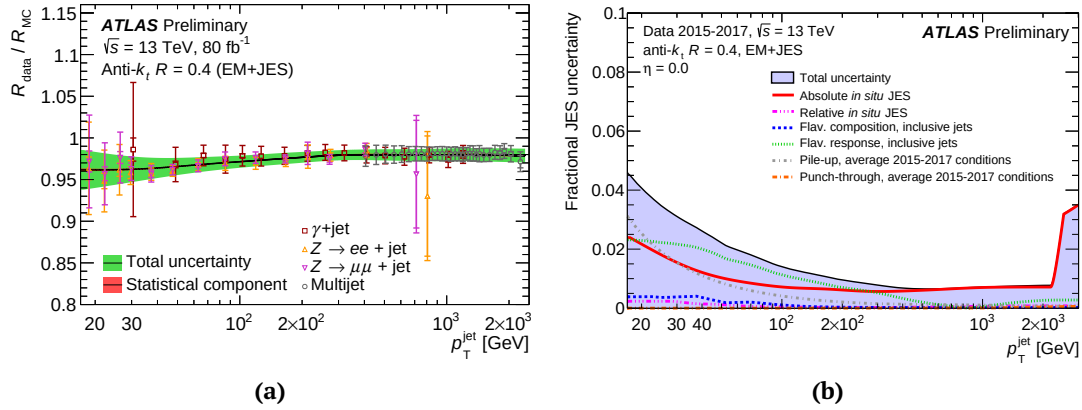


Figure 3.8: Figure (a) shows the ratio of the jet response in data to the one in simulation as a function of jet p_T for the Z +jets, γ +jets, and multi-jet *in-situ* calibrations. The final derived correction (black line) and its statistical (dark blue) and total (light green) uncertainty bands are also shown. Figure (b) shows the fractional jet energy scale systematic uncertainty components as a function of jet p_T for $R=0.4$ anti- k_t jets at $\eta = 0.0$. The total uncertainty (all components summed in quadrature) is shown as a filled region topped by a solid black line [115].

The systematic uncertainties on the jet energy scale measurement come mainly from the *in situ* calibrations and are calculated separately for the Z +jets, γ +jets and multi-jet techniques. Additional contributions come from the η -intercalibrations, the residual mis-modelling of the pile-up in Monte Carlo simulation, the difference in the jet response between jets containing heavy or light flavour quarks, and between quark or gluon-

initiated jets. Furthermore, an uncertainty related to the so called “punch-through” jets, that are high energetic jets exceeding the hadronic calorimeter, is also considered. The total jet energy scale uncertainty is shown in Figure 3.8 (b) together with each single component as a function of the jet p_T .

3.5.2 Jet Energy Resolution

A precise knowledge of the jet energy resolution (JER) is essential for many analyses which study the properties of SM particles decaying into jets, or analyses relying on a well measured missing transverse momentum. The JER measurement is performed using the same technique as for the *in situ* calibration, but using the width of the jet response in p_T instead of its mean. In the central rapidity region, the JER is measured using Z +jets and γ +jets events. In the forward pseudorapidity region and in the high- p_T region, a well defined di-jet system is used [116]. The final JER can be parametrized as:

$$\frac{\sigma(p_T)}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C, \quad (3.6)$$

where N represent the noise term due to the electronic front-end of the detector, S is a stochastic term which collects all the statistical fluctuations on the energy deposit and C is a constant term including energy losses in dead material or corrections to the non-uniform response across the calorimeter.

Figure 3.9 (a) shows the jet energy resolution measured in simulation and in di-jet events as a function of the jet p_T , while Figure 3.9 (b) shows the uncertainty on the jet energy resolution as a function of the jet p_T , separating the uncertainties from the *in situ* measurements and the data/MC simulation difference.

3.5.3 Jet selection

Jets reconstructed in the detector could also be generated by problems in the calorimeter hardware, beam-gas interactions, or showered induced by cosmic rays. These “fake” jets are rejected by applying quality criteria on the shape of the electrical signal collected in the calorimeter cells and the deposited energy.

Furthermore, the topo-clusters reconstructed in the calorimeter can be contaminated by both in-time and out-of-time pile-up. Even though the transverse energy from pile-up contributions is subtracted in the second step of the JES calibration, local fluctuations can still arise due to spurious jets. For this reason, variables able to discriminate between pile-up and hard-scattered jets are combined into a single likelihood function whose output, between 0 and 1, corresponds to the probability of a jet originating from a pile-up or a primary interaction, respectively. Based on the studies presented in Ref. [118], a value of the discriminant “Jet-Vertex-Tagger” (JVT) > 0.59 was set as default for the jet selection, finding a 92% efficiency for hard scattering jets and a 2% fake rate for pile-up jets.

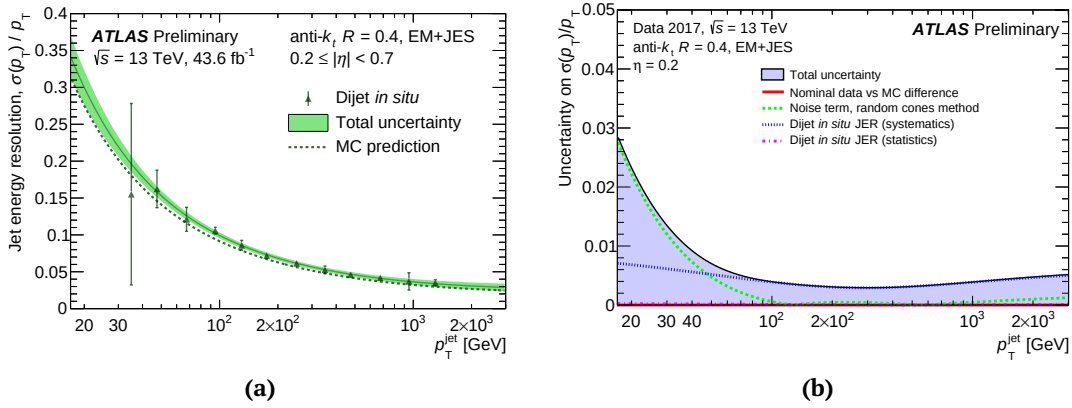


Figure 3.9: Figure (a) shows the relative jet energy resolution $\frac{\sigma(p_T)}{p_T}$ as a function of the jet p_T for anti- k_t jets with $R = 0.4$. The resolution from Monte Carlo simulation is compared to the one in data and the error bars on points indicate the total uncertainties on the derivation of the relative resolution in dijet events, adding in quadrature statistical and systematic components. The result of the combination of the *in situ* techniques is shown as the dark line. Figure (b) shows the uncertainty on the relative jet energy resolution as a function of the jet p_T for anti- k_t jets with $R = 0.4$ and $\eta = 0.2$. The total uncertainty is shown as a filled region topped by a solid black line together with the uncertainties from the *in situ* methods and the uncertainty on the noise term [117].

3.5.4 Truth Flavour Labelling

Jets in simulation can be identified by the truth flavour of the quark from which they were initiated. The labelling scheme is performed at truth level using a ΔR matching scheme. Final state hadrons with $p_T > 5$ GeV and within $\Delta R < 0.3$ from the jet axis are assigned to the corresponding jets. Since the hadron can be assigned to only one jet, the closest jet in ΔR space is chosen. The truth flavour of the jet is then assigned with the following iterative procedure:

- if a truth b -hadron is matched to the jet, the jet is labelled as a b -jet;
- if a truth c -hadron is matched to the jet, and no b -hadron is found, the jet is labelled as a c -jet;
- if a τ -lepton is matched to the jet, the jet is labelled as a τ -jet.
- else, the jet is labelled as a l -jet.

3.6 Missing transverse energy

In collider experiments, the sum of transverse momenta of the collision products must be equal to 0. Any imbalance in the transverse momentum, referred to as *missing transverse energy* E_T^{miss} , can indicate the presence not only of SM neutrinos, but also of weakly interacting particles not predicted by the SM.

Two terms contribute to the reconstruction of the E_T^{miss} , a *hard-term* and a *soft-term*, as explained in Ref. [119]. The hard-term contains only reconstructed and fully calibrated

objects compatible with the hard scatter. Since the calorimeter measurements are not exclusive, the same signal might be used to reconstruct both an electron and a jet, causing a double-counting of the same signal when calculating the E_T^{miss} . Therefore, calorimeter signals are assigned to the physics objects contributing to the E_T^{miss} in the following order: electrons (e), photons (γ), hadronically decaying τ -leptons, jets, and muons (μ).

The soft-term is reconstructed from all the additional detector objects not associated to the hard-term and can be calculated as a *track-based soft term* (TST) or a *calorimeter-based soft term* (CST). The TST term is reconstructed from ID tracks not associated to any physics objects of the hard-term. The main advantage of this method is the ability to disentangle the primary vertex from additional pile-up vertices. On the other hand, the CST term is computed from energy deposits in the calorimeter and has the advantage of accounting for both charged and neutral components and of having a larger coverage. However, given the poor angular resolution of the calorimeter, it is very difficult to associate a cluster to a specific primary vertex, therefore energy deposits due to pile-up are summed in the total energy of the topo-cluster. For this reason, only the TST term is considered in the computation of the E_T^{miss} soft-term in ATLAS.

The $x(y)$ components of the reconstructed E_T^{miss} can be parametrized as follows:

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss},e} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss},\text{jets}} + E_{x(y)}^{\text{miss},\mu} + E_{x(y)}^{\text{miss},\text{soft}}, \quad (3.7)$$

where each term is given by the negative vectorial sum of the momenta of the respective calibrated objects. The final missing transverse energy is obtained as a sum in quadrature of its x and y components:

$$E_T^{\text{miss}} = \sqrt{E_x^{\text{miss},2} + E_y^{\text{miss},2}}. \quad (3.8)$$

Three working points, *Loose*, *Medium* and *Tight* are defined according to the requirements used to select the jets entering the E_T^{miss} reconstruction, as explained in Ref. [120]. Systematic uncertainties on the E_T^{miss} are divided into two main sources: the uncertainties affecting the calibration of the physics objects entering the hard-term and those on the soft-term which are evaluated using $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events with no expected E_T^{miss} , and no jets with $p_T > 20$ GeV. From the width of the E_T^{miss} distribution one gets a measure of the resolution of the E_T^{miss} reconstruction. The main sources of uncertainty on the E_T^{miss} reconstruction come from the JES and the η -intercalibration, while the JER enters in the E_T^{miss} resolution.

3.7 Overlap removal

In the offline reconstruction it is possible that a signal left in a sub-detector is used to reconstruct multiple physics objects. To avoid double-counting, an “overlap removal” procedure using only reconstructed and fully calibrated objects is defined as follows [121]:

Jet-electron If $\Delta R(e, \text{jet}) < 0.2$, the jet is removed as the electron cluster was likely used to reconstruct the jet. Without this criterion, almost all electrons would be double-counted as jets.

Electron-jet If an electron is found within $\Delta R(e, \text{jet}) < 0.4$ from the jet axis, the electron is removed as likely to be originating from a heavy flavour decay inside the jet.

Muon-jet If they lie within a cone of $\Delta R(\mu, \text{jet}) < 0.2$ or the muon track extrapolated to the ID is associated to the jet, the jet is removed if it contains less than 3 tracks with $p_T > 500$ MeV or if the ratio between the muon and jet momentum exceeds 0.5 or if the ratio between the muon p_T and the sum of the p_T of all tracks associated to the jet is less than 0.7. For any surviving jet, the muon is removed if $\Delta R(\mu, \text{jet}) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^\mu)$, as likely to be originated from a heavy flavour decay.

Tau-jet If the τ lies within a cone of $\Delta R(\tau, \text{jet}) < 0.2$ from the jet axis, the jet is removed.

Electron-muon If a combined muon shares an ID track with an electron, the electron is removed. On the other hand, if a calorimeter-tagged muon shares a ID track with an electron, the muon is removed.

Tau-electron If $\Delta R(\tau, e) < 0.2$, the τ is removed.

Tau-muon If $\Delta R(\tau, \mu) < 0.2$, the τ is removed, unless it has $p_T > 50$ GeV and the μ is not a combined-muon.

Flavour Tagging in ATLAS

The identification of jets containing b -hadrons plays an important role in many areas of the ATLAS experiment physics programme. It has been crucial in the observation of the $H \rightarrow b\bar{b}$ decay mode and the $t\bar{t}H$ production mode, and it is a key ingredient for a large number of Standard Model precision measurements or searches for new physics phenomena.

The algorithms developed to identify b -jets from c -jets and l -jets, referred to as b -tagging algorithms, exploit the long lifetime and high mass of b -hadrons, as well as their high decay multiplicity and the properties of the b -quark fragmentation.

Since b -hadrons lifetimes are of the order of 1.5 ps, in the laboratory frame their average decay length is increased by the boost to $\langle l \rangle = \beta\gamma c\tau$. A b -hadron of $p_T = 50$ GeV will travel about 3 mm in the transverse direction before decaying, leading to topologies with at least one decay vertex.

Two classes of algorithms aim at identifying such topologies. The first consists of using the impact parameters of the tracks associated to the charged products of the b -hadron decay. The second approach uses the tracks to explicitly reconstruct the secondary vertex or the full decay chain. The most discriminating observables of these *low-level* algorithms are combined in a multivariate analysis which provides the best separation between the different jet flavours.

This chapter contains an overview of the b -tagging algorithms used by the ATLAS Collaboration and a description of the statistical tagging which is of particular interest for the $VH, H \rightarrow b\bar{b}$ analysis described in Chapter 6.

4.1 Track selection and track-jet matching

The selection of the tracks used in the b -tagging algorithms is designed to reject fake tracks (in which not all hits used for the track reconstruction originate from a single charged particle), tracks from long-lived particles such as K_S or Λ , and material interactions like photon conversion or hadronic interactions. The track selection applied for the impact parameter algorithms requires tracks with $p_T > 1$ GeV, $|d_0| < 1$ mm and $|z_0| \sin\theta < 1.5$ mm, and at least seven silicon hits with at most two silicon holes, at most one of which is in the pixel detector.

The Secondary Vertex (SV1) [122] and the JETFITTER [123] algorithms use a looser selection to maximise the sensitivity of identifying tracks from material interactions or long-lived particles which are then removed for b -tagging purposes. Tracks with $p_T > 500$ MeV, $|d_0| < 3.5$ mm and no cut on z_0 ($|z_0| \sin\theta < 5$ mm) are selected by the SV (JETFITTER) algorithm [124]. Both SV1 and JETFITTER require at least one hit in the pixel detector.

The first step of a b -tagging algorithm is to associate the track from a decay product to the corresponding jet which is reconstructed from topological clusters of energy in the calorimeters. This is done by exploiting the angular separation ΔR between the track momenta and the jet axis, defined as the vectorial sum of the four-momenta of the clusters associated with the jet [125]. The cut on ΔR varies as a function of the jet p_T , as the tracks inside a high- p_T jet are more collimated than in a low- p_T jet. For jets having $p_T = 20$ GeV, $\Delta R_{cut} = 0.45$, while for jets having $p_T = 150$ GeV, $\Delta R_{cut} = 0.26$ [124]. In case more than one jet fulfills this requirement, the closest jet is chosen.

4.2 Low-level b -tagging algorithms

This section summarises the different algorithms used to separate b -jets from c -jets and l -jets. The information carried by the impact parameter (IP) of the reconstructed tracks is used by the IP2D, IP3D and IPRNN algorithms, the SV1 algorithm aims at reconstructing a single secondary vertex in the jet while the JETFITTER technique is able to resolve the topological vertex structure of b -hadron decays inside the jet.

4.2.1 Impact Parameter based Algorithms

The transverse impact parameter d_0 is defined as the distance of closest approach of the track to the primary vertex in the $r - \phi$ plane. The longitudinal impact parameter z_0 corresponds to the z coordinate of the track at the point of closest approach (but $z_0 \sin \theta$ is finally used in the impact parameter based algorithms). A schematic view of the transverse impact parameter is shown in Figure 4.1.

This is a powerful information to discriminate between tracks originating from a b -hadron and tracks coming from the primary vertex. In fact, given the displaced topology of the b -hadron decay, the tracks associated to its decay products will typically have an impact parameter larger than zero in absolute value. On the contrary, the impact parameter of the tracks in a l -jet will be compatible with zero in absolute value. A sign is given to the impact parameter to further discriminate between the b -hadrons children and the tracks coming from the primary vertex. The sign is positive if the track extrapolation crosses the jet direction in front of the primary vertex, and negative otherwise. Therefore, tracks originating from the decay of a b - or a c -quark will have a positive sign. The signed transverse and longitudinal impact parameter significances are used as discriminants between different jet-flavour hypotheses. These are defined as $S_{d_0} \equiv d_0/\sigma_{d_0}$ and $S_{z_0} \equiv z_0/\sigma_{z_0}$, where σ_{d_0} and σ_{z_0} are the uncertainties on the reconstructed d_0 and z_0 . Figure 4.2 shows their distributions for good quality tracks, meaning those having a hit in all pixel and SCT layers (see Ref. [125] for details on the track categorization), selected inside b -jets, c -jets and l -jets. The asymmetric shape of the distribution between the b -jets and c -jets, particularly enhanced in the positive sign region, can be clearly observed.

IP2D and IP3D The IP2D and IP3D algorithms [126] rely on the signed IP significance. They are based on a log-likelihood ratio (LLR) method in which, for each track, the mea-

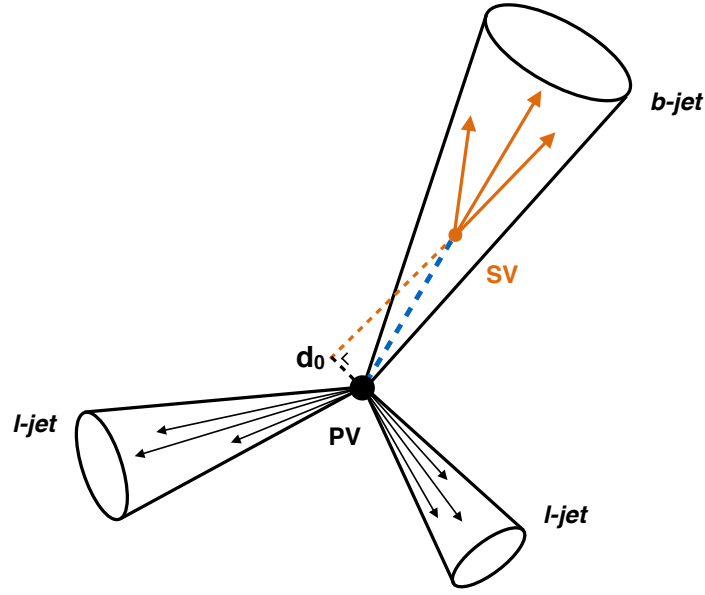


Figure 4.1: Schematic view of the transverse impact parameter d_0 .

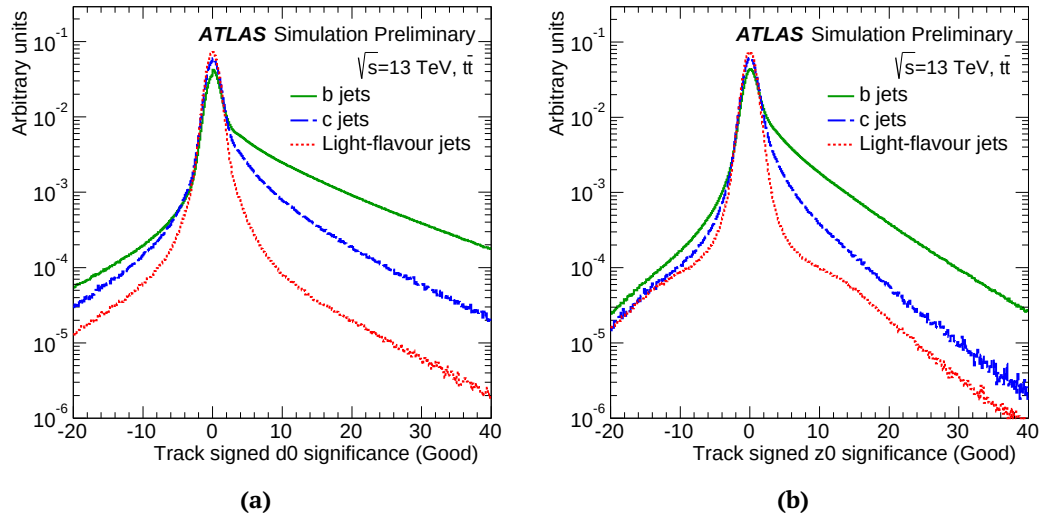


Figure 4.2: Transverse (a) and longitudinal (b) impact parameter significance in $t\bar{t}$ events of tracks associated to b -jets (green), c -jets (blue) and l -jets (red) and belonging to the “Good” category (see Ref. [125] for more details on the track categorization).

sured IP significance is compared to probability density functions (PDFs) obtained from MC simulation for b -, c - and light-flavour-jet hypotheses. Tracks are divided into exclusive categories depending on the hit pattern left in the pixel and the SCT sub-detectors, and a single PDF is derived for each category in order to increase the discrimination power between the different jet flavours. More details on this categorization can be found in Ref. [127]. A $t\bar{t}$ sample is used to derive the PDFs for all track categories except those where no hits are found in the first two layers of the pixel detector; these categories mainly contain b -jets in which the b -hadron decays after the second pixel layer and are better simulated by a $Z' \rightarrow t\bar{t}$ ¹ sample containing a large fraction of high- p_T b -jets. The PDFs are obtained using the 1D distribution of the transverse IP significance in the case of the IP2D tagger and the 2D distribution of the transverse and longitudinal IP significance for the IP3D tagger.

The ratio of PDFs for each track defines the track weight while the jet weight is obtained by summing the logarithms of the individual track weights, so that the final log-likelihood ratio between two different jet flavour hypotheses $P_{i(j)}$ can be written as:

$$\log\left(\frac{P_i}{P_j}\right) = \sum_{k=1}^{k=N} \frac{\text{PDF}_i(\text{IP}_k)}{\text{PDF}_j(\text{IP}_k)}. \quad (4.1)$$

where IP_k corresponds to the impact parameter significance of track k , N is the total number of tracks and $\text{PDF}_{i(j)}$ represents the probability density function for the jet flavour hypothesis $i(j)$.

Three LLRs are built to separate b -jets from c -jets, b -jets from l -jets and c -jets from l -jets for both taggers. The output of the LLRs, which are shown in Figure 4.3, are used as inputs to the high-level taggers.

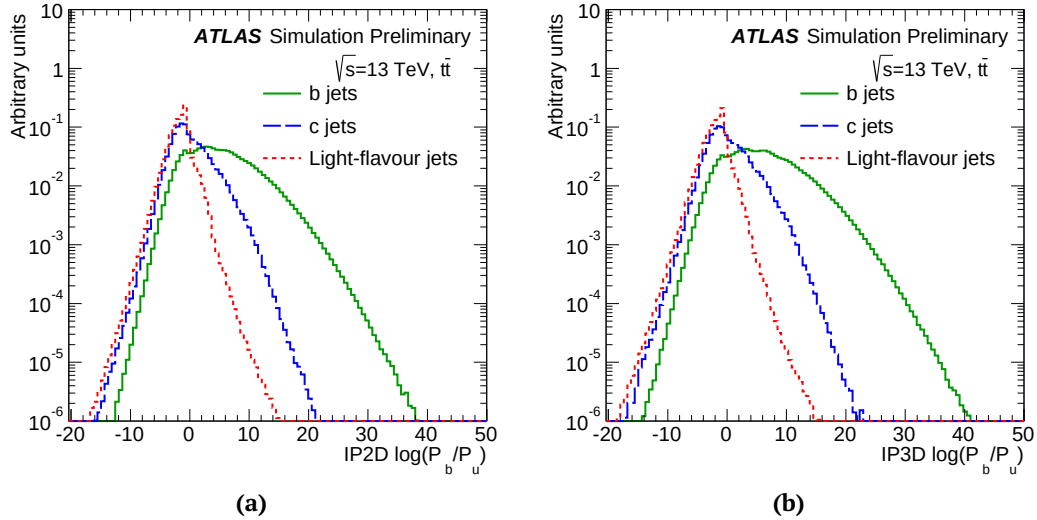


Figure 4.3: Log-likelihood ratios for the IP2D tagger in (a) and IP3D in (b) for b -jets (solid green), c -jets (dashed blue) and l -jets (dotted red) [125].

¹ Z' is a heavy gauge boson predicted by physics beyond the SM theory. In this case, the Z' is a Dark Matter mediator.

IPRNN Many tracks can originate from the decay of a b -hadron inside a jet. One of the main assumptions of the IPXD algorithms is that the per-track flavour conditional likelihood can be computed independently for each track. This does not take into account the fact that if a track is found with a high impact parameter, then it is more likely to find another track with a high impact parameter. If no displaced vertex is present, as in light jets, this correlation between tracks does not exist. This is shown in Figure 4.4, where the 2D distribution of the transverse impact parameter significance S_{d_0} for the leading and subleading transverse impact parameter significance tracks is plotted for b -jets and l -jets.

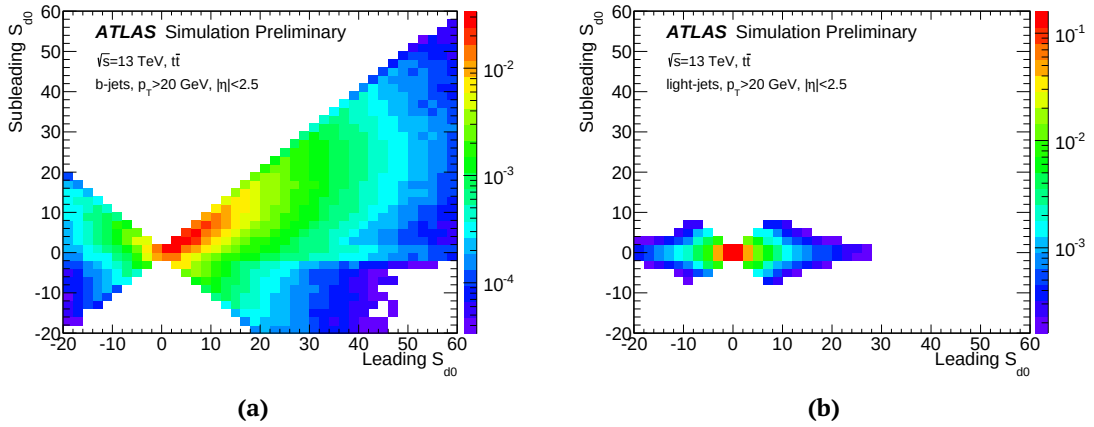


Figure 4.4: Distribution of the d_0 significance for the leading and subleading d_0 significance tracks in b -jets (a) and l -jets (b) in $t\bar{t}$ simulated events [128].

Recurrent neural networks (RNN) are used to overcome this limitation [129]. For each track the transverse S_{d_0} and longitudinal S_{z_0} impact parameter significances, the distance between the track and the jet axis $\Delta R(\text{track}, \text{jet})$, the fraction of transverse momentum of the jet carried by the track and the track category, as defined in the IPXD algorithms, are passed as inputs to the recurrent neural network. The outputs of the algorithm are the probability density functions for a track to be originated by a b -jet (p_b), a c -jet (p_c), a l -jet (p_l) and a τ -jet (p_τ). These are combined into a discriminant function for the b -jet hypothesis:

$$D_{\text{RNN}}(b) = \log \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l}, \quad (4.2)$$

where f_c and f_τ are the c -jet and τ -jet fractions.

The impact parameter RNN (IPRNN) algorithm yields twice the light flavour rejection and about 1.2 times the c -rejection of the IP3D algorithm.

4.2.2 Secondary vertex finding algorithm

The Secondary Vertex algorithm SV1 [122], aims to reconstruct a single displaced secondary vertex in a jet. All reconstructed tracks within a jet undergo a cleaning procedure to remove tracks originating from photon conversions, hadronic interactions, and

long-lived particles like K_s and Λ . Fake vertices originating from pile-up are rejected by considering only the 25 highest- p_T tracks. This criterion preserves a large fraction of tracks from the b -hadron decays while removing most of the fragmentation tracks originating at high energy.

The reconstruction starts with two-track combinations and runs iteratively over the list of tracks that passed the cleaning procedure, adding to the vertex only the tracks giving a reasonable χ^2 . Vertices with large masses are removed as likely not coming from a b -hadron or a c -hadron decay. Vertices with a secondary track having associated detector hits with radii smaller than the radius of the reconstructed SV are also rejected. With this approach, only a single inclusive secondary vertex can be reconstructed and it could contain decay products from both b - and c -hadron decays.

Eight discriminating variables are constructed and serve as inputs to the high-level taggers: the number of tracks associated with the SV1 vertex (N_{TrkAtVtx}), the number of two-track vertex candidates ($N_{2\text{TrkVtx}}$), the invariant mass of the secondary vertex (m), its energy fraction (f_E) (defined as the total energy of all the tracks associated with the secondary vertex divided by the energy of all the tracks associated with the jet), the two-dimensional (L_{xy}) and three-dimensional (L_{xyz}) distance between the primary and secondary vertices (decay length), the three-dimensional decay length divided by its uncertainty (S_{xyz}), and the ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex. The distributions of some of these variables are reported in Figure 4.5 for b -jets, c -jets and l -jets.

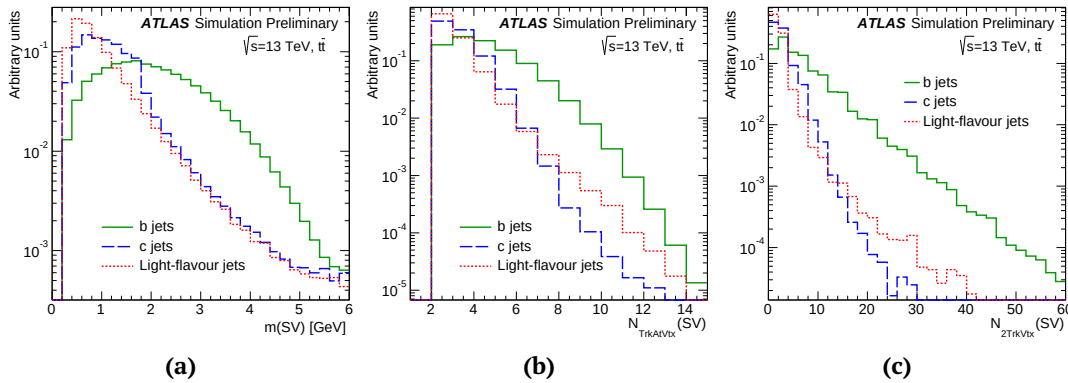


Figure 4.5: Properties of the secondary vertices reconstructed by the SV1 algorithm for b -jets (solid green), c -jets (dashed blue) and l -jets (dotted red) in $t\bar{t}$ events: the invariant mass (a), the number of tracks (b), the number of two-track vertices in the jet (c) [125].

4.2.3 Multi-vertex finding algorithm

The underlying hypothesis of the SV1 algorithm of having a single vertex for both the b -hadron and the c -hadron decays is not always valid: when the distance between the secondary and tertiary decay vertices is larger than the vertex resolution, the tracks from one of the two vertices can be lost in the fit [130]. For this reason, the JETFITTER algorithm [123] was introduced to exploit the topological structure of the b -hadron

and c -hadron decays in order to reconstruct the full b -hadron decay chain. A modified Kalman filter (see [131] for the standard version of the filter) is used to find a common line on which the PV, the b -hadron, and c -hadron decay vertices lie, which corresponds approximately to the b -hadron flying direction. Charged particles originating from either the b -hadron or the c -hadron must intersect this line. With this approach also incomplete topologies with a single track from the b -hadron decay and a single track from the c -hadron decay can be reconstructed.

Eight discriminating variables are constructed and serve as inputs to the high-level taggers: the number of 1-track vertices ($N_{1\text{-trk vertices}}$), the number of two-track vertex candidates ($N_{2\text{TrkVtx}}$), the number of vertices with at least two tracks ($N_{\pm 2\text{-trk vertices}}$), the number of tracks from vertices with at least two tracks (N_{TrkAtVtx}), the average flight length significance of the reconstructed vertices (S_{xyz}), the invariant mass of tracks fitted to one or more displaced vertices (m), the energy fraction (f_E) (defined as the energy from the tracks in the displaced vertices relative to all tracks reconstructed within the jet), and the ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to the displaced vertices. Distributions of some of these variables are reported in Figure 4.6 for b -jets (green), c -jets (blue) and l -jets (red).

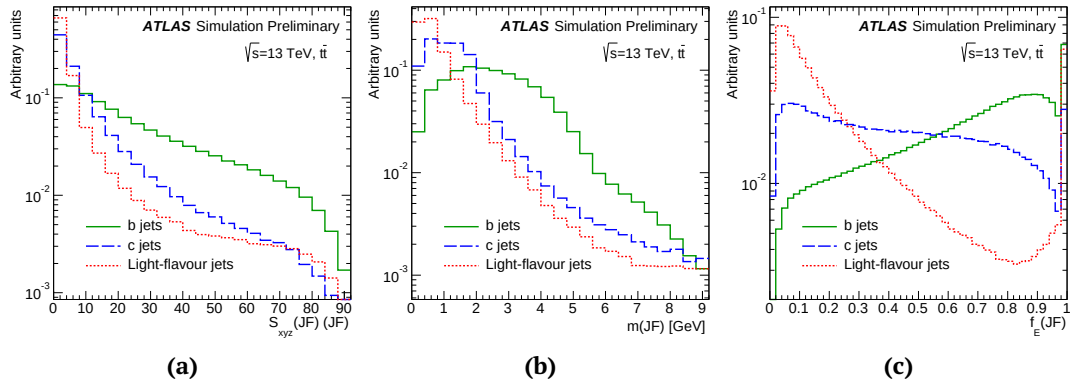


Figure 4.6: Properties of the decay topology reconstructed by the JETFITTER algorithm for b -jets (solid green), c -jets (dashed blue) and l -jets (dotted red) in $t\bar{t}$ events: the average flight length significance of the reconstructed vertices (a), the invariant mass of tracks fitted to one or more displaced vertices (b) and the energy fraction (c) [125].

4.3 High-level tagging algorithms

The high-level taggers combine the inputs from the low-level taggers using a multivariate classifier in order to maximize the b -tagging performance. Three taggers are currently available, the MV2 using a Boosted Decision Tree, and DL1 and DL1r based on deep learning neural networks.

4.3.1 Training samples

The previous versions of the multivariate (MV) high-level taggers used by the ATLAS experiment were trained with $t\bar{t}$ events. However, the spectrum of jets in this sample is steeply falling from 250 GeV making the MV training performance in the medium and high jet p_T region not optimal. To overcome this limitation, a Z' sample with an artificially wider jet p_T spectrum was designed to cope with kinematic regions not accessible by the $t\bar{t}$ sample.

The $t\bar{t}$ sample is simulated with the POWHEG + PYTHIA8 generator and only $t\bar{t}$ decays with at least one lepton from a W boson decay are taken into account. The Z' sample is simulated using the PYTHIA8 generator, where event-by-event reweighting factors are applied in order to broaden the natural width of the Z' , thereby extending the p_T range of jets produced in hadronic decays. The branching fractions of the Z' decays are set to be one third each for $b\bar{b}$, $c\bar{c}$ and light jets. A new *hybrid* sample is produced by including the b -jets from $t\bar{t}$ decays for values of the jet $p_T < 250$ GeV and b -jets from Z' decays for jet $p_T > 250$ GeV.

Figure 4.7 (a) and 4.7 (b) show the jet p_T distribution of the $t\bar{t}$ and the Z' samples for b -jets and l -jets, respectively, while Figure 4.7 (c) shows the b -jet p_T distribution for the final hybrid sample used to train the high-level taggers.

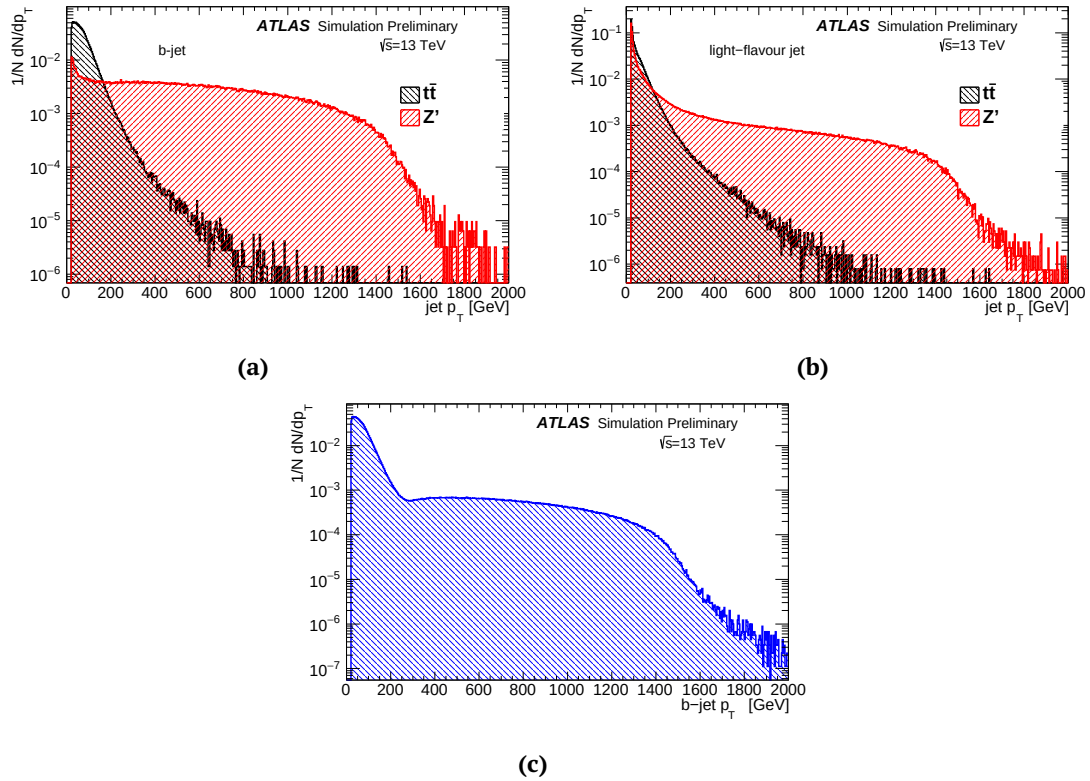


Figure 4.7: Distribution of the jet p_T of the b -jets (a) and l -jets (b) originating from $t\bar{t}$ (red) and Z' (black) decays. The distribution of the b -jet p_T for the hybrid sample is reported in (c) [132].

4.3.2 MV2 discriminant

The MV2 discriminant [132] consists of a Boosted Decision Tree trained on the hybrid $t\bar{t} + Z'$ sample using the ROOT Toolkit for MultiVariate Data Analysis (TMVA) package [133]. Jets stemming from the hadronization of a b -quark are considered as signal while jets produced by c -quarks or l -quarks are treated as background. The outputs of the low-level taggers are combined in a multivariate analysis together with the jet p_T and η to exploit their correlation with the other variables.

To avoid any difference in the kinematic distributions of the b -jets with respect to the c -jets and l -jets being used to discriminate between the jet flavour, the p_T and η distributions of the b -jets are reweighted to match the spectrum of the l -jets ones. In order to enhance the c -jets rejection whilst preserving a high l -jets rejection, the c -jets fraction in the training is set to 7% and the l -jets to 93%. The naming convention of the MV2 algorithm reflects the fraction of c -jets, and consequentially l -jets, present in the training sample. The "MV2cXX" algorithm will have a fraction XX% of c -jets and (100% - XX%) of l -jets. Many configurations have been tested and the MV2c10² one has been found to have the best balance between c -jet and l -jet rejections.

The variables entering the MV2c10 algorithm are listed in Table 4.1 and the output MV2c10 discriminant is shown in Figure 4.8 (a) for b -jets, c -jets and l -jets in $t\bar{t}$ simulated events.

4.3.3 DL1 discriminant

The DL1 tagger [132] is based on a deep feed-forward neural network (NN) trained using Keras [134] with the Theano [135] backend and the Adam optimiser [136]. The input variables used in the training are the same used to train the MV2c10 discriminant with the addition of the c -tagging JETFITTER variables which exploit the specific features of the secondary and tertiary vertices. The full list of input variables entering the training is reported in Table 4.1.

As the jet p_T and η are added as variables in the training, a reweighting procedure similar to the one adopted for the MV2c10 training is performed to match the kinematic distributions of the b -jets to those of the l -jets. The output of the DL1 training corresponds to the probability for a jet to stem from a b -, c - or a light-quark, and can be written as:

$$\text{DL1} = \log \frac{p_b}{p_c \cdot f_c + (1 - f_c) \cdot p_{\text{light}}}, \quad (4.3)$$

where p_b , p_c and p_{light} represent the b -jet, c -jet and l -jet probabilities, respectively, and f_c is the c -jet fraction which can be chosen a posteriori to optimize the performance of the tagger. The output DL1 discriminant is shown in Figure 4.8 (b) for b -jets, c -jets and l -jets in $t\bar{t}$ simulated events.

To enhance the l -jet rejection, the variables used to train IPRNN can be included in the

²The algorithm is not called MV2c07 for historical reasons as it was decided to keep the naming convention from Run 1 where the fraction of c -jet was set to 10%.

Table 4.1: Input variables entering the MV2c10 and the DL1 trainings. The c -tagging JETFITTER variables are used only in the DL1 training [137].

Input	Variable	Description
Kinematics	p_T η	Jet p_T Jet $ \eta $
IP2D/IP3D	$\log(P_b/P_{\text{light}})$ $\log(P_b/P_c)$ $\log(P_c/P_{\text{light}})$	Likelihood ratio between the b -jet and light-flavour jet hypotheses Likelihood ratio between the b - and c -jet hypotheses Likelihood ratio between the c -jet and light-flavour jet hypotheses
SV1	$m(\text{SV})$ $f_E(\text{SV})$ $N_{\text{TrkAtVtx}}(\text{SV})$ $N_{2\text{TrkVtx}}(\text{SV})$ $L_{xy}(\text{SV})$ $L_{xyz}(\text{SV})$ $S_{xyz}(\text{SV})$ $\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass Energy fraction of the tracks associated with the secondary vertex Number of tracks used in the secondary vertex Number of two-track vertex candidates Transverse distance between the primary and secondary vertex Distance between the primary and the secondary vertex Distance between the primary and the secondary vertex divided by its uncertainty ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JETFITTER	$m(\text{JF})$ $f_E(\text{JF})$ $\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$ $S_{xyz}(\text{JF})$ $N_{\text{TrkAtVtx}}(\text{JF})$ $N_{2\text{TrkVtx}}(\text{JF})$ $N_{1\text{-trk vertices}}(\text{JF})$ $N_{\geq 2\text{-trk vertices}}(\text{JF})$	Invariant mass of tracks from displaced vertices Energy fraction of the tracks associated with the displaced vertices ΔR between jet axis and vectorial sum of momenta of all tracks attached to displaced vertices Significance of average distance between PV and displaced vertices Number of tracks from multi-prong displaced vertices Number of two-track vertex candidates (prior to decay chain fit) Number of single-prong displaced vertices Number of multi-prong displaced vertices
JETFITTER c -tagging	$L_{xyz}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $L_{xy}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $m_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $E_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $f_E(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $N_{\text{TrkAtVtx}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$ $Y_{\text{trk}}^{\text{min}}, Y_{\text{trk}}^{\text{max}}, Y_{\text{trk}}^{\text{avg}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Distance of 2^{nd} or 3^{rd} vertex from PV Transverse displacement of the 2^{nd} or 3^{rd} vertex Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex Number of tracks associated with 2^{nd} or 3^{rd} vertex Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex

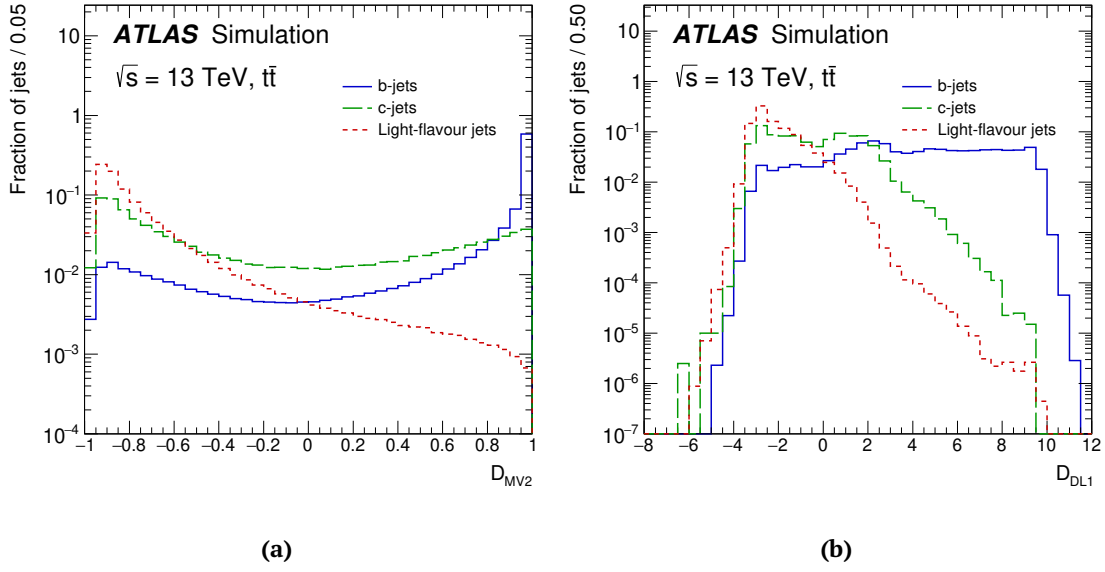


Figure 4.8: Distribution of the MV2c10 (a) and DL1 (b) discriminants for b -jets (solid blue), c -jets (dotted green) and l -jets (dashed red) for $t\bar{t}$ events [137].

DL1 training in order to exploit the spatial and kinematic correlations between tracks originating from the same b -hadron. This approach has shown to improve the tagger performance especially in the high jet p_T regime. The correspondent tagger is called DL1r.

4.4 b -jet tagging efficiency and algorithms performance

The performance of a tagger can be estimated by evaluating its ability to reject the background hypothesis, represented by the c -jets and l -jets components, with respect to the signal hypothesis represented by a jet containing a b -hadron. The efficiency of a tagger is defined as:

$$\varepsilon_j = \frac{\text{Number of jets of flavour } j \text{ passing cuts}}{\text{Number of jets of flavour } j}, \quad (4.4)$$

where the flavour j is assigned based on the quark flavour: a jet is labelled as a b -jet if a b -quark with $p_T > 5$ GeV is found in a cone of size $\Delta R = 0.3$ around the jet direction. If no b -quark is found, the labelling hypotheses of c -jet and τ -jet are tried sequentially and, if none is satisfied, the jet is labelled as a l -jet.

The rejection of a tagger is defined as $1/\varepsilon_j$ and is evaluated at specific values of the efficiency called “working points” (WPs). These are based on a fixed selection requirement on the b -tagging algorithm distribution ensuring a b -tagging efficiency ε_b and are evaluated on $t\bar{t}$ events. A tagger with a higher c -jet and l -jet rejection rate at a specific WP with respect to another has a better performance. Table 4.2 shows the value of the cut on both the MV2c10 and DL1 discriminants which gives a certain efficiency ε_b and the corresponding rejection of the taggers against c -jets, τ -jets and l -jets.

Table 4.2: Fixed cuts defining the efficiency working points and the related c -jet, τ -jet and l -jet rejections for the MV2c10 and DL1 taggers, evaluated on $t\bar{t}$ events [137].

ϵ_b	MV2				DL1			
	Selection	Rejection			Selection	Rejection		
		c -jet	τ -jet	Light-flavour jet		c -jet	τ -jet	Light-flavour jet
60%	> 0.94	23	140	1200	> 2.74	27	220	1300
70%	> 0.83	8.9	36	300	> 2.02	9.4	43	390
77%	> 0.64	4.9	15	110	> 1.45	4.9	14	130
85%	> 0.11	2.7	6.1	25	> 0.46	2.6	3.9	29

The MV2c10 and DL1 distributions can be divided into five “pseudo-continuous” bins whose edges are defined by the efficiency working points 85%, 77%, 70% and 60% listed in Table 4.2. This binned distribution, also called PCBT from “pseudo-continuous b -tagging”, is much easier to calibrate with respect to the full continuous output of the b -tagging algorithm and can be used by physics analyses as a powerful discriminant against backgrounds containing non b -jets, as explained in Chapter 6. The shape of the PCBT distribution is reported in Figure 4.9, separately for b -jets, c -jets and l -jets.

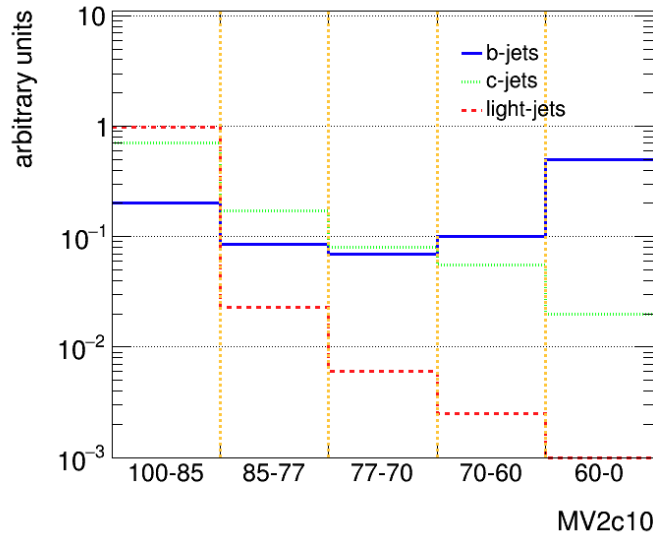


Figure 4.9: Shape of the binned PCBT distribution for b -jets (solid blue), c -jets (dotted green) and l -jets (dashed red) for the MV2c10 tagger. (Courtesy of Ilaria Luise, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE)).

The c -jet and l -jet rejections as a function of the b -tagging efficiency on $t\bar{t}$ events are shown in Figure 4.10 for both the low-level and high-level taggers. It can be observed that combining the information from the low-level taggers into a multivariate analysis improves the rejection of the c -jets and l -jets components.

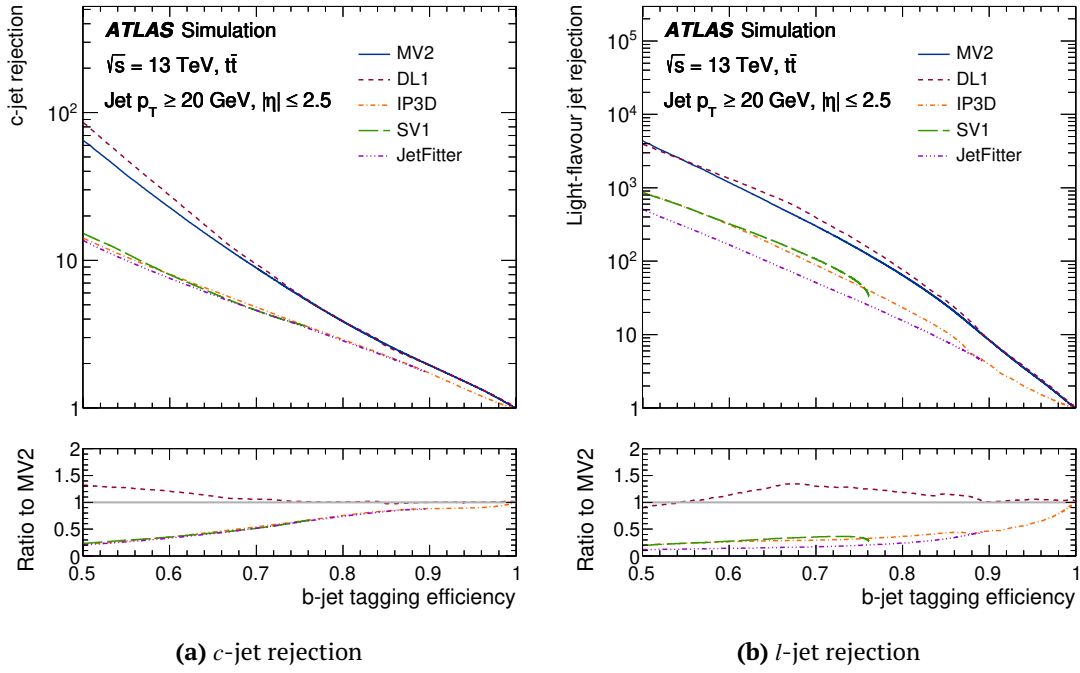


Figure 4.10: c -jet (a) and l -jet (b) rejections as a function of the b -tagging efficiency for the low-level (IP3D, SV1, JetFitter) and high-level (MV2c10, DL1) taggers evaluated on $t\bar{t}$ events [137].

4.5 Truth tagging procedure

If the signal is characterized by the presence of b -jets, as in the case of the $VH, H \rightarrow b\bar{b}$ analysis described in Chapter 6, the rejection capability of the b -tagging algorithms can result in low c -jet and l -jet population during the training. One way to overcome this limitation and reduce the statistical fluctuations in these samples is to avoid removing events failing the b -tagging requirements, and instead apply a weight to each MC event, based on the desired number of b -jets in the event. In practice, this is done by giving events with real b -jets in MC a much higher weight than events having only c -jets or l -jets. This procedure is referred to as “truth tagging” since it makes use of the truth flavour label of the jet to determine the event weight.

If the number of jets m in the event exceeds the number of tagged jets n required by the analysis, a combinatorial approach is used in order to consider all the possible permutations of tagged and non-tagged jets. In this case, the total truth tag weight of the event is defined as the sum of the truth tag weights of each single permutation. Taking for example an event with three jets, if two b -tagged jets are required and each jet has a b -tagging efficiency ε_i , the total truth tagging weight is given by:

$$w_{\text{tot}} = \varepsilon_1 \varepsilon_2 (1 - \varepsilon_3) + \varepsilon_1 (1 - \varepsilon_2) \varepsilon_3 + \varepsilon_2 \varepsilon_3 (1 - \varepsilon_1). \quad (4.5)$$

The b -tagging efficiencies $\varepsilon_{1,2,3}$ are obtained from 2-dimensional maps as a function of the jet p_T , η and the truth flavour label and are derived from simulation for a specific

tagger and a given working point.

In some cases, it is necessary to know which jets are b -tagged among the selected ones, for example to apply specific analysis selections on top of the b -tagging requirement. In these cases, a permutation is chosen randomly among all possible ones, with a probability proportional to its truth tagging weight. In the example of the event with three jets, the probability of choosing the permutation in which the first two jets are tagged is:

$$P_{12} = \frac{\varepsilon_1 \varepsilon_2 (1 - \varepsilon_3)}{w_{\text{tot}}}. \quad (4.6)$$

The higher the weight of the permutation, the higher the probability of choosing it.

Generalizing to the case of an event with m jets and n tagged jets required, the possible permutations are $\binom{m}{n}$. For a given permutation $\binom{m}{n}_i$, the total number of remaining permutations is $\overline{\binom{m}{n}} = \binom{m}{n} - \binom{m}{n}_i$, and the efficiency of all tagged jets and inefficiency of all non-tagged jets are:

$$\varepsilon\left(\binom{m}{n}_i, X\right) = \prod_{j \in n} \varepsilon_X^f(j), \quad (4.7)$$

$$\varepsilon_{\text{in}}\left(\overline{\binom{m}{n}}_i, X\right) = \prod_{j \in m-n} (1 - \varepsilon_X^f(j)), \quad (4.8)$$

where $\varepsilon_X^f(j)$ is the tagging efficiency of the jet with flavour f at the given working point X , derived from the 2-dimensional map in jet p_T and η corresponding to the flavour f . Therefore, the final truth tagging weight to be assigned to the event is:

$$w_{TT}(X) = \sum_i^{|\binom{m}{n}|} \varepsilon\left(\binom{m}{n}_i, X\right) \varepsilon_{\text{in}}\left(\overline{\binom{m}{n}}_i, X\right), \quad (4.9)$$

and the probability of choosing a particular permutation $\binom{m}{n}_i$ is given by:

$$P_i(X) = \frac{\varepsilon\left(\binom{m}{n}_i, X\right) \cdot \varepsilon_{\text{in}}\left(\overline{\binom{m}{n}}_i, X\right)}{w_{TT}(X)}. \quad (4.10)$$

The above formulation for the truth tagging implies that the definition of a tagged or non-tagged jet is done *a priori*, choosing randomly among all possible combinations. However, it can happen that the real b -tagging score of the jets is in contrast with the chosen permutation, meaning that the jets defined as tagged by the truth tagging procedure, have instead a very low b -tagging score. For example, in the case of the three jets event, assuming the jets having MV2c10 scores equal to 0.20, 0.60 and 0.98, respectively, if the truth tagging procedure defines the first two jets as tagged jets at 70% WP, the real scores are not reflecting this choice (as the fixed cut corresponding to the 70% WP of the MV2c10 discriminant is equal to 0.83).

This results in a serious limitation of the procedure, especially when the value of the b -tagging discriminant for each jet is used as variable in the BDT of a physics analysis (as

in the case of the $VH, H \rightarrow b\bar{b}$ analysis in Chapter 6). Therefore, the b -tagging score of the jets are re-calculated according to the chosen permutation. In the three jets case, the first two jets should have a MV2c10 score above the 70% WP requirement, therefore a flat random number is generated between the [70%,60%] and >60% bins, taking into account the non flat shape of the MV2c10 distribution (see Figure 4.9). The same procedure is used to generate the new b -tagging score for the third non-tagged jet in the efficiency bins <70%.

4.6 Calibration of b -tagging efficiency

The efficiency of a b -tagging algorithm is evaluated on $t\bar{t}$ samples and corrected by scale factors (SF) to match the performances in data. The SF is evaluated per jet flavour, working point, and as a function of the jet p_T and η . The SF for a b -tagged jet is defined as:

$$\kappa_j = \frac{\varepsilon_j^{\text{data}}}{\varepsilon_j^{\text{MC}}}, \quad (4.11)$$

where $\varepsilon_j^{\text{data}}$ is the b -tagging efficiency for the jet flavour j extracted from data, and $\varepsilon_j^{\text{MC}}$ is the one evaluated on Monte Carlo simulated events. If the event contains non b -tagged jets, an inefficiency SF is defined as:

$$\nu_j = \frac{1 - \varepsilon_j^{\text{data}}}{1 - \varepsilon_j^{\text{MC}}} = \frac{1 - \kappa_j \cdot \varepsilon_j^{\text{MC}}}{1 - \varepsilon_j^{\text{MC}}}, \quad (4.12)$$

and is applied to the event weight to preserve the total number of events. Furthermore, the Monte Carlo generators can differ in the shower and hadron decay modelling, leading to different efficiencies for different MCs. Therefore, an additional multiplicative correction is applied to the event weight and defined as:

$$\kappa_j = \frac{\varepsilon_j^{\text{data}}}{\varepsilon_{j,\text{ref}}^{\text{MC}}} \frac{\varepsilon_{j,\text{ref}}^{\text{MC}}}{\varepsilon_{j,\text{alt}}^{\text{MC}}}, \quad (4.13)$$

where $\varepsilon_{j,\text{ref}}^{\text{MC}}$ is the efficiency evaluated on the MC generator taken as reference, POWHEG+PYTHIA8+EvtGen in the case of the $t\bar{t}$ sample, while $\varepsilon_{j,\text{alt}}^{\text{MC}}$ is the efficiency evaluated on the alternative generator used in the analysis.

The energy scale and resolution of hadronic jets, the modelling of the $t\bar{t}$ and single-top backgrounds and the data and MC statistics (only in the region $p_T < 40$ GeV) are found to be the main sources of uncertainties in the measurement of the b -tagging efficiency.

Figure 4.11 (a) shows the measured b -tagging efficiency for the 70% working point as a function of the jet p_T for the MV2c10 tagger. The red line corresponds to the efficiency evaluated in $t\bar{t}$ events passing the signal region cuts. The corresponding scale factors are presented in Figure 4.11 (b). The total uncertainty reaches 1% in the $40 \leq p_T < 250$ GeV

range and it is dominated by the modelling of the $t\bar{t}$ backgrounds and the data statistics. Due to the higher impact of the jet energy scale uncertainty in the region $20 \leq p_T < 40$ GeV, the uncertainties increases up to 8%, while for $p_T > 250$ GeV the lack of data statistics gives a total uncertainty of 3%.

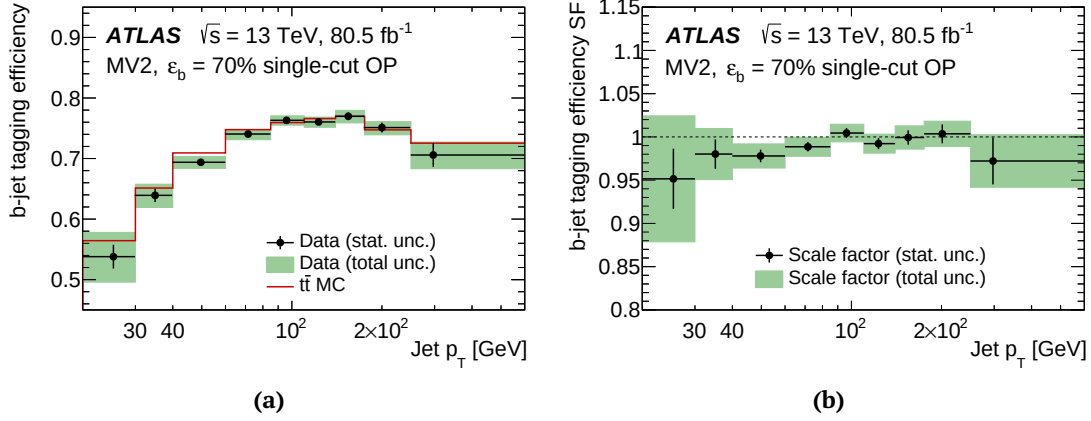


Figure 4.11: b -tagging efficiency as a function of jet p_T (a) and corresponding scale factors (b) for $\epsilon_{WP} = 70\%$. The red line corresponds to the efficiency measured in simulated $t\bar{t}$ events. The error bars show the data statistical uncertainty while the size of green boxes are obtained summing in quadrature the statistical and systematic uncertainties [137].

The b -tagging efficiencies resulting from the log-likelihood fit for the five pseudo-continuous working points are reported in Figure 4.12 for the MV2c10 tagger for the $110 \leq p_T < 140$ GeV range where the b -tagging efficiency reaches its maximum.

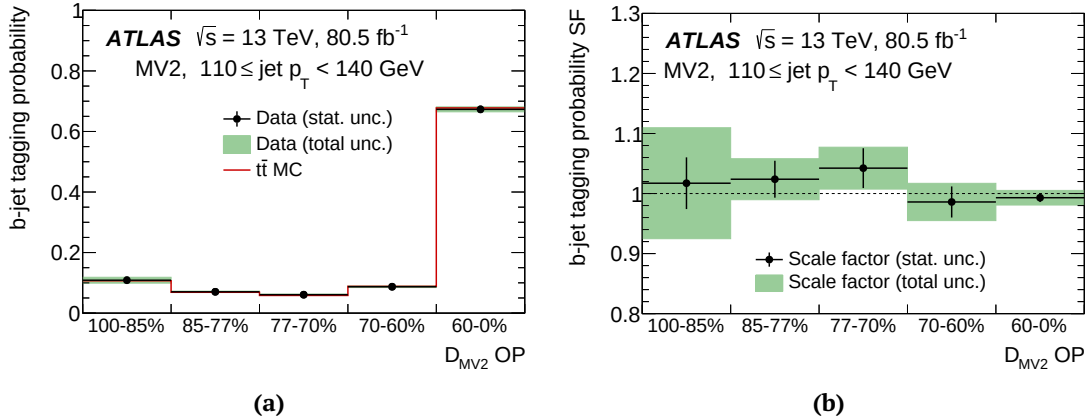


Figure 4.12: b -tagging probabilities (a) and the corresponding scale factors (b) measured in the five pseudo-continuous bins for the MV2C10 tagger for jets with $110 \leq p_T < 140$ GeV. The red line corresponds to the efficiency measured in simulated $t\bar{t}$ events passing the signal region cuts. The error bars show the data statistical uncertainty while the size of green boxes are obtained summing in quadrature the statistical and systematic uncertainties [137].

The misidentification of the c -jets as b -jets plays an important role in many ATLAS analyses, therefore it is convenient to derive a c -tagging efficiency or c -jet mis-tag rate [138]. A semileptonic $t\bar{t}$ sample with a top quark decaying hadronically and the other leptonically is used to estimate the c -tagging efficiency. The choice of this sample is

justified by the fact that 50% of the times, as predicted by the SM, the W boson produced in the top decay, will decay into a c -quark and in most cases a s -quark. A Kinematic Likelihood Fitter (KLFitter) [139] is used to correctly assign the jets to the partonic final state of the $t\bar{t}$ process. The corresponding SFs go from ~ 1.1 to ~ 1.6 from the 85% to the 60% working point while the total uncertainty increases from 6% to 20% as the working point becomes tighter. The scale factor for the 70% working point is shown in Figure 4.13 (a) as a function of the jet p_T , as an example.

The l -jets mis-tag rate is evaluated using the *negative tag method* [140] which is mainly based on the assumption that the l -jets are mis-tagged as b -jets due to the finite resolution of the reconstructed inner detector tracks and their impact parameter. Under this assumption, the impact parameter distribution of l -jets is expected to be symmetric around zero, while the tracks from b -jets and c -jets exhibit a longer tail at high positive values, due to the higher lifetimes (see Figure 4.2). Therefore, the tag rate of l -jets obtained when running the IPXD algorithms after inverting the criterion on the sign of the impact parameter, is expected to be a good approximation of the mis-tag rate of l -jets for these algorithms. A similar feature is expected also in some distributions related to the SV1 algorithm as the signed decay length significance or the secondary vertex mass. However, l -jets can be mis-tagged also due to the presence of tracks from displaced topologies, as the decay of long-lived particles, and material interactions from photon conversions and hadronic interactions. These contributions introduce a bias towards positive values of the impact parameter distribution of l -jets tracks and are taken into account by applying correction factors derived from MC.

The negative tag method use variables with inverted signs as input to the MV2c10 algorithm. The scale factors are derived by comparing the mis-tag rate in data and simulation and are found to be stable around 2 across the full jet p_T spectrum, as shown in Figure 4.13b for the 70% WP.

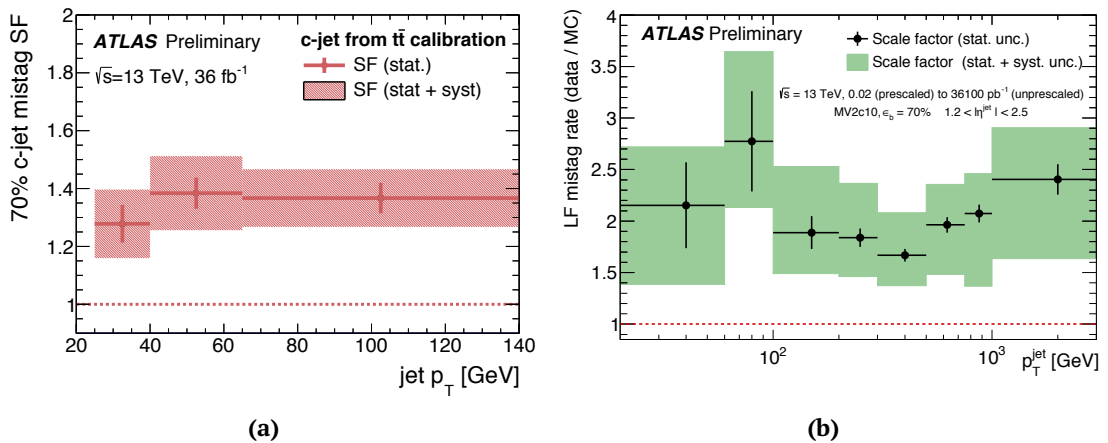


Figure 4.13: Scale factors for the c -jet (a) [138] and l -jet (b) [140] mis-tag rate for the 70% b -tagging working point and as a function of the jet p_T .

4.7 Extrapolation to high- p_T jets

Given the low number of b -jets in $t\bar{t}$ events from $p_T > 400$ GeV, an uncertainty is assigned in the high- p_T region in order to extend the fixed cut WP calibration to the full p_T range explored by many ATLAS analyses. Many sources of uncertainties affect the b -tagging performances at high- p_T , involving the reconstruction of the tracks, the modelling of the b -hadrons and their interactions with the pixel detector. Among the uncertainties related to the tracks, there are the track impact parameter resolution, the fraction of fake tracks, the description of the detector material, and the track multiplicities per jet. The effect of the parton shower simulation and the b -quark fragmentation function are evaluated by changing the Monte Carlo generator to simulate the dilepton $t\bar{t}$ events used in the calibration. At last, in the standard ATLAS MC simulation, interactions with the detector are simulated only for the b - and c -hadron decay products and not for the b - and c -hadron themselves. High momentum b -hadrons are very likely to decay after the first layer of the pixel detector, interacting with the material before decaying. The hit left by the b -hadron can be mistakenly associated to the track of one of the b -hadron decay products bringing differences in the b -tagging efficiency. The size of the uncertainty resulting from this effect was derived on a Z' sample with a flat spectrum in p_T up to 5 TeV to enhance the number of b -hadrons decaying after the innermost layer of the pixel detector. Details on the derivation of the uncertainties accounting for this effect are given in Chapter 5.

The SF measured in the highest p_T bin is extrapolated for $p_T \geq 400$ GeV. The uncertainty is calculated as the sum in quadrature of all the effects described above added in quadrature to the data-based uncertainty in the last bin (250-400 GeV). Figure 4.14 shows the SF obtained for an efficiency WP of 70% for the MV2C10 algorithm.

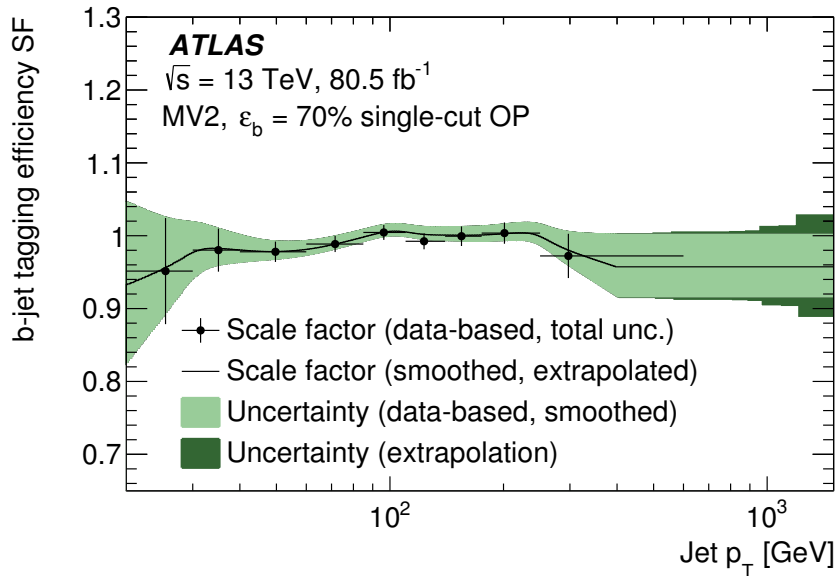


Figure 4.14: b -tagging efficiency SF for the $\epsilon = 70\%$ WP of the MV2C10 algorithm including the extrapolation uncertainty for $p_T \geq 400$ GeV [137].

5.1 The ATLAS Simulation Infrastructure

In order to study a wide range of physics processes, a detailed simulation has been developed to emulate how particles interact with the detector [15]. The simulation of a process can be summarized into four main steps: the generation of the event and the immediate decay of the particles involved, the simulation of the interactions of the stable particles, as defined by the event generator, with the detector material, the digitization of the energy deposited in the sensitive regions of the detector into electronic signals, and the reconstruction of the physics objects which produced the signals inside the detector.

In the *event generation*, a Monte Carlo generator is employed to simulate all phases of the pp scattering, as explained in Section 1.4.2: the hard scattering, the parton shower, the hadronization and the underlying event. In the hadronization step, unstable particles are produced which subsequently decay into stable particles that can interact with the detector. Particles with $c\tau > 10$ mm, with τ the decay time, are considered stable by the event generator, therefore their decays and interactions with the detector are handled by the simulation. On the other hand, particles with $c\tau < 10$ mm are decayed by the event generator, and their possible interactions with the detector material, as well as their curvature inside the magnetic field, are ignored. All stable particles produced by the event generator are referred to as *particle-level* objects.

Particle-level objects are passed to the detector simulation which is based on the GEANT4 software [16]. The complete detector geometry, including misalignments or distortions, is built from databases containing all the information about the detector performances and conditions of a real run of data-taking. The output of the simulation is a list of “hits” representing the energy depositions, positions, and times of each simulated particle.

After the simulation step by GEANT4, the ATLAS *digitization* software takes the hits from the simulated events, such as hard scattering signal, minimum bias, beam halo, beam gas, and cavern background events, and overlays them according to a given pile-up rate. Then, it produces “digits” in case the voltage of a particular channel of the readout system rises above a certain threshold and within a given time window. At this stage, detector noise is also added to the event.

The last step of the ATLAS simulation is the *reconstruction*, which combines the information from the different sub-detectors to reconstruct the actual physics object that produced the signals, as explained in Chapter 3.

5.2 Interactions of b - and c -hadrons with the pixel detector

Particles like b - and c -hadrons have $c\tau < 10$ mm, therefore, their decays are handled by the event generator. However, for high energy b -(c -) hadrons the decay time τ transforms according to the Lorentz transformations as $\tau \rightarrow \tau\gamma$, with γ the Lorentz boost and, in the same way, the decay length transforms as $l = \beta\gamma c\tau$.

For a b -hadron of $p_T = 800$ GeV, the Lorentz boost is $\gamma \sim 150$, and given the decay time of $\tau = 1.6 \mu\text{s}$, the decay length is $l \sim 80$ mm. This means that the b -hadron is very likely to traverse the IBL and the BL before decaying, but its interactions with the material are not simulated as its decay is handled by the event generator and not by the simulation. The most probable decay of a b -hadron is into a c -hadron via the $b \rightarrow cW^{*-}$ transition. With the same approach as before, assuming a c -hadron, e.g. D^+ , of $p_T = 800$ GeV, and given its decay time of $\tau = 1.04 \mu\text{s}$, the D^+ decay length is $l \sim 134$ mm. This means that when a b -hadron decays into a c -hadron, a D^+ in this example, this can potentially traverse all layers of the pixel detector before decaying¹.

Figure 5.1 shows the fraction of b -hadrons traversing the IBL or the BL of the pixel detector as a function of the b -labelled jet p_T . The b -hadron is assumed to traverse the IBL (BL) if its transverse decay length, L_{xy} , defined as the difference in the $x-y$ plane between the b -hadron's decay and production vertices, is larger than the maximum radius of the IBL (BL), as shown Figure 2.6 (a). This value has been chosen in order to ensure that the b -hadron has actually passed the layer for all values of the azimuthal angle ϕ . It can be observed that, already at low b -jet p_T , a non-negligible fraction of b -hadrons traverses the IBL, or even the BL, before decaying.

A b -(c -)hadron traversing one or more layers of the pixel detector can interact with the material before decaying. The interaction can be electromagnetic, in which the heavy flavour loses some energy in the silicon, or it can be an elastic or inelastic hadron-nucleus interaction, where in the latter new particles are produced and/or the internal structure of the hadron or the nucleus is modified.

The current version of the GEANT4 software used in the ATLAS simulation does not take into account the direct interaction of b - and c -hadrons with the detector material, but only for their decay products. Therefore, a new and more accurate simulation, defined as “*quasi-stable*” simulation in the following, has been developed in order to include them. So far, only electromagnetic interactions have been implemented, as heavy flavour hadronic interactions models are very complex and are still under investigations. Furthermore, this new simulation allows the trajectory of both b - and c -hadrons to be bent in the magnetic field. However, the angular deflection in the transverse plane has been estimated to be $\sim 10^{-4}$ rad, corresponding to $2 \mu\text{m}$ on the IBL, which is much smaller than the pixel size. Therefore little impact is expected from the magnetic field in the following analysis.

Many studies have been performed to validate the quasi-stable simulation, in order to

¹The radius of the outermost pixel layer is 122.5 mm.

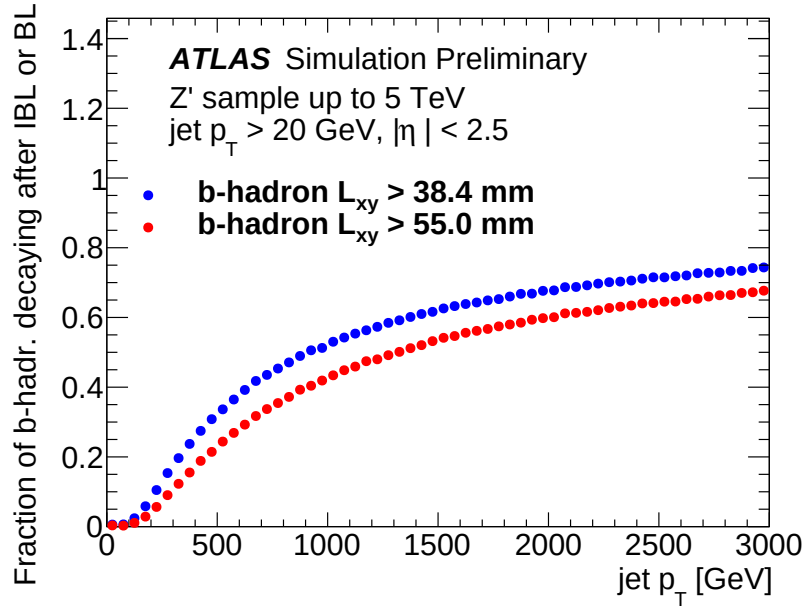


Figure 5.1: Fraction of b -hadrons decaying after the first (blue dots) or the second (red dots) layer of the pixel detector. Results are presented for a simulated dataset of $Z' \rightarrow b\bar{b}, c\bar{c}, l\bar{l}$ and for anti- k_T $\Delta R = 0.4$ particle flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$ [141].

make it available as the default ATLAS simulation starting from the Run 3 data-taking. In particular, the studies presented in this chapter have been carried out to evaluate the change on both the low-level and high-level b -tagging algorithms performance induced by considering the b - and c -hadron interactions with the pixel detector.

In case of an electromagnetic interaction, the hit left by the b -(c)-hadron can be mistakenly associated to the track of one of its decay products, resulting in mismeasured track parameters, as shown in Figure 5.2 in the case of a b -hadron traversing the IBL and decaying before the BL.

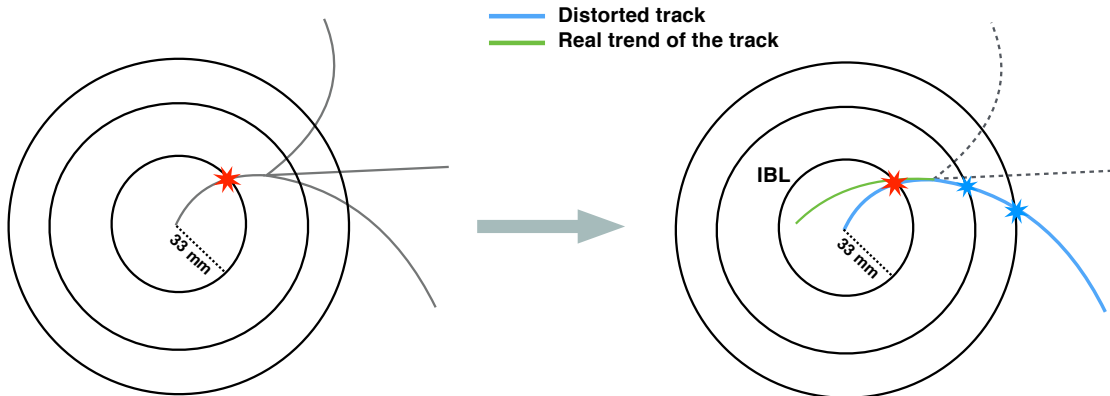


Figure 5.2: Schematic view of a b -hadron traversing the first layer of the pixel detector and leaving a hit in the material before decaying (red star). The hit is mistakenly associated to one of the decay products of the b -hadron resulting in a distortion of the track.

This effect can have a direct impact on the efficiency of tagging a b -jet or a c -jet. The

most correct thing to do would be to derive an additional multiplicative correction to be applied to the efficiency scale factors (see Equation 4.13):

$$\kappa_j = \frac{\varepsilon_j^{\text{data}} \varepsilon_{j,\text{ref}}^{\text{MC}}}{\varepsilon_{j,\text{ref}}^{\text{MC}} \varepsilon_{j,\text{alt}}^{\text{MC}}} \cdot \frac{\varepsilon_{j,\text{QS}}^{\text{MC}}}{\varepsilon_{j,\text{ST}}^{\text{MC}}}, \quad (5.1)$$

where $\varepsilon_{j,\text{ST}}^{\text{MC}}$ is the efficiency derived on a sample generated with no b - or c -hadron interactions (ST stands for “standard”), while $\varepsilon_{j,\text{QS}}^{\text{MC}}$ is the efficiency derived on a sample generated including these interactions (QS stands for “quasi-stable”). However, due to the lack of time, it was decided to simply derive an additional uncertainty taking into account the observed differences in the tagging efficiencies. More details will be given in Section 5.2.3.

The following analysis is focused on studying the impact of the b - and c -hadron material interactions on the b -tagging efficiency, therefore, not only the hits left by the b -hadron can be mistakenly associated to a child track, but also those from the interaction of the c -hadrons coming from the b -hadron decays. In this regard, studying the c -tagging efficiency would also be very interesting as it offers a much clearer topology in which only the hit left by the c -hadron can be mistakenly associated to the track of one of its decay products. This study is already in plan for the next iteration of this analysis.

To have a sufficiently large number of b -hadrons decaying after the IBL, the following studies have been performed using a $Z' \rightarrow b\bar{b}, c\bar{c}, l\bar{l}$ sample with an almost flat spectrum in p_T up to 5 TeV, as shown in blue in Figure 5.3. This sample will be referred to as “extended Z' sample” in the following.

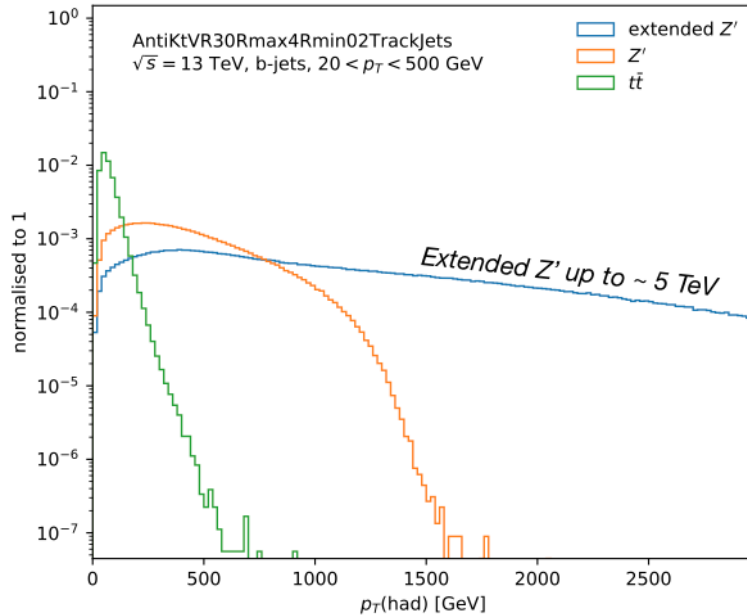


Figure 5.3: Comparison among the $t\bar{t}$ sample (green line) with a p_T spectrum up to 500 GeV, and the Z' (orange line) and extended Z' (blue line) samples with p_T spectrums up to 1.5 TeV and 5 TeV.

Eight million $Z' \rightarrow b\bar{b}, c\bar{c}, l\bar{l}$ events were generated using PYTHIA8, which were then passed to GEANT4, once with and once without the direct interaction of b - and c -hadrons with the pixel detector. The former sample is referred to as “quasi-stable”, and the latter as “standard”. These two samples were then passed to the ATLAS reconstruction in order to build the final physics objects.

The following results are presented for anti- k_T $\Delta R = 0.4$ particle flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$.

5.2.1 Impact on the b -tagging related variables

First, a validation of the quasi-stable sample with respect to the standard one is performed by comparing the b -hadron kinematic properties such as the energy E , the transverse momentum p_T , the values of η and ϕ , and the decay length in the transverse plane, L_{xy} , between the two simulations. These variables are defined at particle-level, therefore, no differences are expected between the standard and the quasi-stable samples. As demonstrated in Figure 5.4, the distributions of E , p_T , η and ϕ for the standard and the quasi-stable samples are compatible within the statistical uncertainties.

However, larger values of the transverse decay length are observed in the case of the quasi-stable simulation. The L_{xy} distributions for the charged and neutral b -hadrons are shown in Figure 5.5 (a) and (b), respectively. Whereas the differences between the two samples agree within the uncertainties in the case of charged b -hadrons, a $\sim 8\%$ increase in the quasi-stable sample is observed for large lifetimes of neutral b -hadrons.

An event by event comparison between the two samples, restricted only to neutral b -hadrons, demonstrated that the larger values of L_{xy} in the quasi-stable simulation came from events with a $B^0 \rightarrow \bar{B}^0$ ($\bar{B}^0 \rightarrow B^0$) oscillation. This issue has not been fully understood yet and is currently under investigation, however, it is not expected to affect the estimation of the additional extrapolation uncertainty on the b -tagging efficiency at high- p_T as only neutral b -hadrons are affected. A more detailed explanation can be found in Appendix A.2.

Figure 5.6 shows the effect of considering the b -hadron interactions with the pixel detector on the track reconstruction efficiency as a function of the b -hadron transverse decay length. For both the standard and the quasi-stable simulation, the further away the b -hadron decays from the primary vertex, the fewer number of clusters available for the track reconstruction, causing a drop in the associated efficiency at each pixel layer. Furthermore, the closer the b -hadron decays to a pixel layer, the smaller the spatial separation of the hits left by its decay products, resulting in a reduced cluster assignment efficiency as shown in Figure 3.2. In addition, in the case of the quasi-stable simulation charged b -hadrons typically leave a hit in each layer they traverse before decaying. This hit can be mistakenly associated to the track of a child and its subsequent distortion might give a bad χ^2 of the track fit, resulting in a possible rejection of the track.

The effect of having an additional hit in the quasi-stable sample is evident in Figure 5.7. The average number of IBL hits per track associated to the b -hadron decay products,

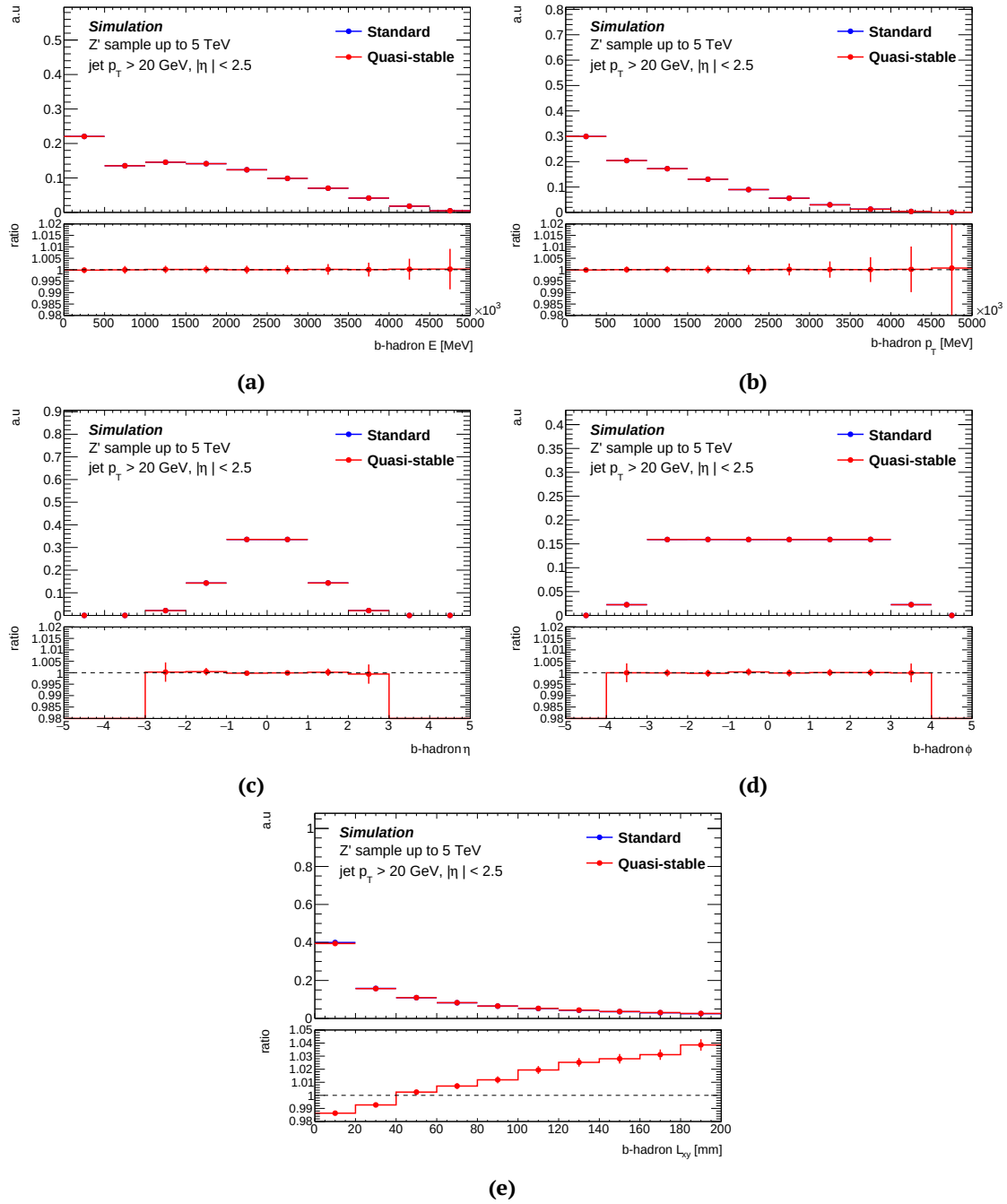


Figure 5.4: Truth value of the b -hadrons' energy E (a), p_T (b), η (c), ϕ (d) and transverse decay length L_{xy} (e). All distributions are normalised to one, therefore, they are shown in arbitrary units a.u. The bottom panel shows the ratio between the quasi-stable and the standard distributions. As expected, no differences are observed in the E , p_T , η and ϕ distributions. However, larger values of the b -hadrons transverse decay length are observed in the case of the quasi-stable simulation.

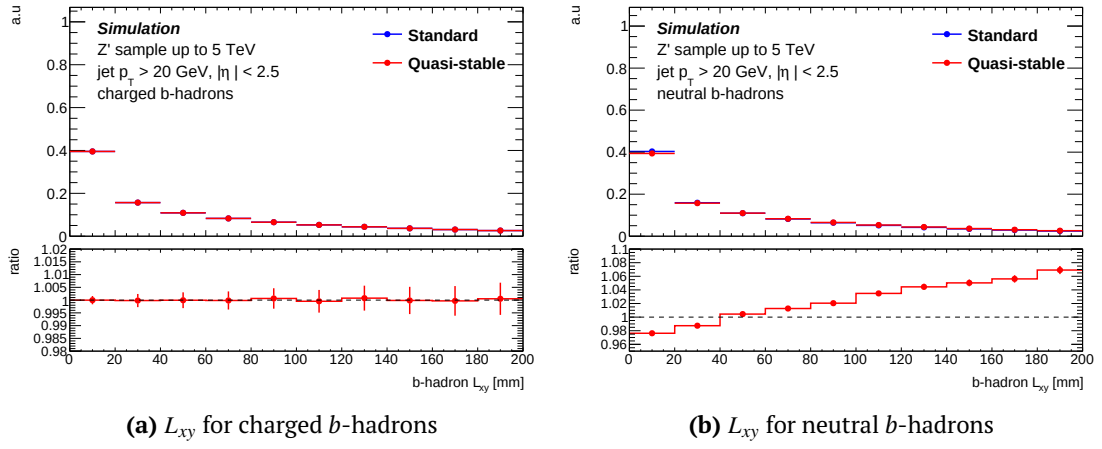


Figure 5.5: Truth value of the transverse decay length L_{xy} in the case of charged (a) and neutral (b) b -hadrons. Both distributions are normalised to one, therefore, they are shown in arbitrary units a.u. The bottom panel shows the ratio between the quasi-stable and the standard distributions. The differences between the standard and the quasi-stable simulations are within the MC statistical uncertainty for the charged b -hadrons, while a $\sim 8\%$ difference at high values of L_{xy} is observed for neutral b -hadrons.

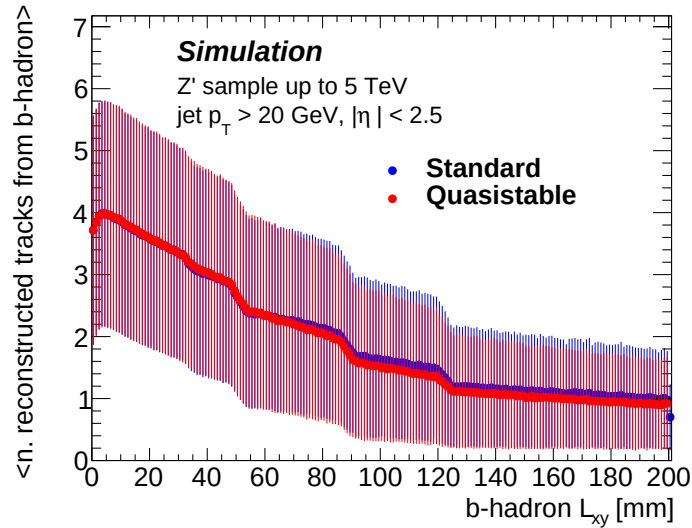


Figure 5.6: Average number of reconstructed tracks from b -hadrons decays as a function of its transverse decay length for the standard (blue dots) and the quasi-stable (red dots) simulations. The band around each contribution represents the RMS of the distribution for each bin. It can be observed that the further the b -hadron decays, the worst is the efficiency in reconstructing the tracks from its decay products.

$\frac{\langle n_{IBLhits} \rangle}{track}$, is shown as a function of the transverse decay length of the *b*-hadron L_{xy} , for the standard and the quasi-stable samples. Figure 5.7 (a) shows the distribution when considering all *b*-hadrons, while Figures 5.7 (b) and (c) show only charged or neutral *b*-hadrons, respectively. The position in the $x-y$ plane of the different layers of the pixel detector is indicated. Some observations can be made:

- the average number of IBL hits per track for the standard simulation is not zero when the *b*-hadron decays between two pixel layers (as one would expect); in such a dense and boosted topology, a hit left by a fragmentation track² can be wrongly assigned to a *b*-hadron decay product. This is related to the drop in the cluster assignment efficiency at the IBL shown in Figure 3.2;
- values larger than 1 are due to the fact that the IBL modules are partly overlapping (see Figure 2.6 for the layout of the IBL), therefore, a track can cross two IBL modules;
- for $L_{xy} < 33.25$ mm, the *b*-hadron decays before the IBL but an increase in the average number of hits in the IBL is observed for the quasi-stable sample for both charged and neutral *b*-hadrons. This is because a *b*-hadron can decay to a charged *c*-hadron before the IBL, that can interact with the IBL material leaving a hit before decaying. The size of this contribution is larger in the case of neutral *b*-hadrons than for charged *b*-hadrons. This is due to the fact that a charged *b*-hadron is more likely to decay to a neutral *c*-hadron, while a neutral *b*-hadron is more likely to decay to a charged *c*-hadron. Also, it has to be noticed that Figure 5.7 (b) for charged *b*-hadrons has a very different scale than Figure 5.7 (c) for neutral *b*-hadrons;
- for $L_{xy} > 33.25$ mm the increase in the number of hits in the IBL is present only for charged *b*-hadrons, as expected;
- for $33.25 < L_{xy} < 88.50$ mm, the *b*-hadron traverses the IBL (and possibly the second and third layers of the pixel detector) leaving a hit before decaying. The hit can be mistakenly associated to the track of one of the *b*-hadron decay products enhancing the distribution for the quasi-stable simulation (see Figure 5.2). It can be observed that the resulting mis-association rate is very large given that there are on average several *b*-hadron decay tracks which can pick-up this hit;
- for $L_{xy} > 88.50$ mm, the mis-association of a IBL hit to the track of a decay product from a *b*-hadron that decays far from the IBL is more reduced.

Similar conclusions can be drawn when looking at the average number of hits in the B-layer and the pixel hits per track associated to the *b*-hadrons decay products as a function of the *b*-hadron L_{xy} , shown in Figure 5.8 and 5.9, respectively.

²A fragmentation track is a track originating from the hadronization of the *b*-quark whose measured IP is compatible with the primary vertex.

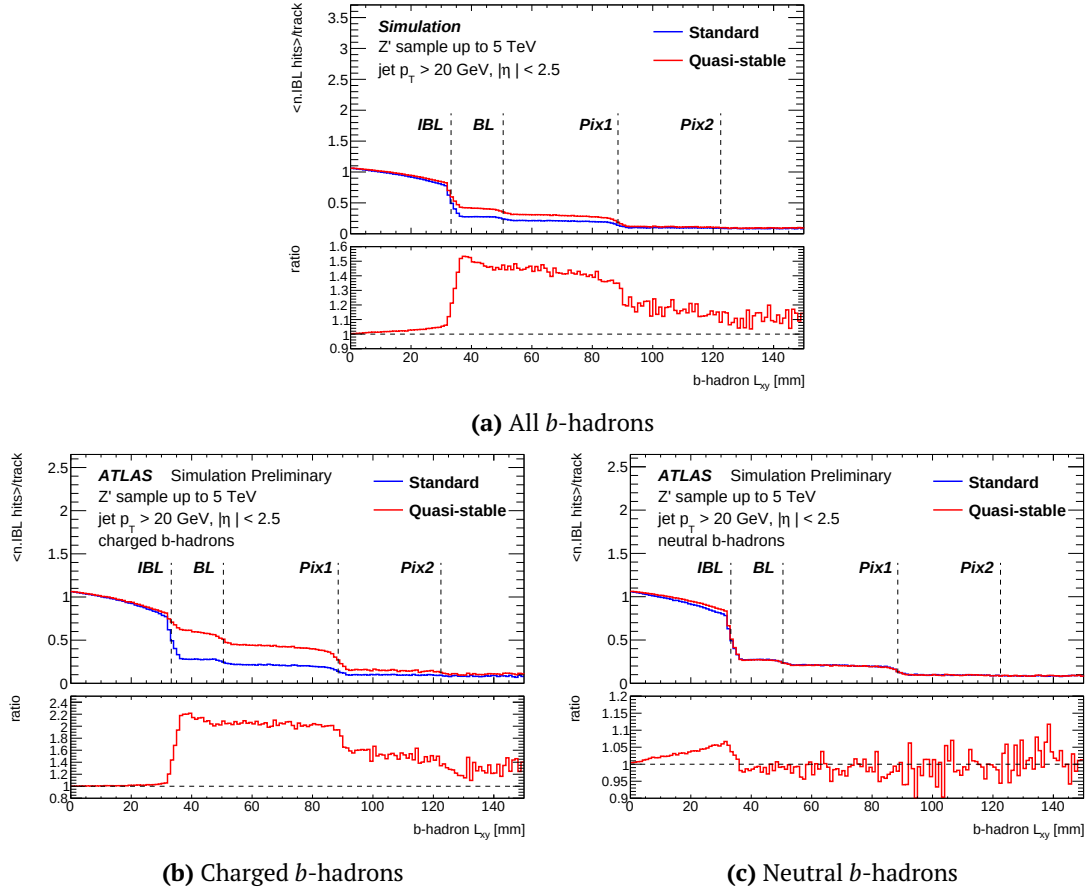


Figure 5.7: Average number of hits per track matched to the b -hadrons decay products in the IBL as a function of the transverse decay length of the b -hadron for the standard (blue) and quasi-stable (red) simulation. Figure (a) shows the distribution when considering all b -hadrons, while in (b) and (c) only the charged and neutral b -hadrons are considered, respectively. The bottom panel of all figures shows the ratio between the quasi-stable and the standard distributions. It can be observed that the additional hit left in the IBL is present only for charged b -hadrons decaying after the IBL, as expected [141].

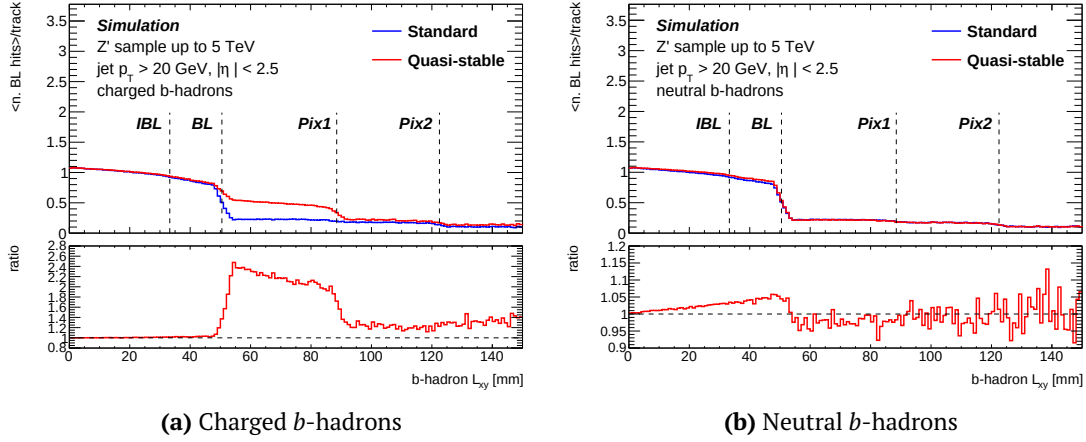


Figure 5.8: Average number of hits per track matched to the b -hadrons decay products in the B-layer as a function of the transverse decay length of the b -hadron for the standard (blue) and quasi-stable (red) simulation. Figure (a) and (b) show the distribution when considering only charged and neutral b -hadrons, respectively. The bottom panel of both figures shows the ratio between the quasi-stable and the standard distributions. It can be observed that the additional hit left in the BL is present only for charged b -hadrons decaying after the BL, as expected.

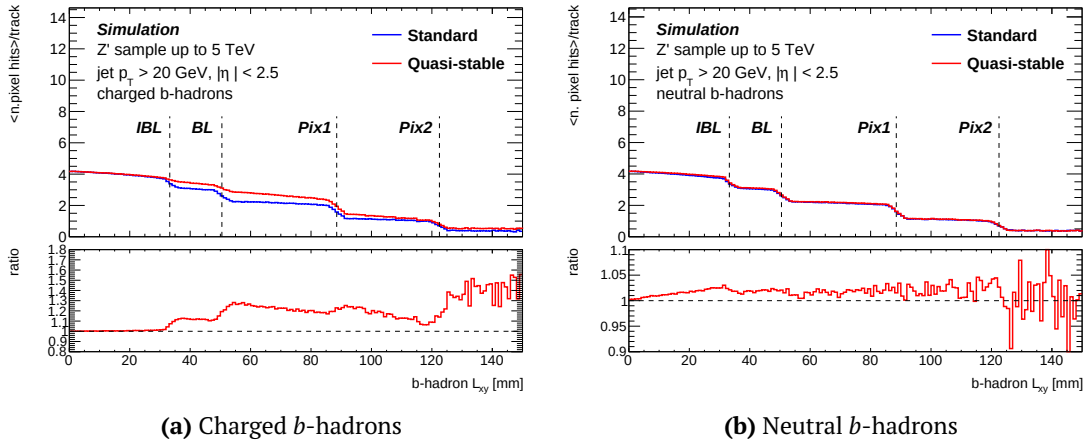


Figure 5.9: Average number of hits per track matched to the b -hadrons decay products in all pixel layers as a function of the transverse decay length of the b -hadron for the standard (blue) and quasi-stable (red) simulation. Figure (a) and (b) show the distribution when considering only charged and neutral b -hadrons, respectively. The bottom panel of both figures shows the ratio between the quasi-stable and the standard distributions.

Figures 5.10 (a) and (b) show the number of hits in the IBL associated to the tracks of the b -hadron decay products for charged or neutral b -hadrons, respectively, decaying before the IBL. In case a charged (neutral) b -hadron decays before the IBL, its decay products can leave a hit in the IBL in both the standard and the quasi-stable sample. The case of zero hits in the IBL can be explained by the drop in the cluster assignment efficiency shown in Figure 3.2, due to the very small distance between the particles traversing the IBL. The case of two hits can be due to a b -hadron passing through the overlap region of two pixel modules or to merged clusters split into two hits from the artificial neural network³ (see Section 3.1). Differences between the standard and the quasi-stable simulations can be observed, especially in the bin corresponding to zero hits. This can be due to b -hadrons decaying into charged c -hadrons before the IBL, that can interact with the IBL material leaving a hit before decaying. This results in a migration from the first to the second bin of the distribution which is more pronounced in the case of neutral b -hadrons as they are more likely to decay to charged c -hadrons.

When a charged b -hadron decays after the IBL, as in Figure 5.10 (b), the corresponding interaction with the pixel detector is not simulated in the standard sample, therefore a large fraction of events enter the first bin of the distribution where no hits are expected in the IBL. However, a residual $\sim 10\%$ of the events enter the second bin of the distribution as the hit left by a fragmentation track gets associated to one of the b -hadron decay products. The bin corresponding to two hits in the IBL associated to the same track corresponds to the case where the NN identifies a cluster as merged and splits it in two hits, but also to b -hadrons passing through the overlap region between two pixel modules.

In the case of the quasi-stable sample, when a charged b -hadron traverses the IBL, it leaves a hit in the pixel sensors that can be mistakenly associated to the track of one of its decay products. This is represented by the migration of events from the first to the second bin of the distribution. The second bin, as in the case of the standard simulation, is populated also by the hits originating from fragmentation tracks that are incorrectly assigned to the b -hadron decay products. The bin containing two hits in the IBL corresponds to the case where the b -hadron passes through the modules overlap region or to split clusters produced by the b -hadron and a fragmentation track.

Finally, if a neutral b -hadron decays after the IBL it does not leave a hit in this layer, neither for the standard nor for the quasi-stable simulation. Therefore, the majority of the events enter the first bin of the distribution. The second bin has the same origin as in the case of charged b -hadrons. The case of two hits can be due to split clusters.

Figure 5.11 shows the number of hits in all pixel layers associated to the tracks of the b -hadron decay products, when a charged or a neutral b -hadron decays before or after the IBL. The distributions for charged and neutral b -hadrons decaying before the IBL are very similar, with a maximum of four reconstructed pixel hits. On the other hand, the

³For a highly boosted b -hadron, several daughter particles are likely to contribute to the same cluster, however, the NN is able to identify a cluster as merged only if less than three tracks contribute to it, otherwise, it gets categorised as shared.

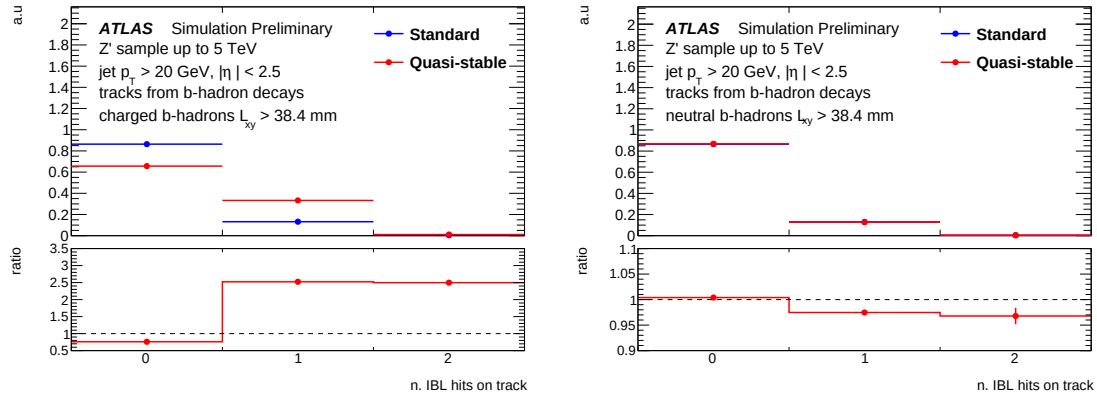
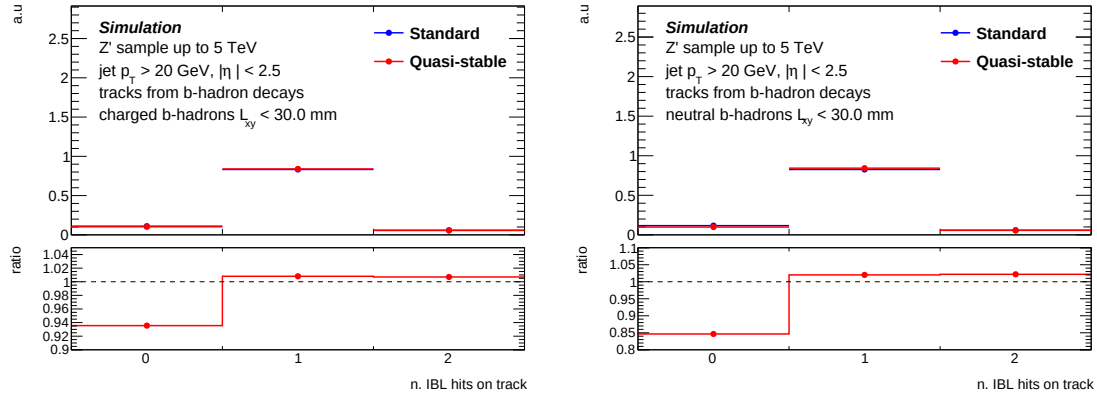
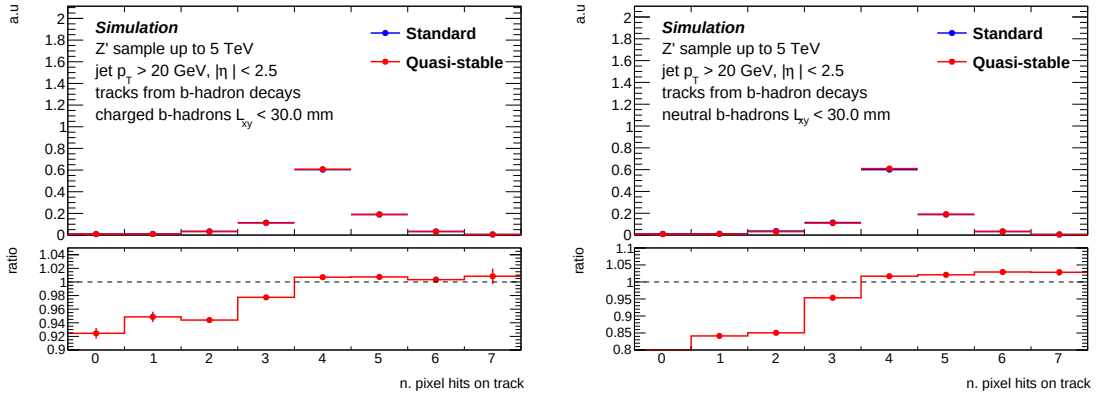
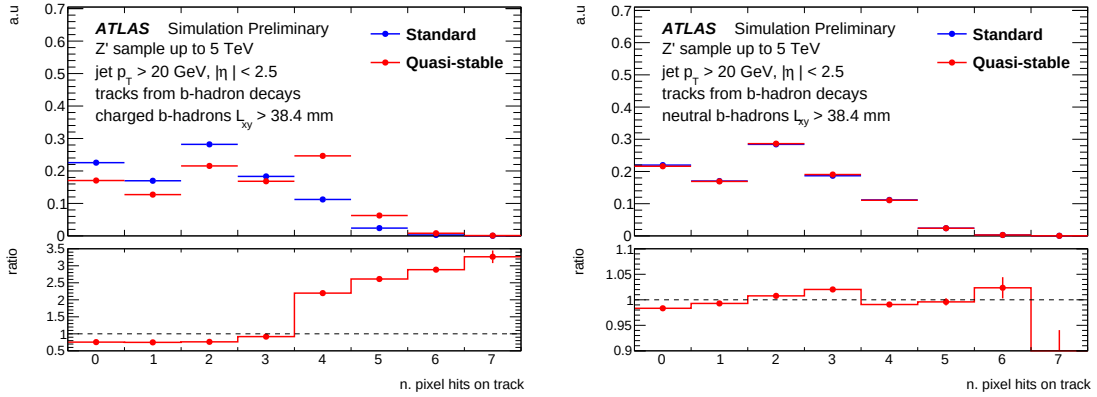


Figure 5.10: Number of hits in the IBL associated to the tracks of the decay products of charged (a) and neutral (b) b -hadrons decaying before (left) or after (right) the IBL [141].

distributions for charged and neutral b -hadrons decaying after the IBL are very different. In the case of neutral b -hadrons no significant differences are observed between the standard and the quasi-stable simulations. Whereas, if a charged b -hadron traverses a pixel layer, it can leave a hit that can be mistakenly associated to the track of a decay product, thereby increasing the number of pixel hits associated to it. Multiple hits are also possible if the b -hadron traverses a layer in the overlap region between two pixel modules, or if a merged cluster gets split into two by the NN during the ambiguity solver stage. A smaller number of associated hits might be related to the inefficiencies in the cluster assignment in very boosted topologies.



(a) Charged (left) and neutral (right) b -hadrons decaying before the IBL



(b) Charged (left) and neutral (right) b -hadrons decaying after the IBL

Figure 5.11: Number of hits in the pixel detector associated to the tracks of the decay products of charged (left) and neutral (right) b -hadrons decaying before (a) or after (b) the IBL [141].

5.2.2 Impact on the b -tagging efficiency

The change in the b -tagging efficiency due to the inclusion of b - and c -hadron interactions with the pixel detector has been studied for both the low-level b -tagging algorithms and high-level discriminants for the 70%, 77%, 85% official efficiency working points. In the case of the high-level taggers, such as MV2c10, DL1 and DL1r, the official WPs are derived on the combined $t\bar{t}$ and Z' sample up to 1.5 TeV described in Section 4.3.1. Therefore,

the corresponding cuts applied to the high-level taggers distributions obtained using the extended Z' sample up to 5 TeV will typically give a lower efficiency. This is shown in Figure 5.12 for the DL1r tagger and for the 70% WP, as an example.

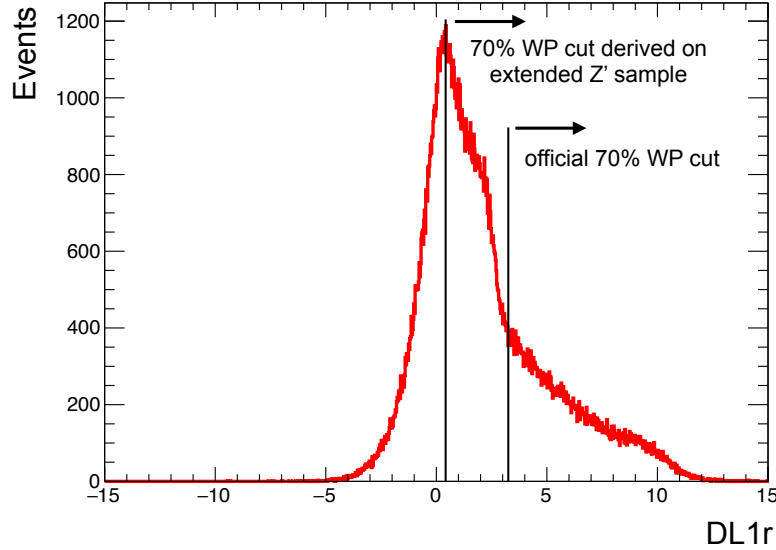


Figure 5.12: DL1r distribution derived on the extended Z' sample. The cut corresponding to the 70% efficiency WP derived on the extended Z' sample is shown. In addition, the cut corresponding to the 70% efficiency WP derived on the hybrid $t\bar{t}$ and Z' sample (official) is also shown. It can be observed that the official cut gives a lower efficiency with respect to the one derived on the extended Z' sample.

In the case of the low-level taggers, such as IP3D, SV1 and JETFITTER, the chosen cut is the one that gives the same b -jet efficiency as DL1r on the extended Z' sample. For all taggers, the performance of the 70% efficiency WP is reported in this section. The results for the remaining WPs and only for the high-level discriminants can be found in the Appendix A.2.

The l -jet efficiency is not affected by the difference between the standard and the quasi-stable simulation, and thus the b -jet efficiency is the only figure of merit. It has been checked that the l -jet efficiency is in fact the same between the two simulations and this is shown in Figure 5.13 for the DL1r tagger at the 70% WP as a function of the b -labelled jet p_T and for all b -hadrons. The compatibility between the two distributions has been quantified by a χ^2 test whose result is $\chi^2/n_{\text{d.o.f}} = 48.04/49 = 0.98$.

In the following, the b -tagging efficiency measured by the tagger is drawn as a function of the jet p_T and the b -hadron L_{xy} , and separately for charged and neutral b -hadrons. The efficiency as a function of the b -hadron L_{xy} is the most interesting distribution based on physics. The distribution of the b -tagging efficiency as a function of the jet p_T is interesting to give an idea of how the efficiency varies as a function of the b -hadron L_{xy} in a given p_T bin, but it is also a relevant quantity for many analyses.

A large number of bins is used in order to reduce the effect of the higher values of L_{xy} observed for the neutral b -hadrons shown in Figure 5.5 and still under investigation.

It has to be noted that when comparing the jet p_T distribution with the b -hadron L_{xy} one,

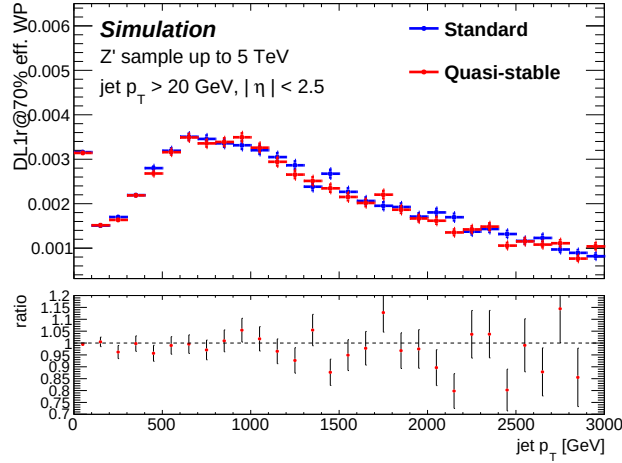


Figure 5.13: l -jet efficiency measured at the 70% WP of the DL1r tagger, as a function of the jet p_T and for all b -hadrons. No significant differences are observed between the two simulations.

one has to remember that the p_T spectrum of the extended Z' sample is not flat over the full range (see Figure 5.3).

Performance of low-level b -tagging algorithms

Impact Parameter based Algorithm IP3D Figures 5.14 (a) and (b) show the b -tagging efficiency of the IP3D algorithm as a function of the jet p_T , separately for charged and neutral b -hadrons. Larger values of the b -tagging efficiency can be observed in the quasi-stable sample mainly for charged b -hadrons.

Figures 5.14 (c) and (d) show the b -tagging efficiency measured by the IP3D algorithm as a function of L_{xy} , separately for charged and neutral b -hadrons. The standard sample exhibits a drop in the efficiency in correspondence with the average radius of each pixel layer, as was shown in Figure 5.6. In fact, for each additional layer traversed by a b -hadron, the number of available clusters for track reconstruction is reduced by one. This yields a smaller number of reconstructed tracks.

In the case of the quasi-stable simulation this drop in the efficiency is not observed due to the different hit pattern of the tracks entering the IP3D algorithm. This causes an increase in the number of tracks used by the IP3D tagger in the case of the quasi-stable simulation with respect to the standard one, which is purely artificial. Figure 5.15 demonstrates this effect by showing the average number of tracks used by the IP3D algorithm as a function of the b -hadron L_{xy} for charged and neutral b -hadrons, respectively.

This behaviour confirms what has been observed in Figures 5.7, 5.8, and 5.9, that the hit left by the b -hadron when traversing the layers is, in the majority of the cases, mis-associated to a child track. As a consequence, also the performance of the track reconstruction is altered and, in the case of the quasi-stable simulation, the reconstructed track can be assigned to a different track category (see Section 4.2.1) with respect to the standard simulation. Figure 5.16 (a) shows the 13 categories in which the reconstructed

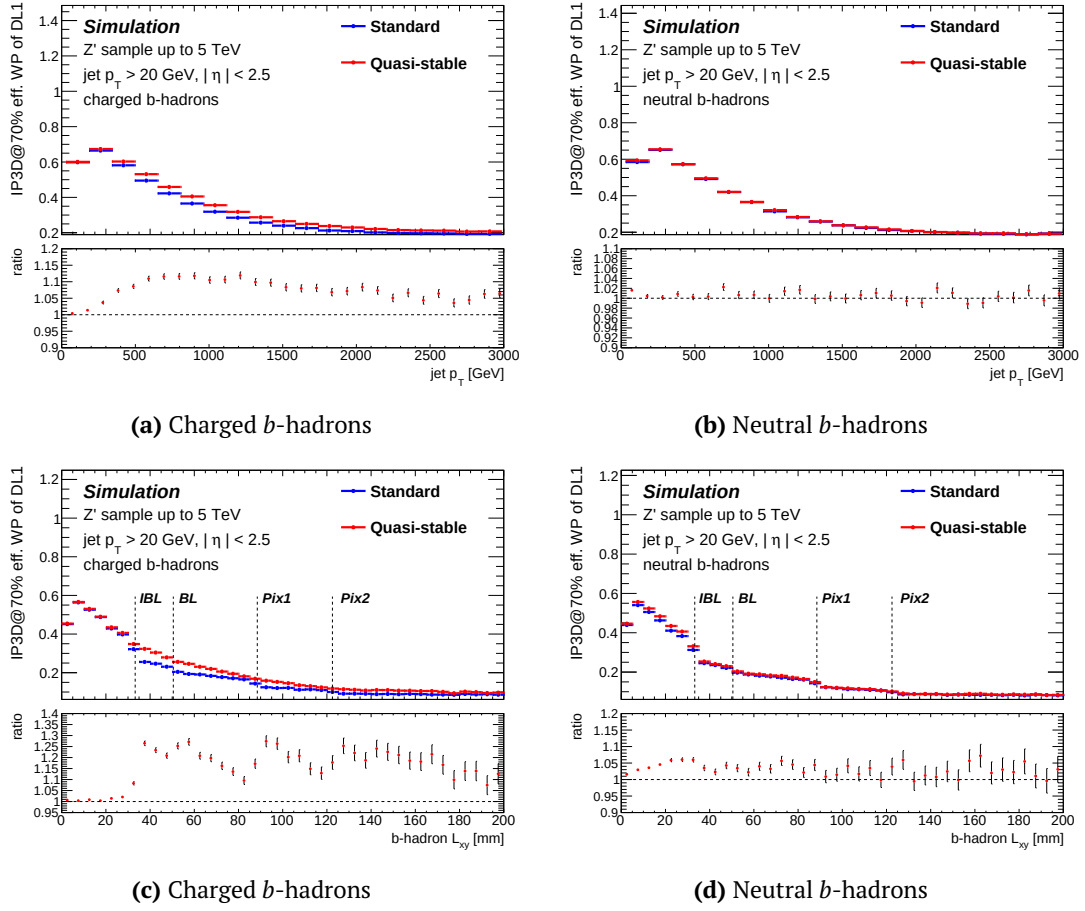


Figure 5.14: b -tagging efficiency measured at the 70% WP of the IP3D tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

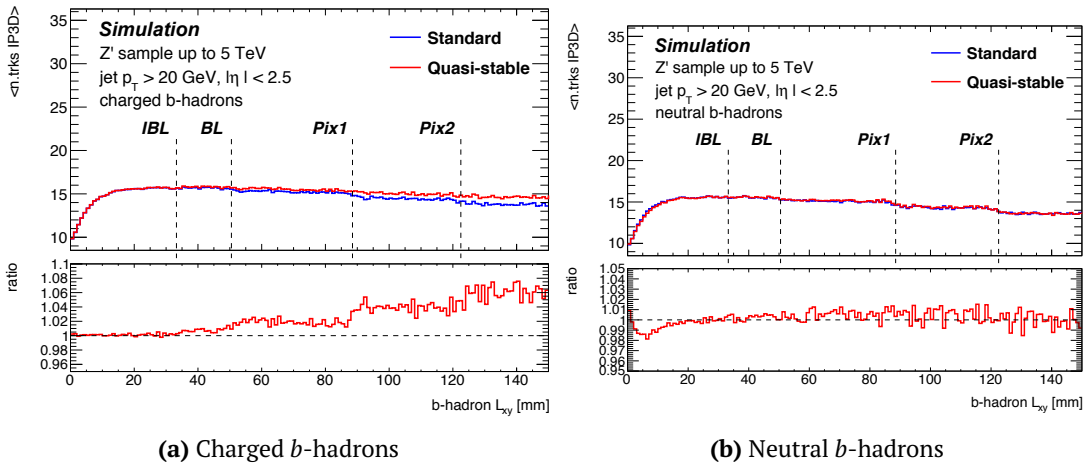


Figure 5.15: Average number of tracks used by the IP3D algorithm as a function of the b -hadron L_{xy} for charged (a) and neutral (b) b -hadrons.

tracks from the b -hadrons decaying after the IBL, and used by the IP3D tagger, are divided. It is interesting to notice that, in the case of the quasi-stable simulation, some of the tracks migrate to the 13th category that has the highest discrimination power between b -jets and l -jets, thus increasing the efficiency of the tagger.

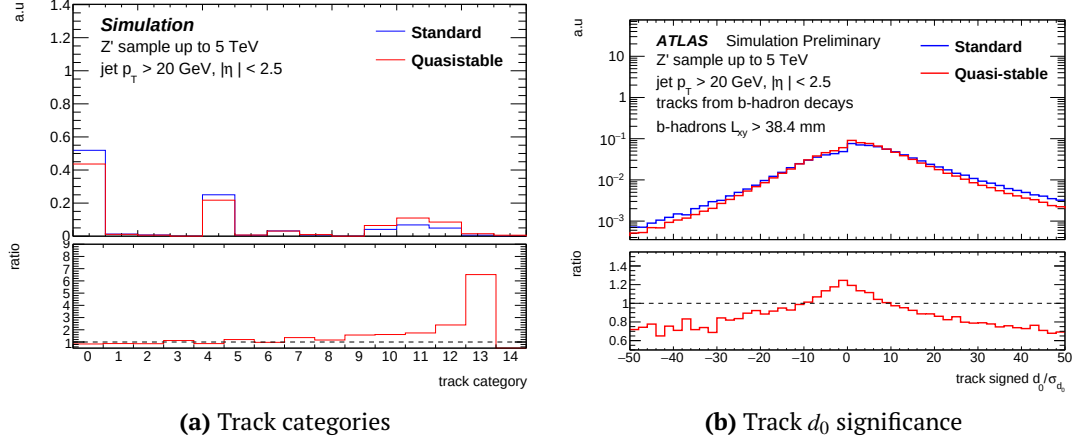


Figure 5.16: Figure (a) shows the reconstructed tracks categorisation for b -hadrons decaying after the IBL and Figure (b) shows the transverse impact parameter significance of the tracks from b -hadrons decaying after the IBL.

Figure 5.16 (b) shows the transverse impact parameter significance for the reconstructed tracks belonging to b -hadrons decaying after the IBL and used by IP3D. It can be observed that in the quasi-stable simulation, the hit left by the b -hadron interaction and mis-associated to a child track biases the IP significance towards smaller values.

Regarding the neutral b -hadron, a slightly higher efficiency can be observed in the quasi-stable sample for b -hadrons decaying before the IBL. This can be attributed to the decay of b -hadrons to charged c -hadrons which, in turn, traverse the pixel layers leaving a hit before decaying.

The l -jet rejection of the IP3D tagger as a function of the b -jet efficiency is shown in Figure 5.17 for charged and neutral b -hadrons, and for the standard and quasi-stable simulations. As the l -jet mis-tag rate is identical between the standard and the quasi-stable simulation, what changes is only the b -tagging efficiency.

Secondary vertex finding algorithm SV1 Figures 5.18 (a) and (b) show the b -tagging efficiency at the 70% WP of the SV1 algorithm as a function of the jet p_T , separately for charged and neutral b -hadrons. A slightly higher efficiency is observed for the quasi-stable sample in the case of charged b -hadrons.

Figure 5.18 (c) and (d) show the b -tagging efficiency at the 70% WP as a function of L_{xy} , separately for charged and neutral b -hadrons.

In the case of neutral b -hadrons, a slightly higher efficiency is observed for the quasi-stable simulation with respect to the standard one. This is due to b -hadrons decaying to c -hadrons which traverse the pixel layers before decaying.

To better understand the behaviour in the case of charged b -hadrons as a function

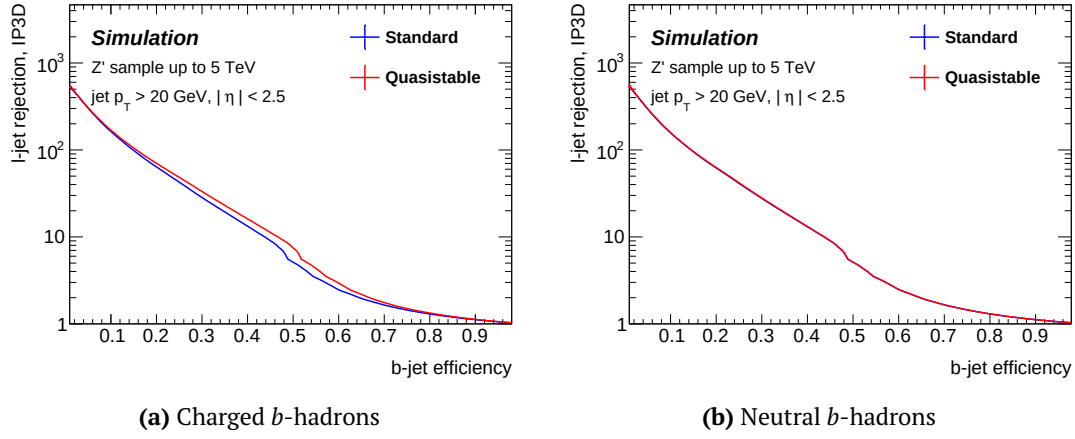


Figure 5.17: l -jet rejection of the IP3D tagger as a function of the b -jet efficiency for charged (a) and neutral (b) b -hadrons, and for the standard (blue line) and the quasi-stable (red line) simulations.

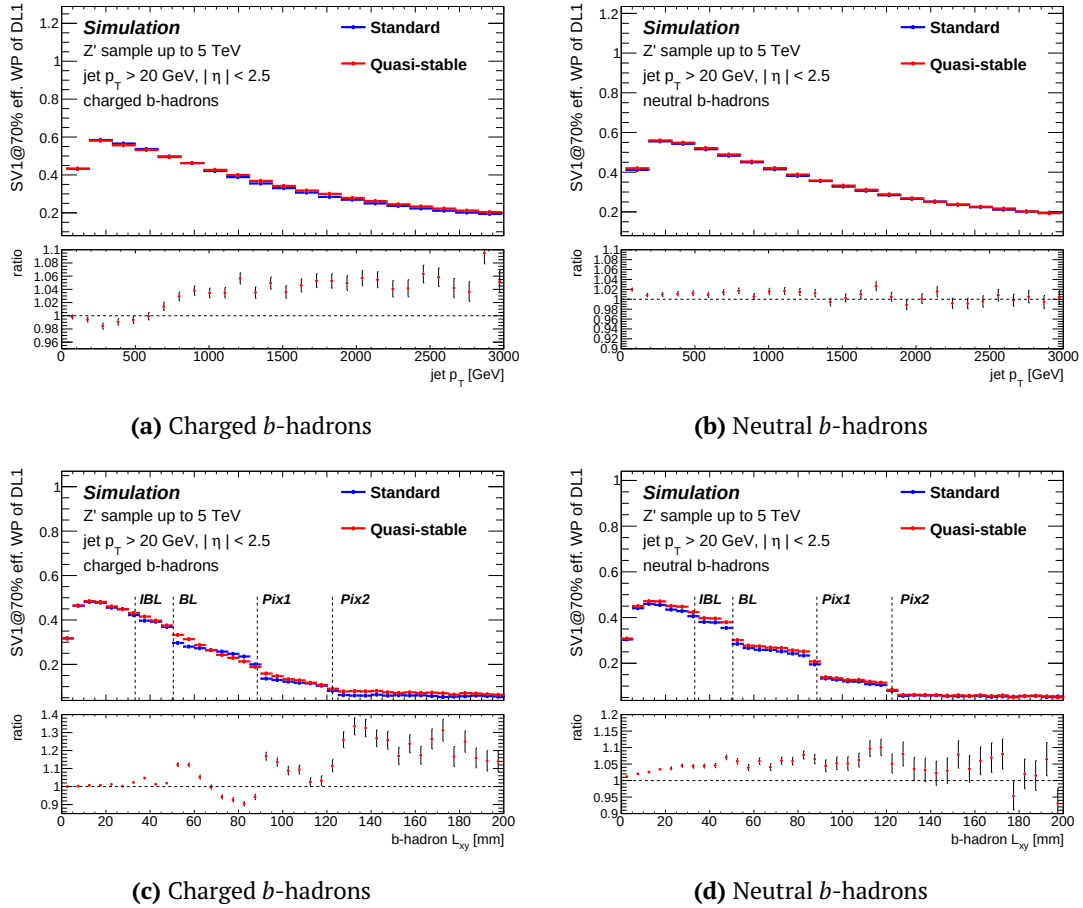


Figure 5.18: b -tagging efficiency at the 70% WP of the SV1 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

of the b -hadron L_{xy} , one can compare the b -tagging efficiency at the 70% WP of the SV1 tagger in Figure 5.18 (c) to the single secondary vertex reconstruction efficiency in Figure 5.19.

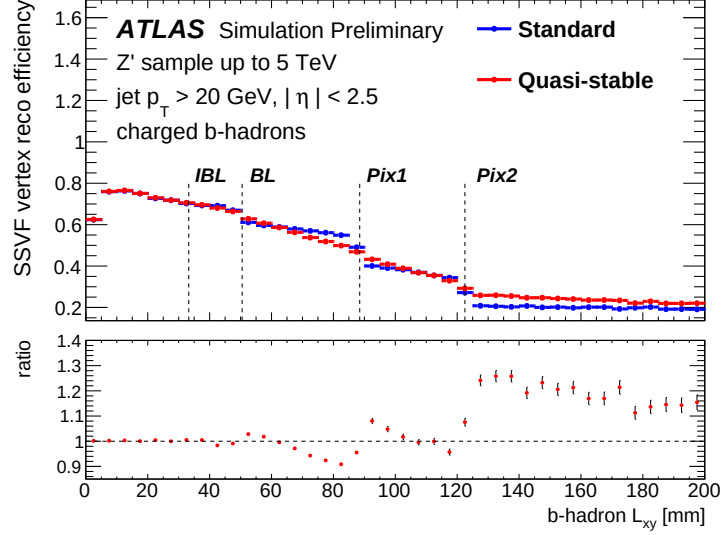


Figure 5.19: Single secondary vertex reconstruction efficiency as a function of the b -hadron L_{xy} for charged b -hadrons only [141].

It can be observed that the efficiency of the SV1 tagger is much smaller than the efficiency of reconstructing a single secondary vertex. This is expected because additional selection cuts on the kinematic variables entering the likelihood ratio method, on which SV1 is based, can only decrease the correspondent efficiency.

Furthermore, a very similar pattern as a function of L_{xy} can be observed in terms of the ratio between the quasi-stable and the standard distributions, however, larger effects are found for the SV1 algorithm. This can be due to differences between the standard and the quasi-stable simulations in the SV1 input quantities which can affect the SV1 output. The distribution of the SV1 input variables for the standard and the quasi-stable simulation are shown in Figure 5.20. In particular, these are:

- the total invariant mass of all particles in the reconstructed secondary vertex;
- the ratio of sum of energies of tracks in the SV to the energy of all tracks in the jet, defined as energy fraction;
- the number of two-track vertices;
- the ΔR between the SV-PV direction and the jet direction.

The presence of an additional hit from the b -hadron can modify the hit pattern of a track which can thus satisfy the SV1 track selection criteria in the case of the quasi-stable simulation but not in the standard one. This results in an increase in the number of reconstructed two-tracks vertices entering the SV1 algorithm, as well as an increase

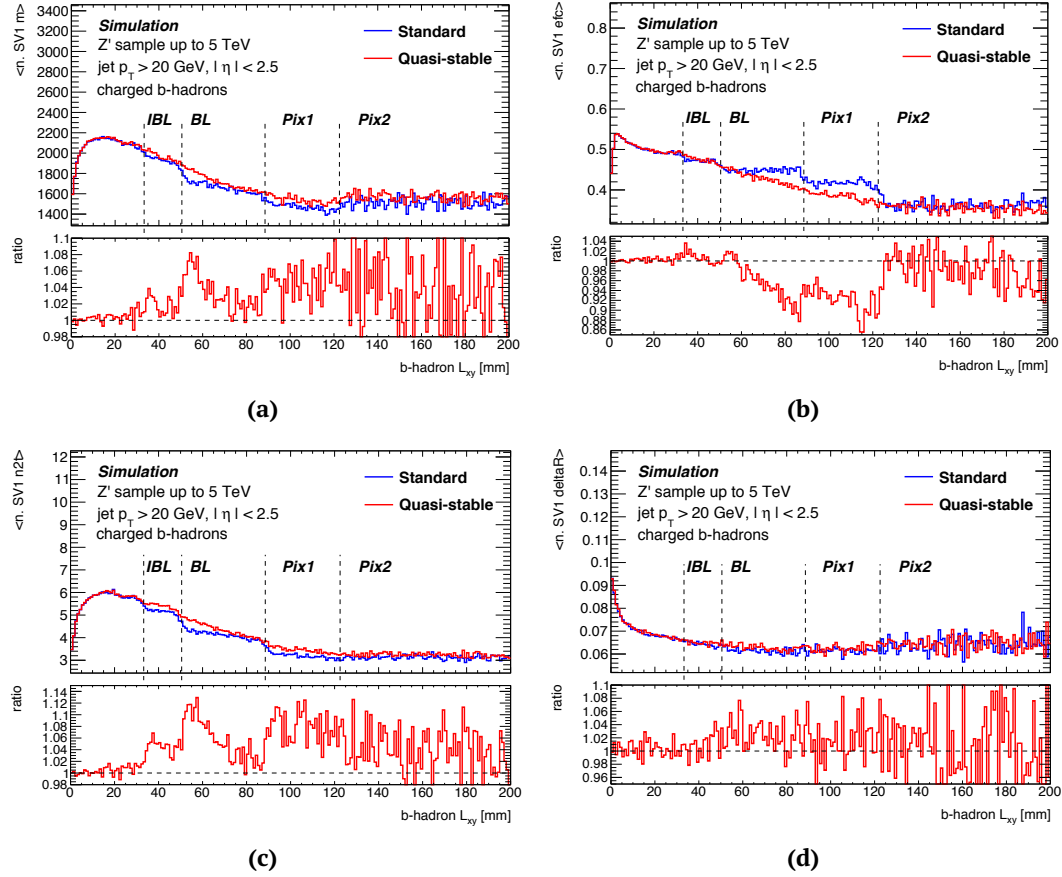


Figure 5.20: Distributions of variables entering the SV1 algorithm for the standard (blue) and the quasi-stable (red) simulations and for charged b -hadrons. The total invariant mass of all particles in the reconstructed secondary vertex is shown in (a), the ratio of sum of energies of tracks in the SV to the energy of all tracks in the jet in (b), the number of two-track vertices in (c), and the ΔR between the SV-PV direction and the jet direction in (d).

in their total invariant mass. In the case of the energy fraction, a decrease between the BL and Pix2 is observed in the case of the quasi-stable simulation with respect to the standard one. The only way this can happen is that not only the number of tracks associated to the SV increases, but also the number of reconstructed tracks in the jet. This is due to different track selection criteria between the nominator and the denominator. No significant differences are observed in the ΔR between the SV-PV direction and the jet direction for the two simulations.

Lastly, it can be observed that close to a pixel layer the efficiency in the case of the quasi-stable sample is higher than in the standard one. This is related to a cleaning cut applied in the SSVF algorithm which requires that “*no secondary track should have associated detector hits with radii smaller than the radius of the SV found by the algorithm*”. This means that in the quasi-stable simulation, if a b -hadron traverses a pixel layer and the hit from its interaction gets associated to a child track, the reconstructed vertex with the track having the wrong hit does not satisfy the cleaning cut and is removed. However, close to the pixel layer, this condition might not be satisfied. The following are two possible explanations of this effect:

- a b -hadron decays to a c -hadron immediately before the pixel layer, and the c -hadron traverses the layer and decays immediately after. The hit left by the c -hadron is then mis-associated to a child track from the b -hadron, but the position of the decay vertex is located before the c -hadron hit. In this case, the cleaning cut of the SSVF algorithm is satisfied, and the vertex is not removed;
- if a b -hadron decays after a pixel layer, but very close to it, the reconstructed vertex from its decay might not be removed due to resolution effects in its reconstructed position.

Multi-vertex finding algorithm JETFITTER Figures 5.21 (a) and (b) show the b -tagging efficiency at the 70% WP of the JETFITTER algorithm as a function of the jet p_T , separately for charged and neutral b -hadrons. Figures 5.21 (c) and (d) show the b -tagging efficiency at the 70% WP as a function of the b -hadron L_{xy} . Similar conclusions to the SV1 tagger can be drawn.

Performance of high-level b -tagging discriminants

The high-level taggers combine the inputs from the low-level taggers using a multivariate classifier in order to maximize the b -tagging performance. Three taggers are currently available, the MV2 using a Boosted Decision Tree, and DL1 and DL1r based on deep learning neural networks.

MV2 discriminant Figure 5.22 shows the b -tagging efficiency at the 70% WP of the MV2c10 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

For charged b -hadrons, a drop in the efficiency for the quasi-stable simulation can be observed. This might be due to the fact that, contrary to the low-level taggers, high-

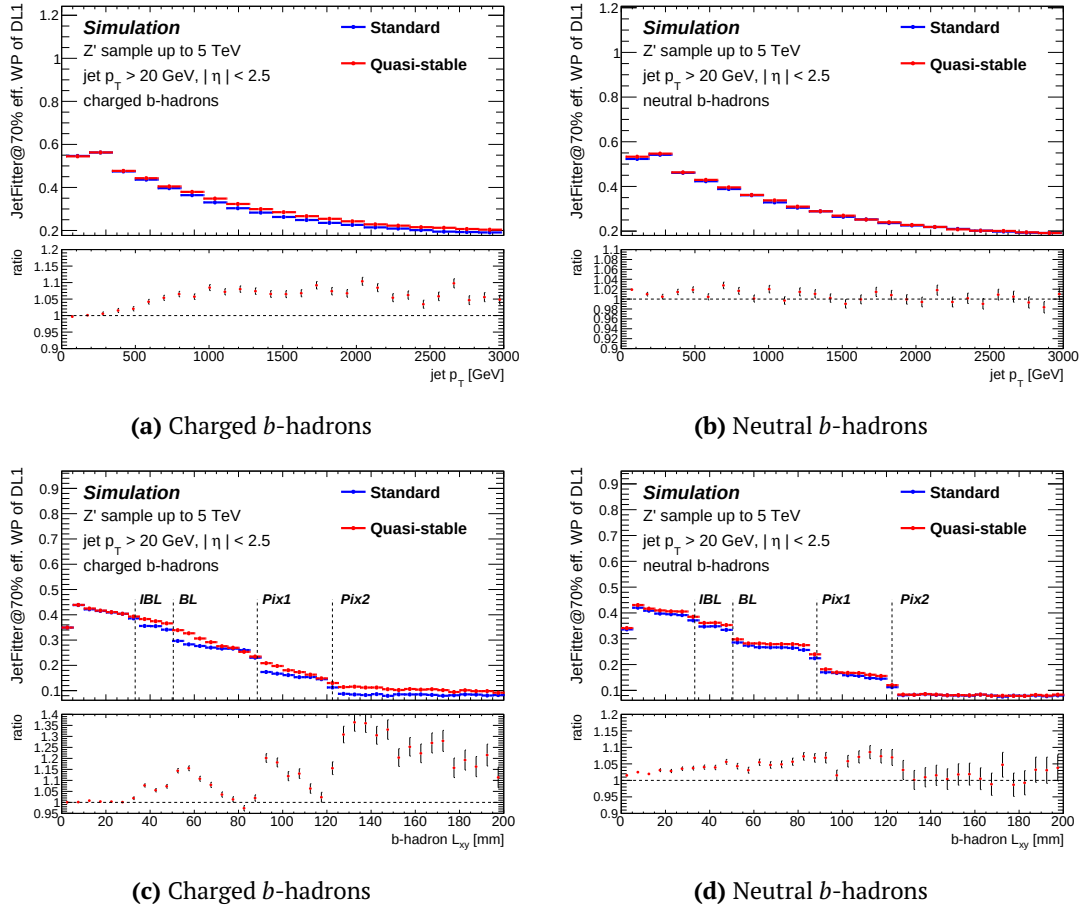


Figure 5.21: b -tagging efficiency measured at the 70% WP of the JETFITTER tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

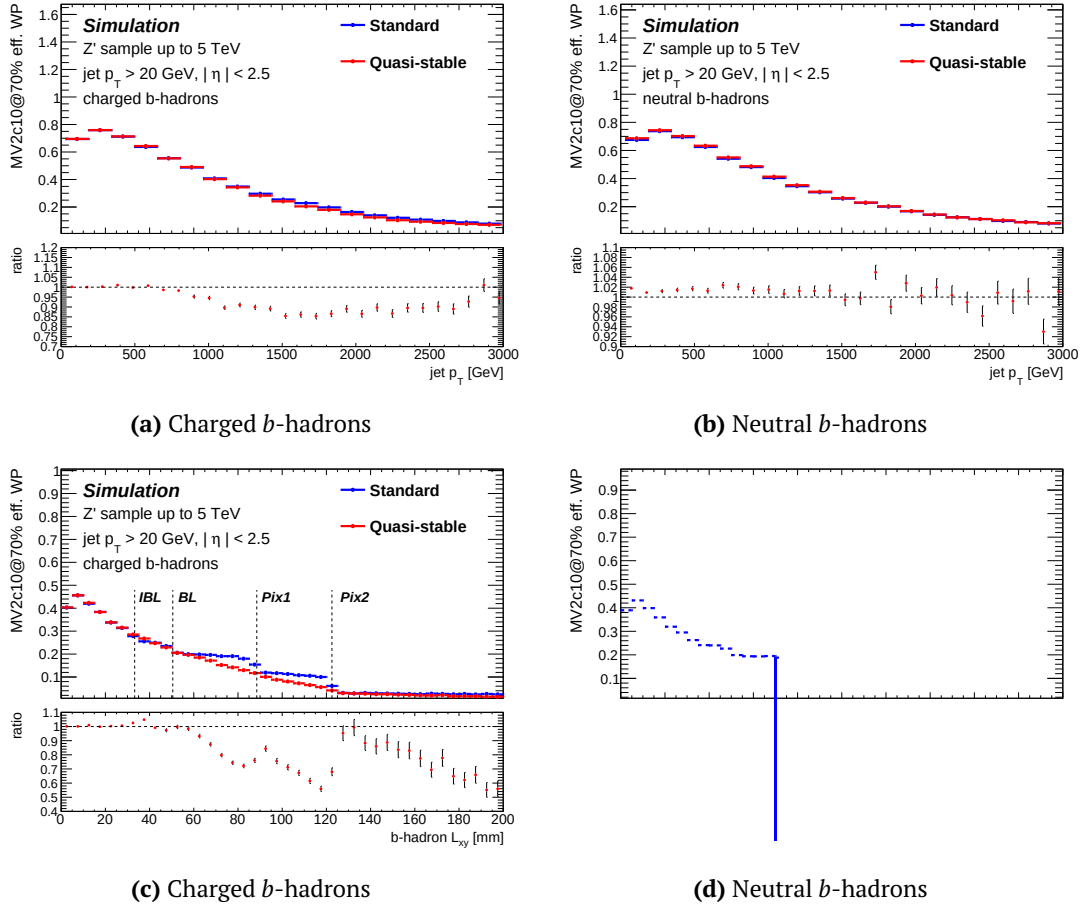


Figure 5.22: b -tagging efficiency measured at the 70% WP of the MV2c10 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

level discriminants like MV2c10, DL1 and DL1r, are supervised training algorithms optimised on the standard simulation which does not have the additional hits from the b -hadrons. Therefore, they are expected to perform worse on the quasi-stable sample due to the different hit pattern with which the tracks are reconstructed. A test was performed by evaluating the efficiency of the IPRNN tagger, trained as well on the standard sample, which is based on a neural network with the track transverse and longitudinal IP significances and the track category as input variables, in order to validate this hypothesis. This efficiency is derived for all b -hadrons as a function of the jet p_T and the b -hadron L_{xy} , and it is shown in Figure 5.23. Even for the IPRNN tagger a drop in the measured b -tagging efficiency for the quasi-stable sample can be observed in both the jet p_T and the b -hadron L_{xy} distributions.

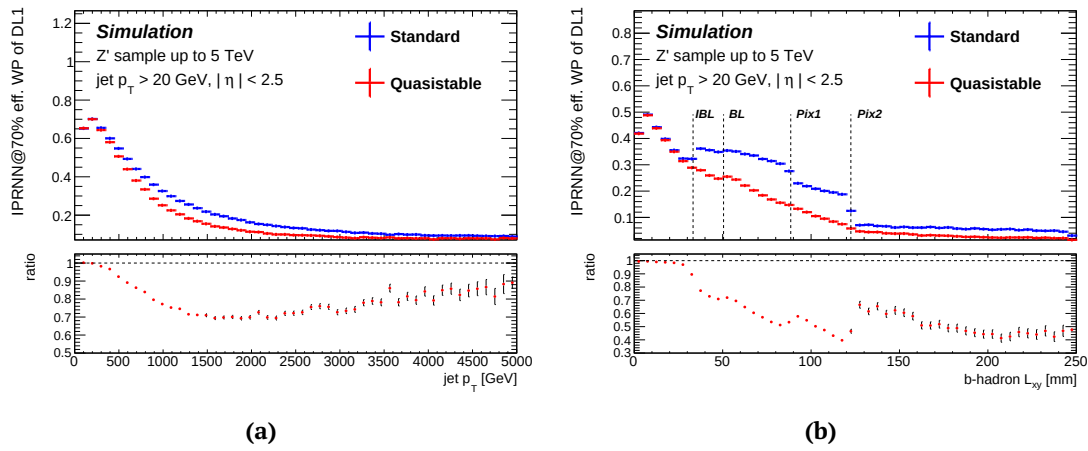


Figure 5.23: b -tagging efficiency measured at the 70% WP of the IPRNN tagger, as a function of the jet p_T (a) and the b -hadron L_{xy} (b) for all b -hadrons.

Without having retrained the MV2c10 discriminant on the quasi-stable sample, as well as DL1 and DL1r, one can only observe that a tagger optimised on a sample with given conditions (no hit from b -hadron interactions), performs worse when the conditions are changed (additional hit from b -hadron interactions).

No other observation can be made on the behaviour of the high-level taggers without retraining them on the quasi-stable sample, as it would be pure speculation.

DL1 discriminant Figure 5.24 shows the b -tagging efficiency at the 70% WP of DL1 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons. Same conclusions as for the MV2c10 discriminant can be drawn.

DL1r discriminant Figure 5.25 shows the b -tagging efficiency at the 70% WP measured by the DL1r discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons. Here the drop in the efficiency caused by the additional hits from the b -hadron interactions is larger than in the MV2c10 and DL1 taggers.

The studies presented in this section triggered an on-going effort within both the Flavour Tagging Working Group [142] and the Simulation Working Group [143], with

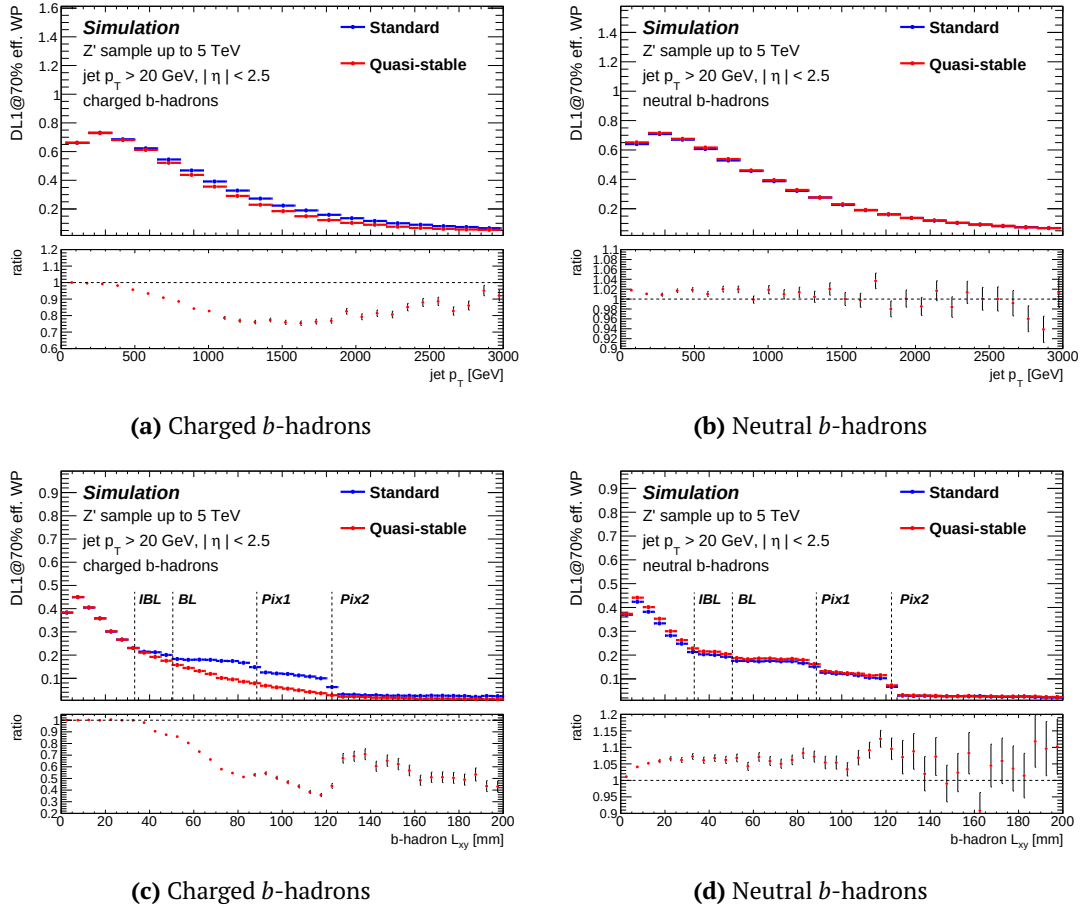


Figure 5.24: b -tagging efficiency measured at the 70% WP of the DL1 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

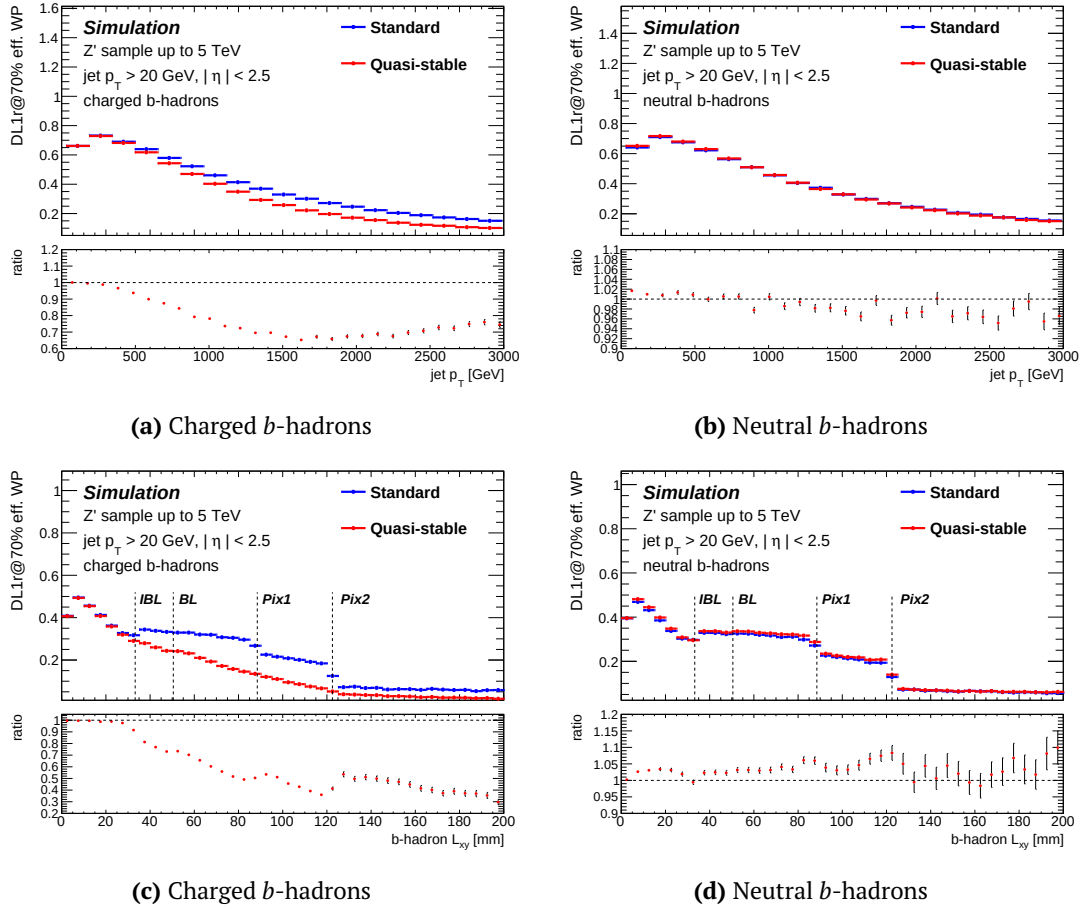


Figure 5.25: b -tagging efficiency measured at the 70% WP of the DL1r tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

the goal of making the quasi-stable simulation the new default simulation used by the ATLAS Collaboration by the start of Run 3.

5.2.3 Extrapolation uncertainty on the b -tagging efficiency

The studies presented in the previous sections led to the derivation of extrapolation uncertainties accounting for the impact of the b -hadron interactions on the b -tagging efficiency as a function of the b -labelled jet p_T . These were derived using the DL1r tagger, as this is the one currently recommended by the Flavour Tagging Working Group. Figure 5.26 shows the b -tagging efficiency measured at all official WPs of the DL1r discriminant, as a function of the jet p_T and for all b -hadrons.

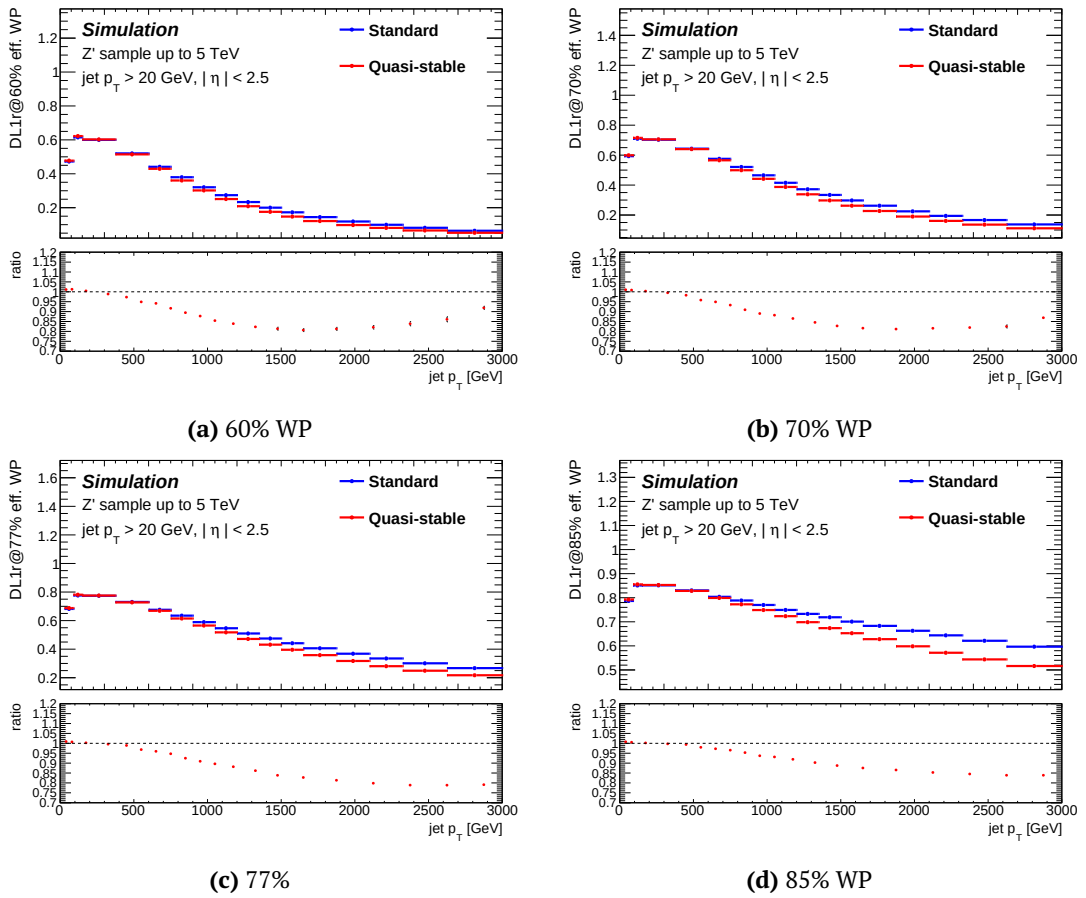


Figure 5.26: b -tagging efficiency measured at the 60% (a), 70% (b), 77% (c), and 85% (d) WP of the DL1r tagger, as a function of the b -labelled jet p_T .

A single extrapolation uncertainty for each jet p_T bin was derived from the ratio between the quasi-stable and the standard distributions, as shown in the bottom panel, of which the final values are also reported in Table 5.1 for all working points. The uncertainties in the first three bins of the distributions are very small and are not reported in the Table.

Table 5.1: Extrapolation uncertainties accounting for the effect of the *b*-hadron interactions on the *b*-tagging efficiency at all official WPs of the DL1r discriminant.

Range [GeV]	Working point			
	60%	70%	77%	85%
250.0 < p_T < 400.0	1.14%	0.51%	0.48%	0.24%
400.0 < p_T < 500.0	2.69%	1.73%	1.03%	0.58%
500.0 < p_T < 600.0	5.13%	4.28%	3.14%	2.01%
600.0 < p_T < 700.0	5.80%	5.17%	4.03%	2.79%
700.0 < p_T < 800.0	8.35%	6.72%	5.25%	3.45%
800.0 < p_T < 900.0	10.67%	9.16%	7.46%	4.78%
900.0 < p_T < 1000.0	12.23%	11.04%	9.13%	6.33%
1000.0 < p_T < 1100.0	14.55%	11.93%	10.35%	6.88%
1100.0 < p_T < 1250.0	16.11%	13.48%	11.79%	8.10%
1250.0 < p_T < 1400.0	17.77%	15.48%	13.90%	9.76%
1400.0 < p_T < 1550.0	18.66%	17.25%	16.23%	11.31%
1550.0 < p_T < 1750.0	19.44%	18.36%	17.34%	12.41%
1750.0 < p_T < 2000.0	18.79%	18.91%	18.79%	13.41%
2000.0 < p_T < 2250.0	18.02%	18.47%	20.23%	14.73%
2250.0 < p_T < 2500.0	16.26%	18.14%	21.24%	15.51%
2500.0 < p_T < 2750.0	14.06%	17.58%	21.25%	16.17%
2750.0 < p_T < 3000.0	8.20%	13.15%	20.92%	16.17%

5.3 Trackless b -tagging for very energetic jets

For very energetic b -hadrons, the decay time and the angle between the decay products transform according to the Lorentz transformations, as $\tau \rightarrow \tau\gamma$ and $\alpha \rightarrow \alpha/\gamma$, resulting in a highly displaced and boosted topology. In such a dense environment, the tracking performance suffers from three main sources of inefficiency. Figure 5.6 shows that the average number of reconstructed tracks from the b -hadron decay products decreases as a function of the b -hadron transverse decay length. In addition, a very small separation between the b -hadron decay products results in a high hit density and a more ambiguous association of hits with tracks, as seen in Figure 3.2. Another source of inefficiency in the track reconstruction is the increase of fragmentation tracks as a function of the p_T of the b -jet, as shown in Figure 5.27. These are tracks originating from the hadronization of the b -quark whose measured IP is compatible with the primary vertex. In contrast, the ability of reconstructing tracks from the b -hadron decay decreases at high- p_T .

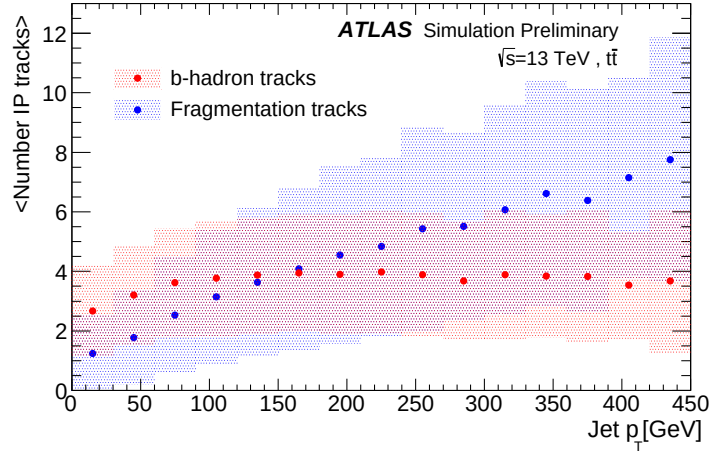


Figure 5.27: Average number of tracks from fragmentation and b -hadron decays selected for the IP algorithm as a function of the jet p_T . The shaded bands represent the RMS of the distribution for each bin [127].

In this regard, it is interesting to see how the clusters originating from b -hadron decays and those from the fragmentation contained in a b -jet evolve as a function of the b -jet p_T . Figure 5.28 shows the average number of clusters from b -hadron decays, and the sum of fragmentation and b -hadron decay clusters in the IBL contained within a cone of $\Delta R = 0.04$ as a function of the jet p_T . Tracks from fragmentation are identified as all tracks not associated to the b -hadron decays, as the probability to have a hit originating from pile-up is negligible, due to the very small cone size. This approach was demonstrated to be pile-up insensitive up to the expected HL-LHC pile-up values [144]. It can be observed that up to 150 GeV, the clusters are mainly coming from the b -hadron decay products. This is because in the low p_T region, the track multiplicity is higher in the case of b -hadrons than fragmentation. After 150 GeV, the abundance of b -hadron decay clusters slowly decreases due to the high boost of the system which can result in two particles

crossing the same silicon sensor, and due to the fraction of *b*-hadrons decaying after the IBL. The fragmentation component, instead, increases at higher jet p_T , as expected from Figure 5.6.

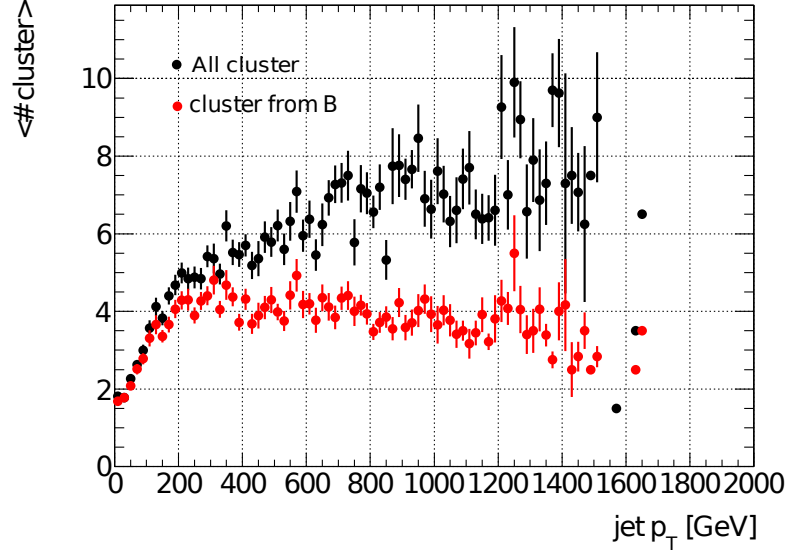


Figure 5.28: The average number of clusters in the IBL originating from *b*-hadrons decays and the sum between fragmentation and *b*-hadrons decay clusters, both contained in a cone of radius 0.04, is shown as a function of the jet p_T .

Since the *b*-tagging algorithms are mainly based on tracking information, the *b*-tagging performance is expected to significantly decrease for highly displaced and boosted topologies. A novel approach was proposed, based on the number of pixel hits left in subsequent layers of the tracker [144]. This approach, referred to as *trackless b-tagging*, is expected to complement the track-based algorithms for very high- p_T *b*-jets. Exploiting the high resolution and granularity of the pixel detector, one can define a *pixel cone* of radius $\Delta R(\text{hit}, \text{jet axis})$ around the jet axis, small enough to contain the highly collimated products from the *b*-hadron decay. Different values of the pixel cone radius have been studied, from 0.04 to 0.01, being an order of magnitude smaller than the radius parameter used in the calorimeter to define the cone size of a jet. Figure 5.29 shows the main concept of the algorithm: a jet containing a *b*-hadron that decays between two pixel layers can be identified by detecting a “jump” in the number of pixel hits contained in the pixel cone at increasing radii.

The relative multiplicity jump f_j at layer j can be defined as:

$$f_j = \frac{N_{j+1} - N_j}{N_j} = \frac{\Delta N_j}{N_j} \quad (5.2)$$

meaning that for $f_j = 1$ there are twice as many hits in layer $j+1$ than in layer j . This discriminant works best when the *b*-hadron decays between two layers of the pixel detector and into many charged daughter particles, so that f_j is high.

Starting from this definition, a more complex discriminant can be built by taking f_j

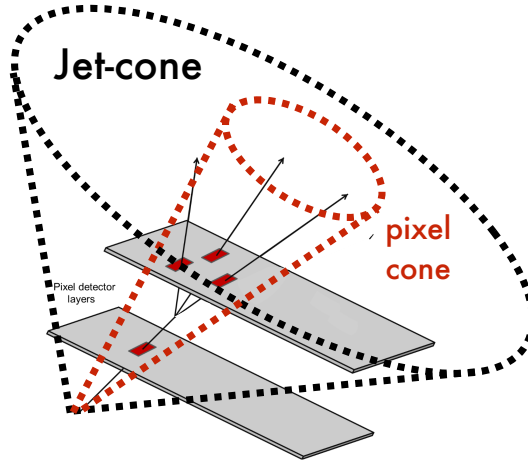


Figure 5.29: A b -hadron decaying between two layers of the pixel detector can be identified by detecting a jump in the number of hits on subsequent pixel layers, contained in a defined pixel cone [144].

giving the maximum jump among all pairs of layers j and $j+1$. This will be referred to as “max_trackless” discriminant in the following and an example of its distribution for b -jets and l -jets is shown in Figure 5.30. The “jumps” are clearly visible at each integer value of the discriminant for the b -jets, corresponding to the case in which $N_j = 1$, while they are less prominent in the case of l -jets.

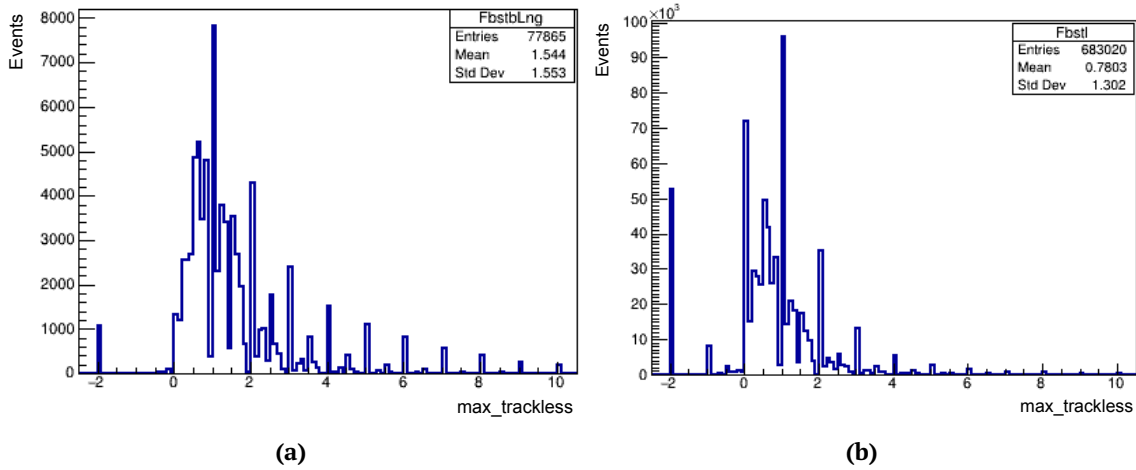


Figure 5.30: Distribution of the max_trackless discriminant for b -jets (a) and l -jets (b). In the cases in which $N_j = 0$, the value of the discriminant is set to -2. (Courtesy of Prof. Todd Huffman, Lady Margaret Hall - University of Oxford).

A more sophisticated version of the trackless b -tagging algorithm can be built by using not only the information on the hit multiplicity, but also the information from different concentric cones around the jet direction to exploit the differences in jet structure between b -jets and l -jets. The possible advantages of using this additional information were studied in MC simulation in the context of the identification of the heavy flavour jets at the hadron Future Circular Collider (FCC-hh) [145]. A multivariate analysis is

performed, based on a BDT classifier and using the TMVA [133] package. Two different input variables are passed to the BDT; the cluster multiplicity in each layer in a single cone for different cone sizes and the cluster multiplicity in each layer contained within a default ΔR of 0.006.

In Figure 5.31 (a) the BDT performance using different cone sizes are compared to the default ($R=0.006$), with the ΔR cone of half the default value having the highest discrimination power. Furthermore, if the information on the cluster multiplicity in all of the four ΔR cone sizes is included in the BDT, the rejection increases by up to 50%.

In Figure 5.31 (b) the l -jets rejection is shown as a function of the b -tagging efficiency in the case the input corresponds to the cluster multiplicity information in different numbers of detector layers. The best discrimination power is achieved when providing the BDT with the information from all layers.

An investigation on this approach is currently on-going within the ATLAS experiment.

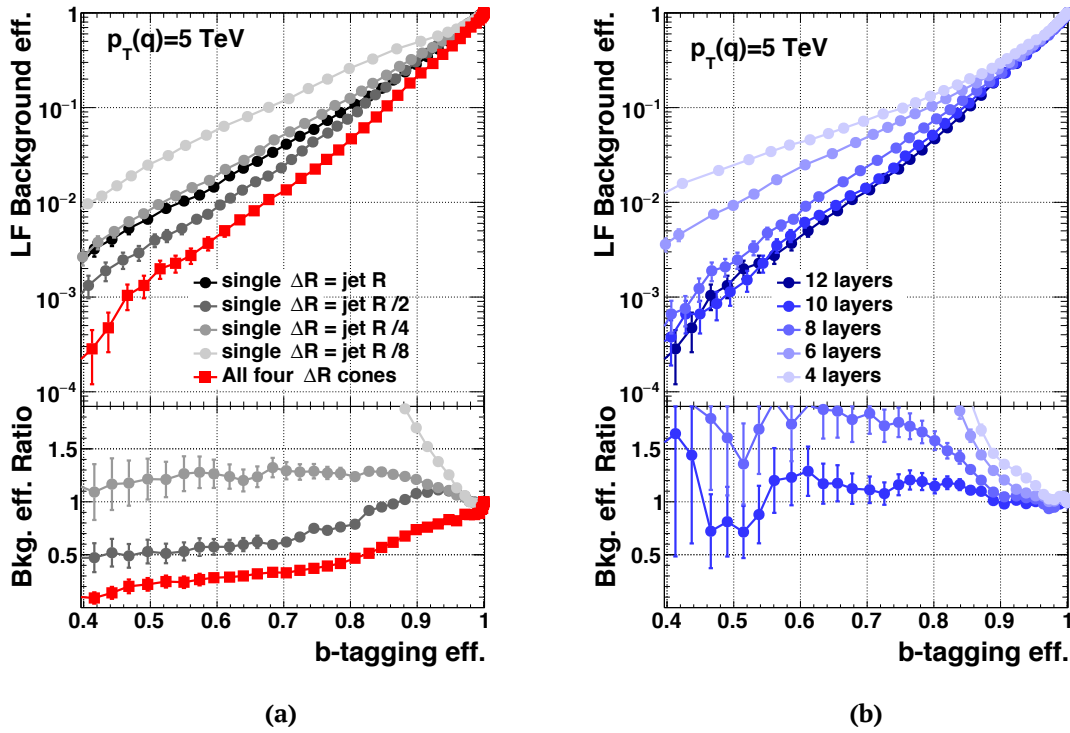


Figure 5.31: l -jets misidentification rate as a function of the b -tagging efficiency. In Figure (a), the BDT rejection power is evaluated using the cluster multiplicity information in a single cone for different cone sizes (in black and in grey) or combining the information in all of the four different cones (in red). In the bottom panel, the ratio is calculated with respect to the performance using a single ΔR cone with default size of 0.006. In Figure (b), the BDT rejection power is evaluated using the cluster multiplicity information in different numbers of innermost layers. In the bottom panel, the ratio is calculated with respect to the performance when including all 12 layers of the FCC-hh pixel detector [145].

5.3.1 Effect of b -hadron interactions on trackless b -tagging performance

As seen in Section 5.2, a very energetic charged b -hadron can leave a hit in each layer of the pixel detector it traverses before decaying. These additional hits can have a direct effect on the performance of the trackless b -tagging algorithm. In fact, based on Equation 5.2, the additional hit from the quasi-stable simulation in one or more layers of the pixel detector would increase the value of N_j , resulting in smaller values of f_j . Consequently, the ability of discriminating between b -jets and l -jets would also be reduced.

To study the effect of b -hadron interactions, the b -tagging efficiency is evaluated with `max_trackless` as discriminant, using the extended Z' sample introduced in Section 5.2 in order to enhance the fraction of b -hadrons traversing the pixel layers, for both the standard and the quasi-stable simulations. One caveat regarding the production of these samples has to be mentioned. In the ATLAS simulation data format, the pixel information is typically removed due to its large size. A large fraction of the work was dedicated to the generation of samples correctly storing the hit information. Given that only the sample produced using the quasi-stable simulation was available, the standard sample was constructed from the quasi-stable one by removing one cluster from the total number of clusters contained in a given pixel layer, in all cases in which a charged b -hadron traversed that layer (b -hadron L_{xy} greater than the pixel radius). It was not possible to check if the b -hadron really left a hit in the pixel, therefore, the difference between the quasi-stable and the derived standard sample is expected to be the largest possible. However, given that the pixel efficiency is $\sim 100\%$, the observed differences are very close to reality.

In case a b -hadron decays into a c -hadron before a pixel layer, and the c -hadron traverses the layer before decaying, the cluster possibly left by the c -hadron is not removed. This means that, before the IBL, the standard and quasi-stable distributions as a function of the b -hadron L_{xy} are expected to be identical.

Figure 5.32 shows the b -tagging efficiency of `max_trackless` at the 50% WP as a function of the b -hadron L_{xy} and the jet p_T , for the standard and the quasi-stable simulation and for all b -hadrons. The large number of fragmentation tracks at high jet p_T (as shown in Figure 5.27), is expected to leave the same number of hits in two consecutive layers. Therefore, the `max_trackless` variable is equal to zero for a large fraction of the b -jets, resulting in a b -tagging efficiency limited to $\sim 50\%$.

An almost flat $\sim 9\%$ difference can be observed between standard and quasi-stable simulation as a function of the jet p_T . The reduction in the tagging efficiency as a function of the b -hadron L_{xy} after the IBL is due to the fact that the value of N_j in Equation 5.2 is bigger in the quasi-stable than in the standard simulation, as it contains an additional hit in the j^{th} pixel layer coming from the b -hadron interaction.

Although the b -tagging efficiency reaches $\sim 80\%$ at high- p_T , the corresponding light jet rejection of the tagger is poor, as shown in Figure 5.32 (c) for the quasi-stable simulation. Therefore, it is natural to evaluate how much one can effectively profit from the use of

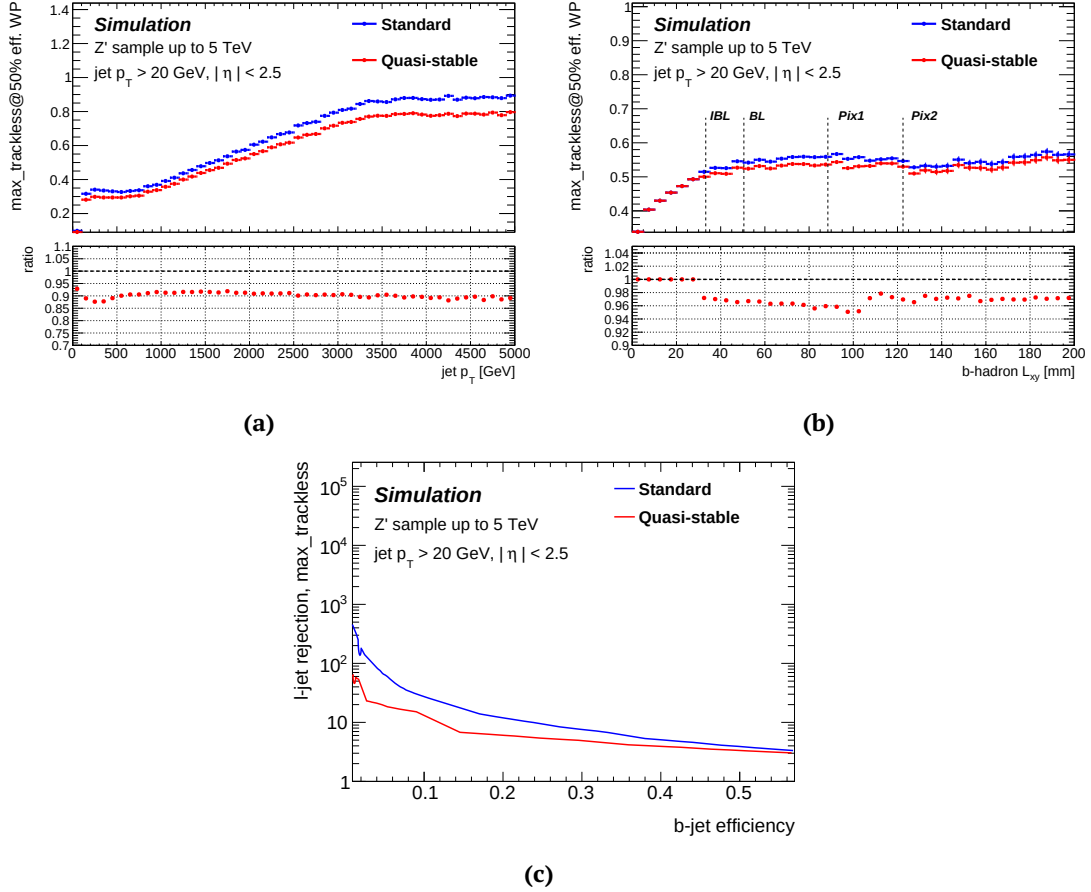


Figure 5.32: b -tagging efficiency at the 50% WP of $\max_trackless$ as a function of jet p_T (a) and b -hadron L_{xy} (b) for the standard (blue) and the quasi-stable (red) simulations. In the bottom pad, the ratio between the quasi-stable and the standard distributions is shown. The statistical uncertainties are not shown because given with the procedure used to derive the standard sample, the two samples are no longer statistically independent. Figure (c) shows the light jet rejection of the tagger as a function of the b -jet efficiency.

this tagger to identify b -jets. The following studies are performed using the quasi-stable sample, in order to overcome the limitation of the derivated standard sample.

Figure 5.34 (a) and (b) show the fraction of b -jets that have been tagged by `max_trackless` at the 50% WP (in blue), those tagged by `max_trackless` at the 50% WP but not by `DL1r` at the 70% WP (in red), those tagged by `DL1r` at the 70% WP but not by `max_trackless` at the 50% WP (in magenta), those tagged by both `max_trackless` at the 50% WP and `DL1r` at the 70% WP (in yellow), and those tagged by `DL1r` at the 70% WP without any requirement on `max_trackless` (in green). A scheme to clarify which b -jets enter in which curve is reported in Figure 5.33.

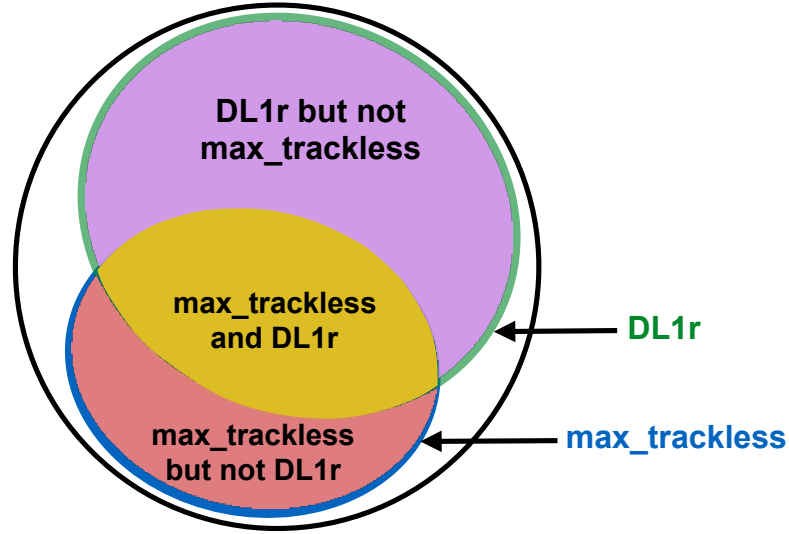


Figure 5.33: Among all b -jets, those tagged by `max_trackless` enter the blue oval which overlaps with the one containing the events tagged by `DL1r` shown in green. The events that are tagged by `max_trackless` but not by `DL1r` enter the red portion, while those tagged by `DL1r` but not by `max_trackless` enter the magenta portion. The events that are tagged by both `DL1r` and `max_trackless` enter the yellow portion while those that are identified neither by `max_trackless` nor by `DL1r` belong to the remaining white part of the entire set.

At low p_T , `DL1r` is expected to have better performances with respect to `max_trackless`. On the other hand, it can be observed that starting from ~ 1 TeV many of the jets not tagged by `DL1r` get recovered by the new discriminant, which can be observed also when evaluating the efficiency as a function of the b -hadron L_{xy} .

These studies triggered an on-going effort within the Trackless b -tagging Working Group for including the hit information in the training of `DL1r`. This will be a key study to evaluate how much the new trackless tagger would help the standard discriminant in the identification of b -jets at very high- p_T .

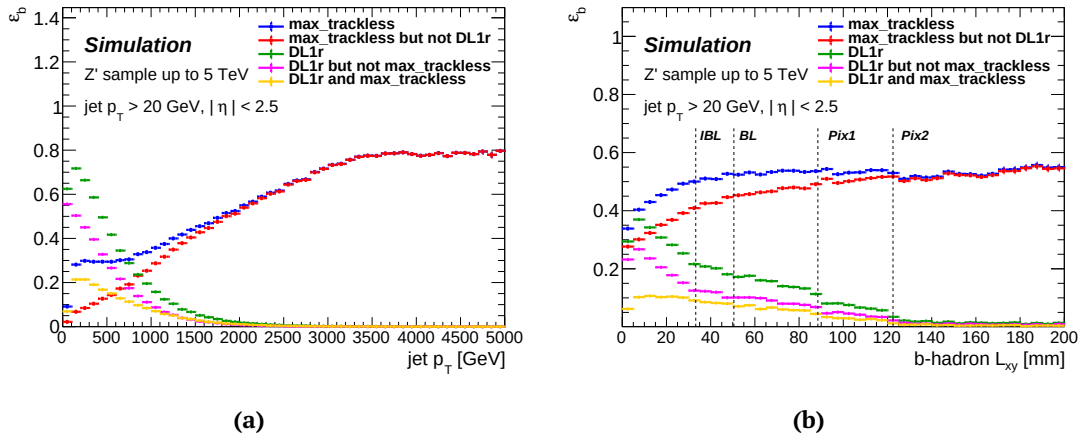


Figure 5.34: b -tagging efficiency for all the events tagged by max_trackless at the 50% WP (blue), those tagged only by max_trackless at the 50% WP and not by DL1r at the 70% WP (red), those tagged by DL1r at the 70% WP but not by max_trackless at the 50% WP (in magenta), those tagged by DL1r at the 70% WP (in green), those tagged by DL1r at the 70% WP but not by max_trackless at the 50% WP, and those tagged by both DL1r at the 70% WP and max_trackless at the 50% WP. The three curves are plotted as a function of jet p_T (a) and b -hadron L_{xy} (b).

Measurement of the $H \rightarrow b\bar{b}$ decay in association with a Vector Boson

6

As described in Chapter 1, the Higgs field predicted by the Brout-Englert-Higgs mechanism gives mass to the charged fermions (quarks and leptons) through interactions involving the Yukawa couplings, with strength proportional to the fermion mass. For this reason, the largest couplings of the Higgs boson are expected to be with the b - and t -quarks. However, the t -quark is too heavy to allow for the decay of an on-shell Higgs boson to a pair of t -quarks, thus the determination of its coupling can be directly probed only via the $t\bar{t}H$ and the tH production modes and, indirectly, via the gluon-gluon fusion process.

On the other hand, the decay of the Higgs boson into a pair of b -quarks has the largest predicted branching ratio ($\sim 58\%$) [11, 12] among all its decay modes, providing the most powerful way to probe the coupling of the Higgs boson to the down-type quarks.

The observation of the $H \rightarrow b\bar{b}$ decay mode is a mandatory step to confirm, or not, the mass generation mechanism as predicted by the Standard Model, but also to set strong limits on possible beyond the Standard Model phenomena in the Higgs sector.

However, this channel is very challenging: the Higgs boson production via gluon-gluon fusion (ggF), having the largest cross section, leads to a pair of low momentum b -quarks which is almost impossible to distinguish from the inclusive QCD $b\bar{b}$ -pair production, whose cross section is about $\sim 10^7$ larger than the ggF one.

To overcome this limitation, it is necessary to study processes that have smaller cross sections but exhibit features not present in the QCD processes. The second largest cross section for the production of the Higgs boson is given by the vector boson fusion (VBF), where the Higgs boson is produced in association with two energetic jets in the forward and backward regions. However, given its fully hadronic final state, both trigger and background modelling remain very challenging.

The solution is given by the associated production of the Higgs boson with a vector boson (VH), W or Z , whose leptonic decays can be exploited for triggering and to highly suppress the QCD background, enhancing the signal sensitivity. Therefore, the VH associated production has been chosen as the main channel for the search of the $H \rightarrow b\bar{b}$ decay mode at the LHC.

In this chapter, the analysis strategy that provided the observation of the $H \rightarrow b\bar{b}$ decay and the VH production mode [13] is summarised in Section 6.1. This will serve as a basis for the definition of the strategy adopted by the $VH, H \rightarrow b\bar{b}$ analysis performed with the full Run 2 dataset of 139 fb^{-1} , which is the main subject of this thesis. The data and the Monte Carlo samples used to simulate the signal and background processes are summarized in Section 6.2, the object reconstruction and the event selection can

be found in Section 6.3 and Section 6.4, respectively. In Section 6.5 the multivariate analysis technique is described, followed by the definition of the statistical procedure used to extract the signal strength in Section 6.6. From Section 6.7 to Section 6.12 the signal and background modelling and the experimental uncertainties entering the fit model are summarised, together with a description of the so called “*smoothing*” and “*pruning*” procedures applied to the systematic uncertainties. Particular attention is given to the description of the modelling of the W +jets background and to the fit studies and statistical model in the 1-lepton channel, as they represent the main contributions of the author to the analysis. Section 6.13 summarises all the cross-checks performed at the fit stage that led to a major change of the fit model itself. In Section 6.14, the results of the main $VH, H \rightarrow b\bar{b}$ analysis will be presented, followed by the results of the diboson and di-jet mass cross-check analyses.

6.1 Observation of the $H \rightarrow b\bar{b}$ decay and VH production

In 2018 the ATLAS Collaboration announced the observation of the Higgs boson decaying to a pair of b -quarks produced in association with a W or a Z boson [13]. The data, corresponding to an integrated luminosity of 79 fb^{-1} , were collected in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ during the first half of Run 2. The combination of the results from the $VH, H \rightarrow b\bar{b}$ analysis and other searches in both Run 1 and Run 2 for the $H \rightarrow b\bar{b}$ decay mode yielded an observed (expected) significance of 5.4 (5.5) standard deviations. Additionally, a combination of Run 2 analyses searching for the Higgs boson produced in association with a vector boson, provided the observation of the VH production mode with an observed (expected) significance of 5.3 (4.8) standard deviations. In the following sections, this result will be addressed to as the “ $H \rightarrow b\bar{b}$ observation analysis”.

Utilising the three leptonic decays of the vector boson, three analysis channels are defined. The 0-lepton channel targets events where the Z boson decays into a pair of neutrinos, $Z \rightarrow \nu\nu$, the 1-lepton channel targets events where the W boson decays into a lepton and a neutrino, $W \rightarrow l\nu$, and the 2-lepton channel identifies events where the Z boson decays into a pair of leptons, $Z \rightarrow l^+l^-$. The corresponding Feynman diagrams of the three analysis channels are illustrated in Figure 6.1.

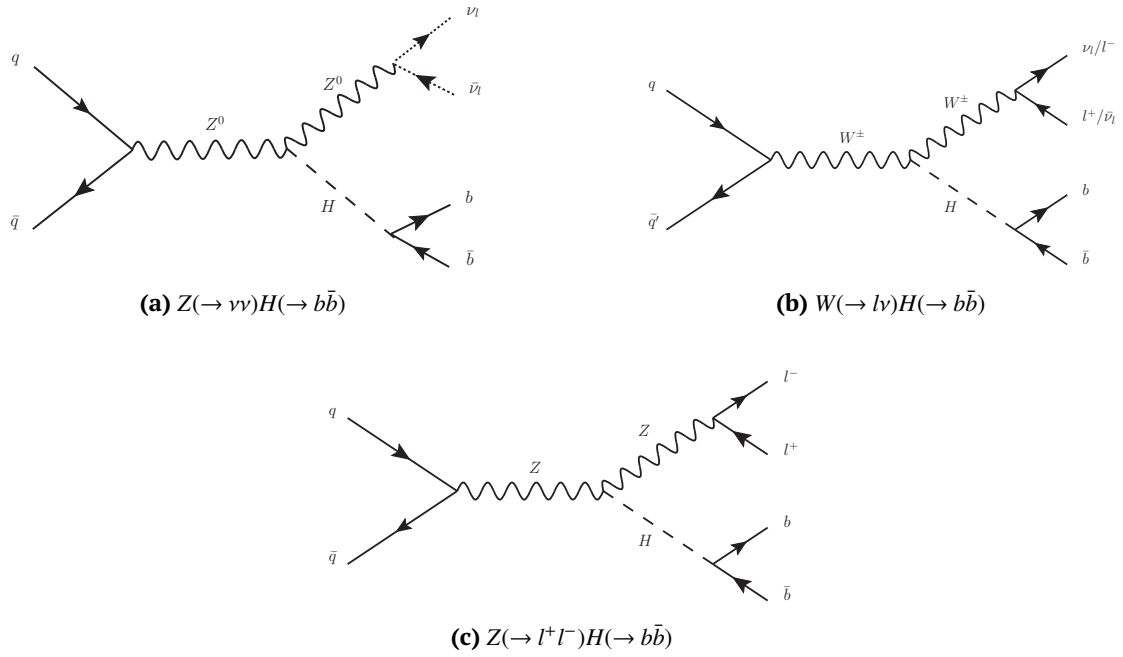


Figure 6.1: Feynman diagrams of the three lepton channels in which the $VH, H \rightarrow b\bar{b}$ events are categorized.

Whilst the QCD background is greatly reduced thanks to the leptonic signature provided by the VH production mode, several SM backgrounds remain. The main backgrounds are the production of top-quark pairs, $t\bar{t}$, the leptonically decaying vector boson, W or Z , produced in association with jets, V +jets, the production of a single top-quark and the diboson production, WW , WZ and ZZ , where one boson decays leptonically and the other decays hadronically. Some Feynman diagrams of the background processes are shown in Figure 6.2.

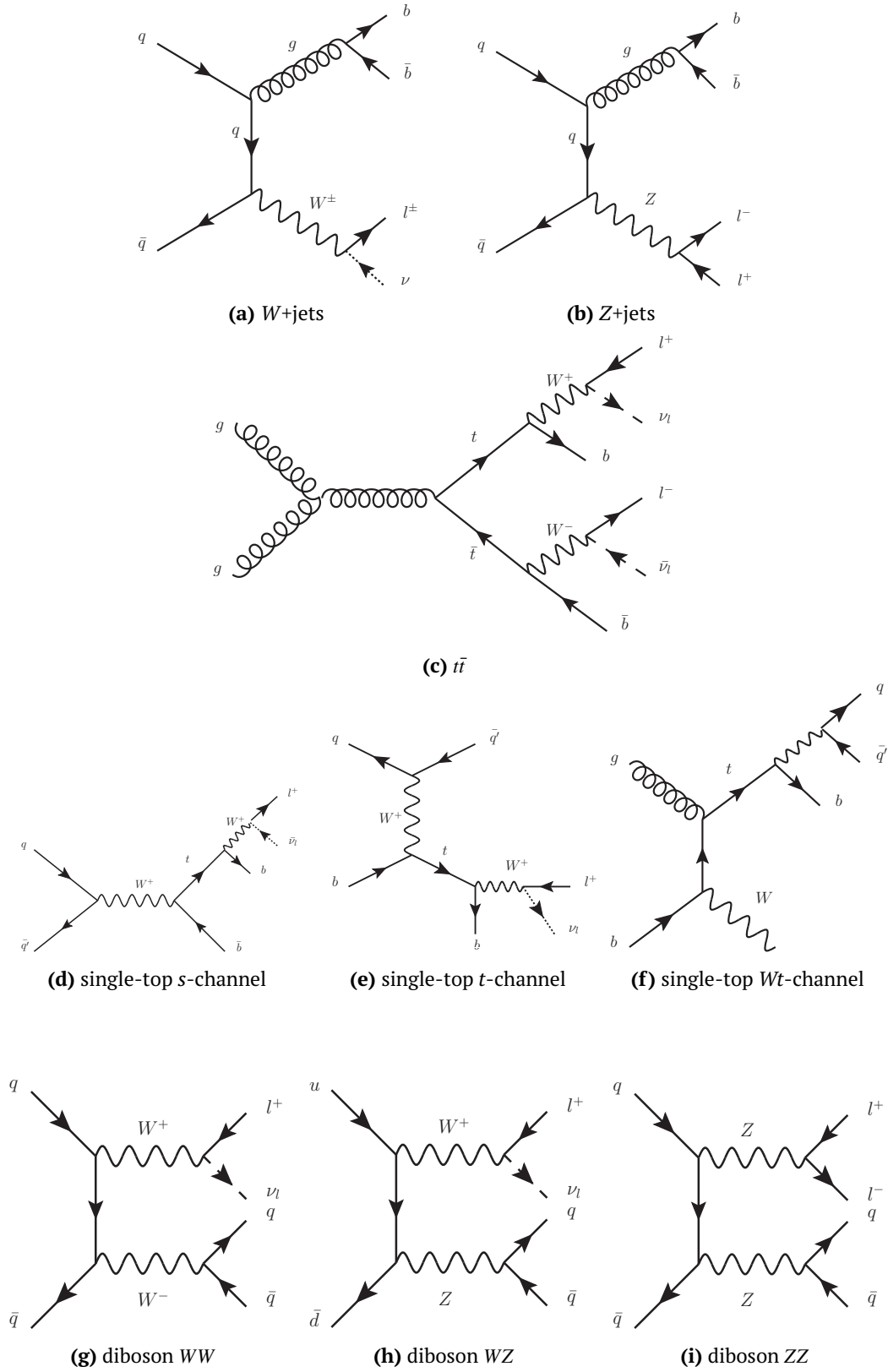


Figure 6.2: Feynman diagrams of the main backgrounds of the $VH, H \rightarrow b\bar{b}$ analysis.

To reduce the large background contributions, the $VH, H \rightarrow b\bar{b}$ analysis is performed in the high p_T regime in order to have a better signal-to-background ratio, as shown in Figure 6.3 by the spectra of the vector boson transverse momentum, p_T^V , in the 2-lepton channel and in the region with two selected jets, both identified as b -jets.

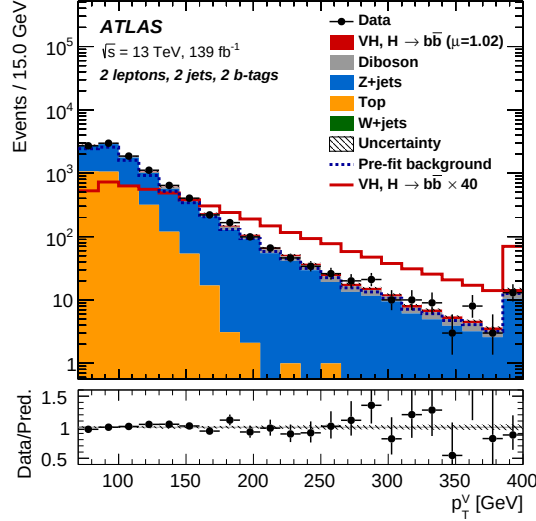


Figure 6.3: Distributions of the p_T^V variable in the 2-lepton channel for 2 b -tagged jets, 2-jet events and inclusively in the $p_T^V > 75$ GeV regions. The data points in black with the associated error bars are compared to the Monte Carlo prediction of the backgrounds which are shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The bottom pad shows the ratio of the data to the combined signal and backgrounds. It can be observed that at the p_T^V spectra of the signal (continuous red line) at high p_T is harder than the background one [146].

The main handle to discriminate between the signal from the sum of all backgrounds is the invariant mass, m_{bb} , of the two b -jets in the final state. The analysis performed using this single variable as discriminant will be referred to in the following as *di-jet* mass analysis. Figure 6.4 shows the m_{bb} distribution where the WH and the ZH channels are combined and all the backgrounds, except for the diboson, are subtracted from data. A clear peak can be observed at the mass of the Higgs boson which is consistent in shape and rate with the theoretical prediction for the $H \rightarrow b\bar{b}$ decay process.

However, the sensitivity of the di-jet mass analysis with 79 fb^{-1} is not sufficient to claim the observation of the $H \rightarrow b\bar{b}$ decay. Therefore, it is used as a cross-check to a multivariate technique using a Boosted Decision Tree (BDT) to enhance the signal over background ratio by making use of other kinematic variables that show significant differences between the signal and the backgrounds. Two sets of BDT are constructed, the nominal BDT_{VH} is used to separate the $VH, H \rightarrow b\bar{b}$ signal from the sum of the expected backgrounds, while the second one (BDT_{VZ}) is used to validate the analysis strategy and the background modelling of the main analysis by extracting the $VZ, Z \rightarrow b\bar{b}$ process. In both cases, the output distributions of the multivariate analyses are used as discriminants

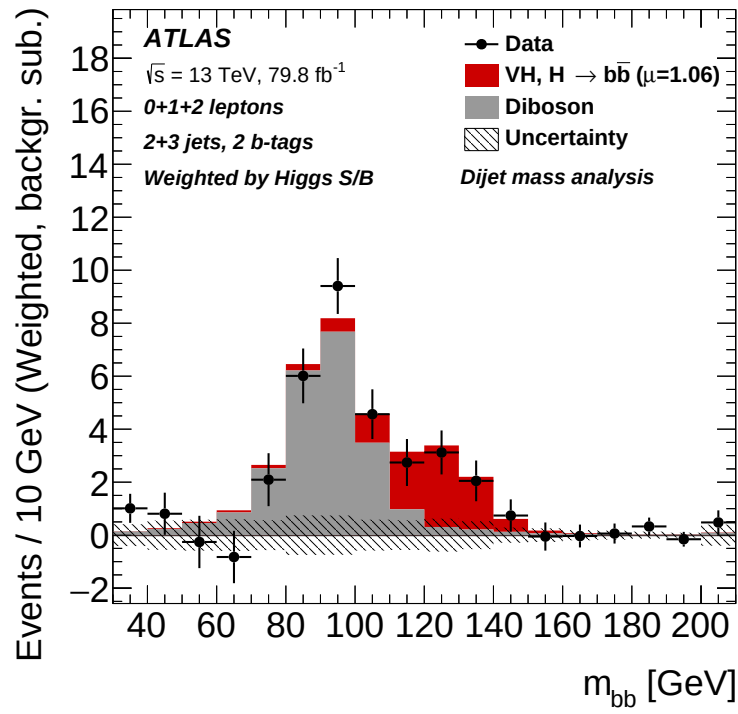


Figure 6.4: Distribution of m_{bb} from all search channels combined after subtraction of all backgrounds except for WZ and ZZ production. The data points in black with the associated error bars are compared to the Monte Carlo expectations from the production of WZ and ZZ (in grey) and of WH and ZH (in red) [13].

to a binned maximum-likelihood fit, referred to as the global likelihood fit, which allows the signal yield and the background normalisations to be extracted.

The same analysis strategy has been adopted for the full Run 2 $VH, H \rightarrow b\bar{b}$ analysis which will be described in details in the next sections.

6.2 Data and simulated samples

The proton-proton collision data used in this analysis was collected by the ATLAS detector during the full Run 2 of the LHC and corresponds to a dataset of 139 fb^{-1} at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. Events have to pass a filter requirement called Good Run List (GRL) to ensure their quality and that all systems of the ATLAS detector were operating well when the events were recorded.

All backgrounds are simulated using Monte Carlo (MC) generators, with the exception of the multi-jet background, present only in the 1-lepton channel, and the top background ($t\bar{t}$ and single-top) in the 2-lepton channel, which are estimated with data-driven techniques. All simulated processes are normalised to the most accurate theoretical prediction of their cross-section and were generated to next-to-leading-order (NLO) accuracy at least, except for the $gg \rightarrow ZH$ and $gg \rightarrow VV$ processes which are generated at LO. Table 6.1 summarises the matrix element (ME), the parton shower (PS) and the underlying event (UE) models, the Parton Distribution Function (PDF) sets and the corresponding cross-section for each signal and background process. All samples of simulated events are passed through the ATLAS detector simulation [15] which is based on the GEANT4 software [16].

Table 6.1: Monte Carlo generators used for the simulation of the signal and background processes and corresponding cross-sections at $\sqrt{s} = 13$ TeV. The acronyms ME, PS and UE stand for matrix element, parton shower and underlying event, respectively.

Process	ME generator	ME PDF	PS and Hadronisation	UE model tune	Cross-section order
Signal, mass set to 125 GeV and $b\bar{b}$ branching fraction to 58%					
$q\bar{q} \rightarrow WH \rightarrow \ell\nu b\bar{b}$	Powheg-Box v2 [147] + GoSam [148] + MiNLO [149, 150]	NNPDF3.0NLO [56]	Pythia8.212 [151]	AZNLO [152]	NNLO(QCD)+ NLO(EW) [153–159]
$q\bar{q} \rightarrow ZH \rightarrow \nu\bar{\nu} b\bar{b} / \ell\ell b\bar{b}$					
$g\bar{g} \rightarrow ZH \rightarrow \nu\bar{\nu} b\bar{b} / \ell\ell b\bar{b}$					
	Powheg-Box v2	NNPDF3.0NLO	Pythia8.212	AZNLO	NLO+NNLL [160–164]
Top quark, mass set to 172.5 GeV					
$t\bar{t}$	Powheg-Box v2 [165]				NNLO+NNLL [167]
s -channel single top	Powheg-Box v2 [168]	NNPDF3.0NLO [56]	Pythia8.230 [61]	A14 [166]	NLO [169]
t -channel single top					
Wt	Powheg-Box v2 [170]				Approximate NNLO [171]
Vector boson + jets					
$W \rightarrow \ell\nu$	Sherpa 2.2.1 [64, 172, 173]	NNPDF3.0NLO [56]	Sherpa 2.2.1 [174, 175]	Default	NNLO [176]
$Z/\gamma^* \rightarrow \ell\ell$					
$Z \rightarrow \nu\nu$					
Diboson					
$q\bar{q} \rightarrow WW$	Sherpa 2.2.1 [64, 172, 173]	NNPDF3.0NLO [56]	Sherpa 2.2.1 [174, 175]	Default	NNLO [176]
$q\bar{q} \rightarrow WZ$					
$q\bar{q} \rightarrow ZZ$					
$g\bar{g} \rightarrow VV$					

6.2.1 Signal Processes

Nominal Monte Carlo samples The VH signal can be divided into three signatures depending on the number of charged leptons in the final state: $Z(\rightarrow \nu\nu)H(\rightarrow b\bar{b})$, $W(\rightarrow l\nu)H(\rightarrow b\bar{b})$ and $Z(\rightarrow l^+l^-)H(\rightarrow b\bar{b})$ where $l = e, \mu, \tau$. All $q\bar{q}$ -initiated samples are simulated using the POWHEG generator for the ME with the MiNLO (Multileg Improved NLO) procedure applied [150], interfaced to the PYTHIA8 [151] generator for the PS and applying the AZNLO tune [152] with the NNPDF3.0 PDF [56] set. The computation of the one-loop amplitudes is performed with the GoSam automated software [148]. The gg -initiated ZH contribution is simulated using the POWHEG prediction with the matrix element computed at LO in QCD, interfaced with PYTHIA8 and applying the AZNLO tune [152] with the NNPDF3.0 PDF [177] for both the matrix element and the parton shower. The cross sections for WH and ZH are calculated at NNLO in QCD [154] [156] and NLO in electroweak (EW) [178] for the quark-induced processes and at NLO+NLL QCD for the gluon-induced ones.

Alternative Monte Carlo samples In order to derive the uncertainties to model the signal processes, the nominal POWHEG + MiNLO + PYTHIA8 samples are compared to the following alternative samples:

- POWHEG + MiNLO + HERWIG7 for the UE and PS models;
- POWHEG + PYTHIA8 with variations of the multiple parton interaction (MPI) cut-off compared to the nominal value;
- POWHEG + PYTHIA8 with six variations for the renormalization scale μ_R and the factorization scale μ_F ;
- POWHEG + PYTHIA8 with variations of two sets of eigentunes for the AZNLO [152] parton shower tune covering primordial k_t and initial state radiation (ISR) cut-off variations [152].

6.2.2 Top-quark Pair Production

The top-quark pair production is the largest background in the $VH, H \rightarrow b\bar{b}$ analysis, being present in each of the three lepton channels. However, its treatment differs between the 0-/1-lepton and 2-lepton channels as the firsts contain $t\bar{t}$ events with at least one non-reconstructed object, while the $t\bar{t}$ events in the 2-lepton channel are mostly fully reconstructed. In addition, studies on fully reconstructed $t\bar{t}$ events in the 0- and 1-lepton channels have shown that the events with two b -tagged jets consist mainly of $b\bar{b}$ -pairs but also a non-negligible amount of bc events.

As a consequence, a MC based prediction is used in the 0-/1-lepton and a fully data-driven approach is utilized in the 2-lepton channel, making use of a dedicated “top- $e\mu$ ” control region 99% pure in the $t\bar{t}$ and single-top background. More details about the construction

of this control region will be given in Section 6.4. The following MC samples are used to describe the $t\bar{t}$ background.

Nominal Monte Carlo samples The default MC generator for the $t\bar{t}$ process is POWHEG v2 + PYTHIA8, which utilises the POWHEG NLO matrix element generator [165, 179] interfaced to PYTHIA8.210 [61] using the A14 tune [166] for the parton shower, hadronization, underlying event, and multiple parton interactions. In addition to this, the sample utilises the NNPDF3.0 (NLO), and NNPDF2.3 (LO) PDF set in the ME calculation and PS, respectively.

Alternative Monte Carlo samples To assess modelling uncertainties for the $t\bar{t}$ background, the nominal generator POWHEG + PYTHIA8 is compared to the following alternative samples:

- POWHEG + HERWIG7: same setup for POWHEG as for the nominal sample, but with PS, hadronisation, and UE simulated with HERWIG7.0 [60, 180] with the H7UE tune [60];
- MADGRAPH5_aMC@NLO + PYTHIA8.2: using MADGRAPH5_aMC@NLO [181] for the hard scattering generation at NLO precision and PYTHIA8.2 for the PS, hadronisation, UE, and MPI simulation;
- POWHEG + PYTHIA8: same setup as for the nominal sample but with doubled μ_R and μ_F , $\text{hdamp}^1 = 1.5 \cdot m_{top}$ and A14 tune down variation for the *low radiation samples* and halved μ_R and μ_F , $\text{hdamp} = 3.0 \cdot m_{top}$ and A14 up variation for the *high radiation samples*².

A filtering at generator level is applied using particle-level information. Samples defined as *non-all-hadronic* are required to have at least one W boson decaying leptonically, while the *di-lepton $t\bar{t}$* samples are required to have both W bosons decaying leptonically. The $t\bar{t}$ samples are then normalised using cross sections calculated at NNLO+NNLL [167].

6.2.3 Vector Boson + Jets Production

The production of a V ($V=W$ or Z) boson in association with jets is the second main background in the 0- and the 1-lepton channels.

Nominal Monte Carlo samples The V +jets samples are simulated using SHERPA 2.2.1 [182] interfaced with the NNPDFs [177] for both the ME calculation and the PS tuning. SHERPA 2.2.1 adopts a 5-flavour scheme in the ME where both b -quarks and c -quarks are treated as massless. Massive quarks can be generated at parton shower level relying on the algorithms provided internally by the generator. The precision of the simulation depends

¹hdamp corresponds to the transverse momentum scale around which the Sudakov resummation becomes negligible. Smaller values of hdamp correspond to a larger suppression of the hardest emission from the matrix element to the cross section.

² m_{top} is defined as the invariant mass of the lepton, the reconstructed neutrino (identified as the E_T^{miss}) and the b -tagged jet that yields the lowest mass value.

on the parton multiplicity: SHERPA 2.2.1 is NLO in the ME up to two extra partons and LO with 3 or 4 extra partons. The merging of different parton multiplicities is achieved through a matching scheme based on the CKKW-L [183, 184] technique using a matching scale of $Q_{cut} = 20$ GeV.

In order to generate a sufficient number of events with large transverse momentum, V +jets samples are split according to the p_T of the vector boson and the transverse energy of the event H_T , introducing a cut at generation level and producing samples for different slices in $\max(H_T, p_T^V)$. An exception is made for $Z(\rightarrow \nu\nu) + jets$ samples that are produced with only a p_T^V filter, since it was found to be more efficient in enriching the high p_T^V phase space than the nominal $\max(H_T, p_T^V)$ filter. Moreover, to obtain a sufficient number of events with heavy-flavour final states, flavour filters are applied at the generator level: *b-filter* samples require at least one *b*-jet, *c-filter* + *b-veto* samples contain at least 1 *c*-jet, and a veto is imposed on events passing the *b-filter*, and *c-veto* + *b-veto* veto the events passing the *b-filter* and the *c-filter* + *b-veto* one.

Alternative Monte Carlo samples MADGRAPH 5 [181] interfaced to PYTHIA8 is used as an alternative sample to simulate the hard-scattering, the parton shower, and the underlying event. It is LO in QCD and it merges together matrix-element calculations with different parton multiplicities, up to 4 additional jets (higher jet multiplicities are modelled by the parton shower algorithm). The merging scheme applied to combine different parton multiplicities is the CKKW-L scheme, as for the nominal sample, but using a merging scale of $Q_{cut} = 30$ GeV.

SHERPA 2.2.1 generator with the same setup as the nominal but with doubled or halved values of the renormalization and factorization scales, or with a different PDF set (specifically, MMHT2014nnlo68cl and CT14nnlo) is used as alternative sample.

SHERPA 2.2.1 samples with different values of the resummation and ckkw scales³ (not yet included as internal weights in the present version of SHERPA 2.2.1) with respect to the nominal were not produced in time to be used in the analysis, but their effect is expected to be small from studies performed for the $H \rightarrow b\bar{b}$ observation analysis [13].

6.2.4 Single-Top Production

The single-top background consists of three channels: *s*-channel, *t*-channel and Wt -channel. The *s*-/*t*-channel events are generated requiring at least one leptonically decaying W , while the Wt -channel events are produced both applying a di-lepton filter (both W s decay leptonically) and inclusively. However, this analysis uses only the inclusive samples.

Nominal Monte Carlo samples The POWHEG v2 generator is used to model the hard scattering and it is interfaced to PYTHIA8 for the simulation of the parton shower, the

³The ckkw scale refers to the matching scale between matrix element and parton shower. The resummation scale corresponds to the scale at which the large logarithms due to the soft gluon emission are absorbed at all perturbative orders.

underlying event and the multiple parton interactions.

Alternative Monte Carlo samples The nominal samples are compared to POWHEG + PYTHIA6 samples with the following alternative generators:

- POWHEG + PYTHIA6 with doubled or halved μ_R and μ_F ;
- POWHEG + PYTHIA6 with “*diagram subtraction*” scheme for the matrix element instead of the nominal “*diagram removal scheme*”⁴;
- POWHEG + HERWIG++ to assess effects on the parton shower model;
- MADGRAPH5_amc@NLO + HERWIG++ that is compared to POWHEG + HERWIG++ to assess the effect on the matrix element.

6.2.5 Diboson Production

The diboson background consists of three processes: ZZ , WW and WZ . The SHERPA 2.2.1 generator interfaced with the NNPDF 3.0 PDF sets is used as nominal generator for both the matrix element and the parton shower and it is compared to:

- POWHEG v2 + PYTHIA8 to assess effects on both the matrix element and the parton shower;
- POWHEG + HERWIG++, generated using the same setup as POWHEG v2 + PYTHIA8 but using HERWIG++ for a different PS, hadronisation, UE and MPI models;
- SHERPA 2.2.1 internal weights for the renormalization μ_R and factorization μ_F scales, which are doubled or halved with respect to the nominal value.

Each diboson process is normalised according to its MC generator predicted cross-section.

6.2.6 Multi-jet Background

The QCD multi-jet background arises either from jet energy mismeasurements, especially in the 0-lepton channel, or from “fake leptons” in the 1-lepton and 2-lepton channels. Fake leptons can be produced by hadronic jets misidentified as electrons, by electrons from photon conversions ($\gamma \rightarrow e^+e^-$) or by electrons or muons from non-prompt weak decays. Studies performed for the $H \rightarrow b\bar{b}$ observation analysis [13] have shown that the contribution of the multi-jet background in the 0-lepton and 2-lepton channels is negligible. In the 1-lepton channel, its contribution is small but non-negligible with respect to the other backgrounds.

Due to the small number of multi-jet events simulated in the analysis phase space and

⁴The NLO diagrams of the Wt process interfere with $t\bar{t}$ at LO and can be thought as a production of a $t\bar{t}$ pair. This raises up the question whether or not it makes sense to consider Wt as an independent process. In the *diagram removal* scheme, the resonant $t\bar{t}$ effects are removed at the amplitude level, while in the *diagram subtraction* scheme, these effects are removed at the cross-section level. By comparing these two approaches one can give an estimation of the interference between Wt and $t\bar{t}$.

to the inaccurate description of fake leptons, a fully data-driven approach is used to estimate the multi-jet background using a template fit method. Details of the estimation of the multi-jet background will be provided in Section 6.10.

6.3 Object Reconstruction

The event topologies considered in the $VH, H \rightarrow b\bar{b}$ analysis contain zero, one or two charged leptons and two b -tagged jets. In the following, the reconstruction of the physics objects for each channel is described.

Triggers Events in the 0-lepton channel are selected using the lowest unscaled E_T^{miss} triggers with online thresholds of 70 GeV for data recorded in 2015, 90 GeV and 110 GeV for data recorded in 2016 and 110 GeV for those recorded in 2017 and 2018.

Events in the 1-lepton electron sub-channel and in the 2-lepton channel are triggered using the lowest unscaled single lepton trigger having p_T thresholds increasing from 25 GeV in 2015 to 26 GeV in 2016-2018. In the 1-lepton muon sub-channel channel, the single muon trigger has an efficiency of 80%. This is mainly due to the reduced $|\eta|$ ($|\eta| < 2.4$ plus gaps at $|\eta| < 0.1$) coverage of the muon chambers. To recover for the loss of efficiency in the 1-lepton channel, the same E_T^{miss} trigger as the 0-lepton channel is used to trigger events with one single muon. In fact, these triggers use only the information from the electromagnetic calorimeter firing the trigger even for high p_T muons which release only a small amount of energy inside the calorimeter cells.

Leptons The $VH, H \rightarrow b\bar{b}$ analysis is categorized into three lepton channels depending on the number of leptons (e or μ) in the final state.

The *VH-loose* selection criteria are designed to maximize the electron and muon selection efficiency. Both are required to have $p_T > 7$ GeV, the electron (muon) should be in the range of $|\eta| < 2.47$ ($|\eta| < 2.7$), the impact parameter significance should be less than 5σ (3σ) and $|\Delta z_0 \sin \theta| < 0.5$ mm to reject tracks from pile-up. Loose track isolation ([106] for electrons, [108] for muons) is also applied to reduce jets faking electrons (muons). Events with zero *VH-loose* leptons end up in the 0-lepton channel while for the 1-lepton and 2-lepton channels, tighter selection criteria are applied to reduce the multi-jet component. In the 2-lepton channel, the *ZH-signal* requires one or more electrons (muons) with $|\eta| < 2.5$ and $p_T > 25$ GeV ($p_T > 27$ GeV). If only one *ZH-signal* lepton is found, the other one is a *VH-Loose* lepton. In the 1-lepton channel, the selection of isolated leptons is crucial to reduce the large QCD multi-jet background: exactly one electron (muon) is required with tight (medium) identification and calorimeter-based (track-based) tight isolation criteria [106, 108] keeping 95% of signal efficiency.

Table 6.2 (6.3) presents a summary of electron (muon) selection criteria, for *VH-loose*, *ZH-signal* and *WH-signal*, respectively.

Table 6.2: Electron selection criteria; the events with zero VH-Loose leptons end up in the 0-lepton channel. In the 1-lepton channel, exactly one electron with WH-signal selection is required, in the 2-lepton channel, one or more electrons with ZH-signal selection are required. If only one, the other satisfies the VH-loose selection. The isolation requirements are those defined in Section 3.2.4 for the electrons.

Electron Selection	p_T	η	ID	d_0^{sig} w.r.t. B-layer	$ \Delta z_0 \sin \theta $	Isolation
VH-Loose	>7 GeV	$ \eta < 2.47$	LH Loose	< 5	< 0.5 mm	Fixed-Cut Loose
ZH-Signal	>27 GeV	$ \eta < 2.47$	LH Loose	< 5	< 0.5 mm	Fixed-Cut Loose
WH-Signal	>27 GeV	$ \eta < 2.47$	LH Tight	< 5	< 0.5 mm	HighPtCaloOnly

Table 6.3: Muon selection criteria; the events with zero VH-Loose leptons end up in the 0-lepton channel. In the 1-lepton channel, exactly one muon with WH-signal selection is required, in the 2-lepton channel, one or more muons with ZH-signal selection are required. If only one, the other satisfies the VH-loose selection. The isolation requirements are those defined in Section 3.3.3 for the muons.

Muon Selection	p_T	η	ID	d_0^{sig} w.r.t. B-layer	$ \Delta z_0 \sin \theta $	Isolation
VH-Loose	>7 GeV	$ \eta < 2.7$	LH Loose	< 5	< 0.5 mm	Fixed-Cut Loose
ZH-Signal	>27 GeV	$ \eta < 2.5$	LH Loose	< 3	< 0.5 mm	Fixed-Cut Loose
WH-Signal	>27 GeV	$ \eta < 2.5$	LH Medium	< 3	< 0.5 mm	HighPtTrackOnly

Jets Jets are reconstructed from topological clusters [185] using the anti- k_t algorithm [113] with radius $R = 0.4$ (see Chapter 3). Jets can be identified as forward or signal jets. Signal jets are required to have $p_T > 20$ GeV, $|\eta| < 2.5$, and to satisfy the cuts on the JVT needed to reject low p_T jets likely originating from pile-up. Forward jets are required to have $p_T > 30$ GeV and $2.5 \leq |\eta| < 4.5$.

b -tagging requirements The Higgs boson candidate is reconstructed from jets arising from b -quarks, therefore it is mandatory to be able to identify the b -jets with respect to the light-, c - and τ -jets. The algorithm used to perform b -jet tagging is the MV2c10 algorithm which was described in Section 4.3.2. A fixed-cut requirement at the 70% working point selects the 2-tag events entering the signal region. In the case of the 3-jet category, a veto is imposed on the third jet so that it does not pass the tagging requirement. A truth tagging procedure, as described in Section 4.5, is applied to the WW samples⁵ and to V +jets events not containing b -jets. The $t\bar{t}$ and V +jets events containing b -jets undergo a *hybrid tagging* procedure which consists of applying direct tagging to b -jets and truth tagging to non b -jets components. The procedure can be described in three main steps:

⁵To mimic the signal signature, one W has to decay hadronically and the other leptonically. Therefore, only one of the two jets from the hadronically decaying W is a real b -jet. This is not the case for the ZZ and the WZ backgrounds in which the hadronic decay of the Z is dominated by two real b -jets.

- Divide the jets of each events in two groups, depending on their truth flavour: one with only b -jets and one with only non b -jets;
- All b -jets in the first group are direct tagged;
- The c -jets and l -jets in the second group are truth tagged imposing a number of tagged jets equal to the difference between the number of required jets in the signal region and the number of b -jets in the first group passing the b -tagging requirement (cut on the MV2c10 discriminant).

With this procedure, a reduction of 60% and 90% in the statistical uncertainty of the $V + bc$ and $V + bl$ samples, respectively, is obtained.

b -jet energy corrections Dedicated b -jet energy corrections are applied to improve the signal mass resolution which is deteriorated by the inaccuracies in the b -jet measurements.

The b -quarks can decay directly leptonically ($b \rightarrow l\nu X$) $\sim 22\%$ of the times or indirectly leptonically through the $b \rightarrow c \rightarrow l\nu X$ chain, l being μ or e , $\sim 20\%$ of the times. Therefore, about 20% of the reconstructed b -jets have a semileptonic decay of the b -hadron which involves a muon. Being a minimum ionizing particle (MIP), the energy lost by the muon inside the calorimeters is very small. Therefore, if a reconstructed muon having $p_T > 4$ GeV and passing the medium quality cuts is found within $\Delta R(j, \mu) < \min(0.4, 0.04 + 10/p_T^\mu)$ from the b -jet axis, the four momentum of the muon is added to the total momentum of the jet and the energy deposited by the muon in the calorimeter is removed. This correction is called *muon-in-jet* correction.

An additional correction applied to the b -jets is called *PtReco* and is applied after the *muon-in-jet* correction. This is evaluated from the ratio of the particle-level jet energy to the reconstructed one.

A *kinematic fit* (KF) is applied in the 2-lepton channel where no E_T^{miss} is present and the kinematics of the $Z(\rightarrow l^+ l^-)H(\rightarrow b\bar{b})$ events can be fully reconstructed. Since the di-lepton mass resolution is much better than the one of the di-jet system, it is possible to improve the b -jet energy correction by constraining the final state objects to be balanced in the transverse plane. To do this, the total transverse momentum is constrained to be zero with a width of 9 GeV mass determined by measurements in simulated ZH events based on particle-level information. The leptons and b -jets are then smeared by Gaussian distributions correspondent to the expected resolution while, if present, additional jets are smeared with dedicated asymmetric transfer functions.

The final di-jet mass resolution is shown in Figure 6.5 where the effect of sequentially b -jet energy corrections is summarised.

Vector Boson The missing transverse momentum of the vector boson corresponds to the E_T^{miss} in the 0-lepton channel, the vectorial sum of the transverse momentum of the charged lepton and the E_T^{miss} in the 1-lepton channel and the vectorial sum of the transverse momenta of the leptons in the 2-lepton channel.

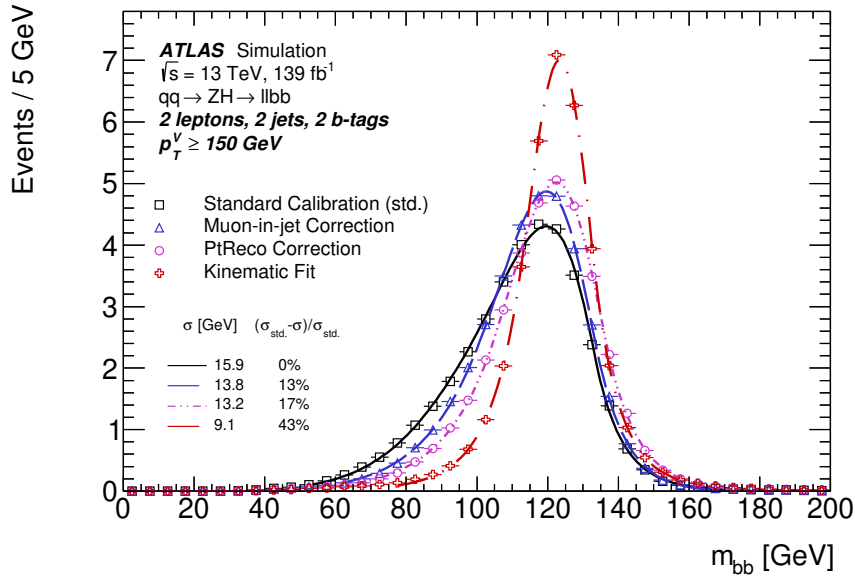


Figure 6.5: Comparison of the m_{bb} distributions as additional corrections are applied to the jet energy shown for simulated events in the 2-lepton channel with 2-jets and $p_T^V > 150$ GeV inclusively in signal and control regions. A fit to a Bukin function [186] is superimposed on each distribution to highlight the resolution improvements [187].

6.4 Event Selection

The $VH, H \rightarrow b\bar{b}$ analysis is categorized into three lepton channels depending on the number of leptons (e or μ) in the final state. In all channels, the events are required to have exactly two b -tagged jets which form the Higgs boson candidate. At least one b -tagged jets is required to have $p_T > 45$ GeV. Events with exactly two signal jets form the 2-jet category while events with exactly three jets form the 3-jet category. In the 0-lepton and 1-lepton channel, only one additional jet is allowed in order to suppress the $t\bar{t}$ background. In the 2-lepton channel any number of additional jets is accepted to increase the signal acceptance.

The vector boson transverse momentum p_T^V corresponds to the E_T^{miss} in the 0-lepton channel, therefore a cut on $E_T^{\text{miss}} > 150$ GeV is required to select events containing a Z boson decaying to a pair of neutrinos. The analysis is split into three p_T^V regions to increase the sensitivity because of different signal over background ratios: a medium p_T^V region only in the 2-lepton channel, $75 < p_T^V < 150$ GeV, a high p_T^V region $150 < p_T^V < 250$ GeV and an extreme p_T^V region $p_T^V > 250$ GeV. The split at $p_T^V = 250$ GeV was introduced in order to synchronize the reconstructed categories of the analysis with the truth categories of the Simplified Template Cross Section (STXS) bins in which the cross section of the $H \rightarrow b\bar{b}$ process is calculated. Details of the STXS scheme will be provided in Section 6.8.1.

In the 0-lepton channel high E_T^{miss} in multi-jet events arises from mismeasured jets in the calorimeters. Such events are efficiently removed by the *anti-QCD* requirements reported in Table 6.4. As explained in Chapter 3, the missing transverse momentum, E_T^{miss} , is reconstructed as the negative vector sum of the transverse momenta of the hard objects

(leptons, photons, hadronically decaying τ -leptons and jets) using both calorimeter-based and track-based information, and a track-based soft-term calculated as the negative vector sum of the p_T of tracks matched to the primary vertex but not associated with a reconstructed lepton or jet. The track-based missing transverse momentum, $E_{T,\text{trk}}^{\text{miss}}$ in Table 6.4 differs from the nominal E_T^{miss} as it is calculated using only tracks reconstructed in the inner tracking detector and matched to the primary vertex. This variable is less sensitive to pile-up effects but cannot account for neutral particles. The values of the cuts are tuned in such a way that the remaining fraction of multi-jet contamination is of the order of 1% of the total background. In the 1-lepton channel a cut on $E_T^{\text{miss}} > 30$ GeV is applied in the electron sub-channel.

Table 6.4: Anti-QCD cuts required to suppress the multi-jet background in the 0-lepton channel. The Higgs boson candidate is indicated by the bb .

$\min[\Delta\phi(E_T^{\text{miss}}, \text{jets})]$	$> 20^\circ$ (2 jets) , $> 30^\circ$ (3 jets)
$\Delta\phi(E_T^{\text{miss}}, bb)$	$> 120^\circ$
$\Delta\phi(b_1, b_2)$	$< 140^\circ$
$\Delta\phi(E_T^{\text{miss}}, E_{T,\text{trk}}^{\text{miss}})$	$< 90^\circ$

To better constrain the modelling of the backgrounds, three control regions are defined. The first two, enriched in V +jets and $t\bar{t}$, were built in the 1-lepton channel using continuous cuts on $\Delta R(\vec{b}_1, \vec{b}_2)$ (ΔR_{bb} for simplicity) between the two selected b -jets as a function of p_T^V . A lower and an upper contour cut separate the signal region from a Low- ΔR CR and a High- ΔR CR, respectively. An illustration of the procedure is shown in Figure 6.6.

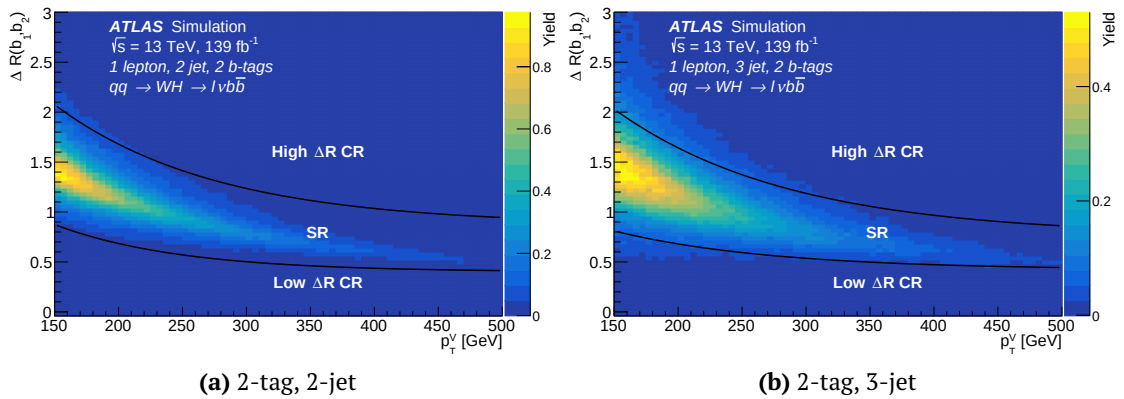


Figure 6.6: Signal yield distribution of the ΔR_{bb} between the two b -tagged jets as a function of the p_T^V in the 1-lepton channel for 2-tag jets events, in the 2-jet (a) and exactly 3-jet (b) categories in the high- p_T region. The black lines show the continuous lower and upper cuts on ΔR_{bb} used to categorise the events into the signal and control regions [146].

In order to have a consistent definition among the three channels, the same contours

also define these two CRs in the 0-lepton and 2-lepton channels. The upper cut which defines the High- ΔR CR is designed in the 1-lepton channel to keep up to 95% of the signal in the 2-jet category and up to 85% in the 3-jet category. As a result, this region is enriched in $t\bar{t}$ in the 0-lepton channel, in $t\bar{t}$ and single-top in the 1-lepton channel and in Z +jets events in the 2-lepton channel. Further control on the modelling of the $t\bar{t}$ and single-top backgrounds is provided by the split in p_T^V regions.

The VZ background is a resonant process as the VH signal but with a smaller mass scale, therefore, the VZ peak sits at the left of the signal peak in the ΔR_{bb} distribution, as shown in Figure 6.7 for the three lepton channels.

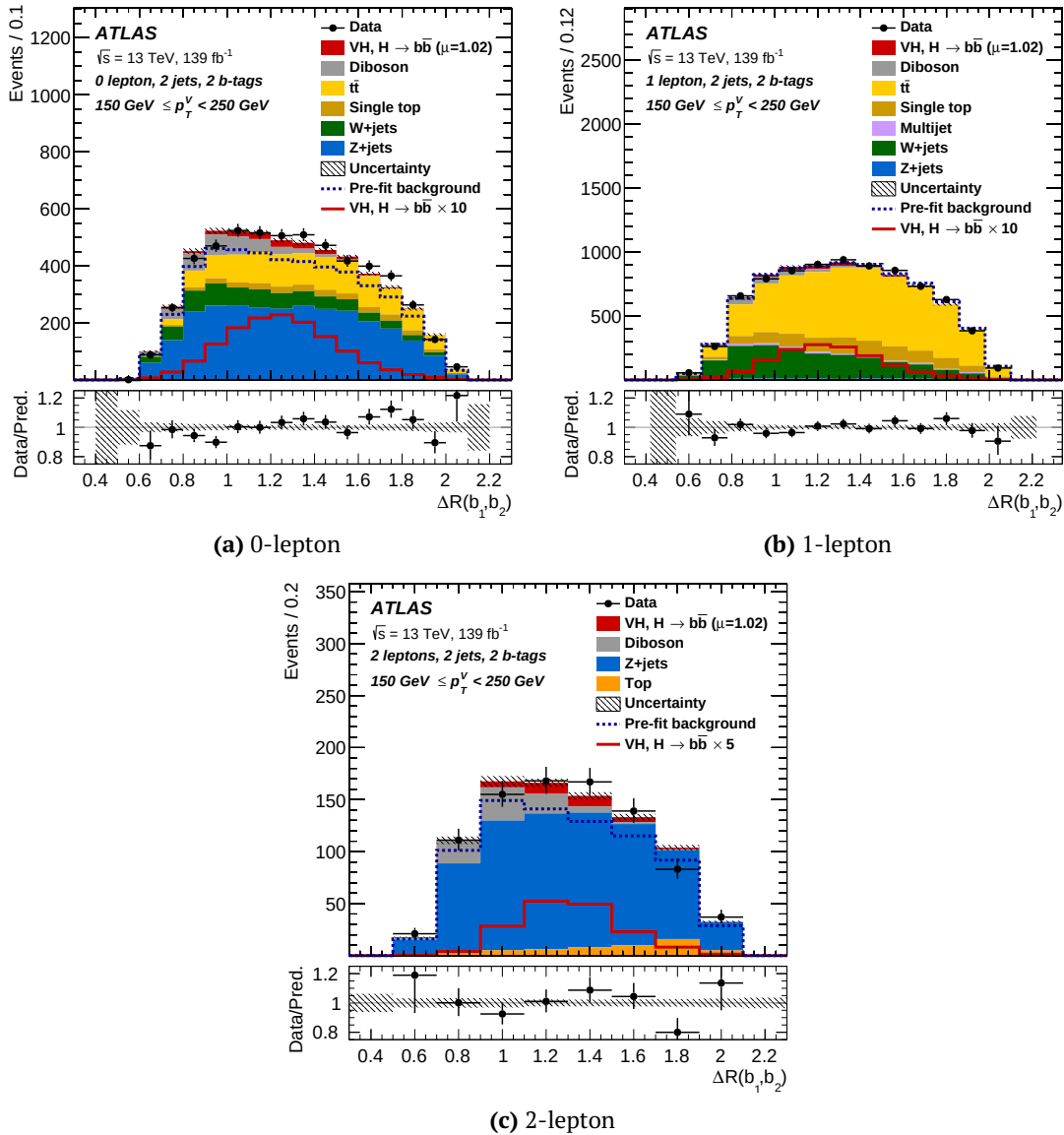


Figure 6.7: Distribution of ΔR_{bb} showing the resonant peaks of the VH signal and the VZ background [146].

The lower cut which defines the Low- ΔR CR is designed to keep the majority of diboson events in the signal region without compromising the signal purity. This is motivated by

the fact that the same analysis regions are also used in the diboson cross-check analysis. This region is enriched in W +jets events in the 1-lepton channel and in Z +jets events in the 2-lepton channel. The contribution of the W +jets background is enhanced when considering the 0-lepton channel. Also in this region, the split in p_T^V regions helps to constrain the modelling of the V +jets backgrounds.

Whereas in the $H \rightarrow b\bar{b}$ observation analysis the di-jet mass cross-check analysis required further cuts to remove the backgrounds under the m_{bb} peak, this new definition of SR and CRs allows to harmonise the di-jet mass analysis with the multivariate one without compromising its sensitivity.

The $t\bar{t}$ background is one of the dominant backgrounds in this analysis and it is flavour symmetric. Therefore, in addition to the Low- ΔR CR and High- ΔR CR, a CR enriched in top background called “top $e\mu$ -CR” is built in the 2-lepton channel by applying the same event selection of the analysis but requiring di-lepton events with different flavours ($e\mu$ or μe). This way, the only difference between the phase spaces of the analysis and the top control region is the lepton flavour; it has been demonstrated that the $t\bar{t}$ kinematic distributions of the two phase spaces are highly compatible with each other.

The requirement on the lepton flavour eliminates most of the backgrounds so that the CR is 99% pure in top ($t\bar{t}$ and Wt) events. The top $e\mu$ -CR is used to extract the $t\bar{t}$ yield in the 2-lepton channel through a fully data-driven approach. This eliminates all the systematic uncertainties related to the $t\bar{t}$ modelling in the 2-lepton channel leaving only the data statistics. Details of the data-driven top background estimate are given in Section 6.10.

A summary of the object reconstruction and event selection of the $VH, H \rightarrow b\bar{b}$ analysis is shown in Table 6.5.

Table 6.5: Summary of the event selection and categorisation in the 0-, 1- and 2-lepton channels. Here H_T stands for the scalar sum of the p_T of the jets in the event.

Selection	0-lepton	1-lepton		2-lepton
		e sub-channel	μ sub-channel	
Trigger		Single lepton	E_T^{miss}	Single lepton
Leptons	0 loose leptons	Exactly 1 <i>tight</i> electron 0 additional <i>loose</i> leptons $p_T > 27$ GeV	Exactly 1 <i>tight</i> muon 0 additional <i>loose</i> leptons $p_T > 25$ GeV	Exactly 2 <i>loose</i> leptons $p_T > 27$ GeV Same-flavour Opposite-sign charges ($\mu\mu$)
E_T^{miss}	> 150 GeV	> 30 GeV	–	–
m_{ll}	–	–	–	$81 \text{ GeV} < m_{ll} < 101 \text{ GeV}$
Signal jets		> 20 GeV for $ \eta < 2.5$		
Forward jets		$\text{JVT} > 0.59$ for $ \eta < 2.4$ and > 0.11 for $2.4 < \eta < 2.5$ if $p_T < 120$ GeV		
b -jets		> 30 GeV for $2.5 < \eta < 4.5$		
Leading b -tagged jet p_T		Exactly 2 b -tagged jets > 45 GeV		
Jet categories	Exactly 2 / Exactly 3 jets	Exactly 2 / Exactly 3 jets		Exactly 2 / ≥ 3 jets
H_T	> 120 GeV (2 jets), > 150 GeV (3 jets)	–		–
$\min[\Delta\phi(E_T^{\text{miss}}, \text{jets})]$	$> 20^\circ$ (2 jets), $> 30^\circ$ (3 jets)	–		–
$\Delta\phi(E_T^{\text{miss}}, b\bar{b})$	$> 120^\circ$	–		–
$\Delta\phi(b_1, b_2)$	$< 140^\circ$	–		–
$\Delta\phi(E_T^{\text{miss}}, \vec{p}_T^{\text{miss}})$	$< 90^\circ$	–		–
p_T^V regions	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	$75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$
Signal regions	$\Delta R(b_1, b_2)$ signal selection			
Control regions	High and low $\Delta R(b_1, b_2)$ side-bands			

6.5 Multivariate Algorithms

The $VH, H \rightarrow b\bar{b}$ analysis is a Multivariate Analysis (MVA). The algorithm preferred by the $VH, H \rightarrow b\bar{b}$ analysis consists of a Boosted Decision Tree (BDT) designed to discriminate the signal from the sum of all backgrounds. As already explained in the introduction of this chapter, the nominal analysis, whose BDT will be addressed to as BDT_{VH} , is validated by a cross-check analysis in which the VH signal is replaced by the VZ signal and the correspondent BDT will be addressed to as BDT_{VZ} .

Multivariate approaches are particularly useful when no single variable, describing the analysis phase space, shows a clear separation power between signal and background. A set of variables $x_{i...n}$ able to discriminate between signal and background are passed as input to the multivariate algorithm which combines them into a single output variable y having an enhanced separation power. The relation between the $x_{i...n}$ variables and the final output y is given by:

$$y = f(x_i, \dots, x_n). \quad (6.1)$$

The procedure used to determine the function f is called *training*, or *supervised learning* if the events are previously labelled as signal and background events, as in the case of the $VH, H \rightarrow b\bar{b}$ analysis. The supervised learning needs to be performed on a large number of events otherwise the multivariate technique can be sensitive to *overtraining* effects, meaning that the function f learns the features of the particular *training* sample making its response not optimal when applied to an alternative *testing* sample. One possibility to overcome this issue and maximize the number of training events would be to divide the initial sample into fully orthogonal training and testing sets. Each training set will yield an output distribution y which can be cross-checked applying it to the correspondent testing sample. For the $VH, H \rightarrow b\bar{b}$ analysis a BDT algorithm with two sub-sets (folds) has been chosen, given the large available statistics of the signal and background MC samples.

The BDT is a combination of Decision Trees (DT). A decision tree takes a set of variables as input and each variable is used to build a decision node; each decision node can be recursively split into two branches (or children) until some stopping condition is reached ending the branch into a leaf. A representation of a DT is shown in Figure 6.8.

The core of a decision tree algorithm resides in how a node is split into two [188]. The function used to optimized the cuts in the tree is the *Gini index* which can be thought as the probability of misclassifying an event. The *Gini index* is maximal when the signal and background content in the node is identical (no separation), minimal for nodes with either only signal or only background events (perfect separation) and symmetric in signal and background purities. The *Gini index* is defined as:

$$Gini = \left(\sum_i^{N_{events}} w_i \right) p(1 - p) \quad (6.2)$$

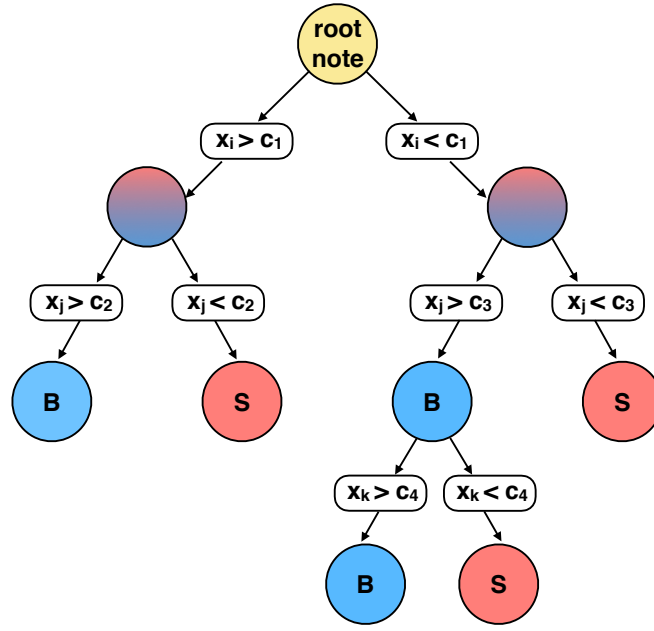


Figure 6.8: Schematic view of a decision tree. Starting from the first (root) node, a sequence of binary splits is applied to the data using the discriminating variables x_i . Each split uses the variable that at this node gives the best separation between signal and background. The leaf nodes at the bottom end of the tree are labelled “S” for signal and “B” for background depending on the majority of events that end up in the respective nodes.

where $\left(\sum_i^{N_{events}} w_i\right)$ is the weighted sum of events in the node and p is the signal purity of the node, $p = \frac{n_s}{n_s + n_b}$, with n_s and n_b number of signal and background events. The event weight w_i of each signal and background event in the $VH, H \rightarrow b\bar{b}$ analysis is defined from the MC simulation and contains all the corrections described in Section 6.2 and the normalisation of the sample to the correct cross-section.

The *Gini index* is calculated for both the parent and the children and the optimal cut is chosen to maximize the difference between the *Gini index* of the mother node and the one of the two children nodes:

$$\tilde{G} = Gini_{parent} - Gini_{child_1} - Gini_{child_2} \quad (6.3)$$

The growth of the node is stopped when specific conditions are met:

- **minimum node size:** if the number of events falling into a child node is below a certain threshold;
- **max depth:** if the distance from the initial node is equal to a certain threshold;
- **gain:** if $\tilde{G}$ from adding an additional layer is less than 0.

At this point, the node becomes a leaf and the event is classified as *signal* or *background* acquiring a score that is either the *signal purity* defined above or, as in the case of the

$VH, H \rightarrow b\bar{b}$ analysis, +1 if the leaf contains a majority of signal events, or -1 if it contains a majority of background events.

6.5.1 Boosted Decision Tree

A single tree is highly susceptible to statistical fluctuations in the training set. One way to overcome this issue is to build a “forest” of trees using the *boosting* technique. The idea is to subsequentially build many “weak” classifiers each of which learns from the mistakes of the previous one. All classifiers are then combined into an averaged classifier which smooths out the statistical fluctuations each single tree suffers from.

Historically, the $VH, H \rightarrow b\bar{b}$ analysis has always adopted AdaBoost as boosting decision algorithm, however the full Run 2 version of the analysis uses GradientBoost as it was shown to bring an improvement of 2%-5% according to the region. Some details of the GradientBoost are given in the following.

Given a set of N training events $\{y_i; \mathbf{x}_i\}_1^N$, with $\mathbf{x}_i = \{x_1 \dots x_n\}_i$ the values of the input variables for the event i , and y_i the correspondent response of the classifier, the goal of the boosting technique is to build an improved predictive model which maps $\mathbf{x}_i$ to y_i . The idea is to first train a classifier with the full set of training events and then, for the next iteration, use a modified training sample in which the previously mis-classified events are given a larger weight [189, 190]. The final classifier is therefore a linear combination of M weak classifiers of the form:

$$F(\mathbf{x}) = \sum_{m=0}^M \beta_m f(\mathbf{x}; \mathbf{a}_m), \quad (6.4)$$

where $f(\mathbf{x}; \mathbf{a}_m)$ is the corrected response of a single classifier, β_m is the weight of the classifier in the final average, and $\mathbf{a}_m$ are the parameters specifying the various node splits of the classifier m . Both the expansion coefficients $\{\beta_m\}_0^M$ and the parameters $\{\mathbf{a}_m\}_0^M$ are derived at each iteration in the training (therefore for each weak classifier) by fitting them to the training data such that the *loss function* between the predictive model and the response y of the classifier to the current training sample is minimised event per event:

$$(\beta_m, \mathbf{a}_m) = \arg \min_{\beta, \mathbf{a}} \sum_{i=1}^N L(y_i, \sum_{m=0}^M \beta_m f(\mathbf{x}_i; \mathbf{a}_m)). \quad (6.5)$$

The current implementation of the GradientBoost in the TMVA package [133] used in this analysis uses a *binomial loss function*:

$$L(F, y) = \ln(a + e^{-2F(\mathbf{x})y}). \quad (6.6)$$

Giving good results already with small trees of 5-10 nodes, GradientBoost is typically less susceptible to overtraining that is why it was chosen as the default algorithm in the full Run 2 iteration of the $VH, H \rightarrow b\bar{b}$ analysis.

6.5.2 The multivariate technique in the $VH, H \rightarrow b\bar{b}$ analysis

The training is performed on the full phase space of the analysis before the event classifications into signal and control regions, in order to enhance the training statistics and avoid overtraining. Due to the different background composition and kinematic variables able to discriminate between signal and backgrounds, the training is divided into eight training regions, according to the lepton channel, the jet category and the p_T^V region, as shown in Table 6.6. All categories are defined in the 2-tag region.

Table 6.6: The eight regions in which the training is divided.

Channel	Jet category	p_T^V region
0-lepton	2-jet	$E_T^{\text{miss}} > 150 \text{ GeV}$
	3-jet	
1-lepton	2-jet	$p_T^V > 150 \text{ GeV}$
	3-jet	
2-lepton	2-jet	$75 < p_T^V < 150 \text{ GeV}$
		$p_T^V > 150 \text{ GeV}$
	≥ 3 -jet	$75 < p_T^V < 150 \text{ GeV}$
		$p_T^V > 150 \text{ GeV}$

Contrary to the definition of the analysis regions, the events are not subject to a further cut at $p_T^V > 250 \text{ GeV}$ as studies have proven that a unique training in the $p_T^V > 150 \text{ GeV}$ has equivalent performances as two separate trainings (in $150 < p_T^V < 250 \text{ GeV}$ and $p_T^V > 250 \text{ GeV}$).

The input variables are chosen following an iterative procedure defined in Run 1 [191] and most of them have been kept unchanged for the $H \rightarrow b\bar{b}$ observation publication and the full Run 2 analysis. Starting from m_{bb} and ΔR_{bb} , a new variable is added only if it improves the sensitivity gain of the BDT training, which is calculated on a bin-by-bin basis for a given distribution as:

$$S = \sqrt{\sum_{i=1}^N \left\{ 2 \times \left[(s_i + b_i) \times \log \left(1 + \frac{s_i}{b_i} \right) - s_i \right] \right\}}, \quad (6.7)$$

with N the total number of bins in the distribution, and s_i (b_i) the signal (background) yield in bin i . The procedure is repeated for all variables until no further improvement is observed. This is done separately in each lepton channel and the input variable entering the final MVA are summarized in Table 6.7 with the associated description.

Table 6.7: Variables entering the multivariate discriminant in each of the lepton channels and jet category. In the 0-lepton channel, p_T^V is not an additional input variable as it corresponds exactly to E_T^{miss} .

Variable	0-lepton	1-lepton	2-lepton
m_{bb}	×	×	×
$\Delta R(\vec{b}_1, \vec{b}_2)$	×	×	×
$p_T^{b_1}$	×	×	×
$p_T^{b_2}$	×	×	×
p_T^V	×	×	×
$\Delta\phi(\vec{V}, \vec{bb})$	E_T^{miss}	×	×
MV2(b_1)	×	×	×
MV2(b_2)	×	×	×
$ \Delta\eta(\vec{b}_1, \vec{b}_2) $	×	×	×
m_{eff}	×	×	×
$p_T^{\text{miss},st}$	×	×	×
E_T^{miss}	×	×	×
$\min[\Delta\phi(\vec{\ell}, \vec{b})]$	×	×	×
m_T^W	×	×	×
$ \Delta y(\vec{V}, \vec{bb}) $	×	×	×
m_{top}	×	×	×
$ \Delta\eta(\vec{V}, \vec{bb}) $	×	×	×
$E_T^{\text{miss}} / \sqrt{S_T}$	×	×	×
m_{ll}	×	×	×
$\cos\theta(\vec{\ell}^2, \vec{Z})$	×	×	×
angle between negatively charged lepton and Z-boson flight direction in Z-boson rest frame			
Only in 3-jet events			
$p_T^{\text{jet}3}$	×	×	×
m_{bbj}	×	×	×
transverse momentum of the third jet of the 3-jet system			
invariant mass of the 3-jet system			

The variables m_{top} and $|\Delta\eta(\vec{V}, b\bar{b})|$, entering the 1-lepton channel BDT training, need the neutrino longitudinal momentum to be determined by imposing a constraint on the W boson mass, m_W , to solve the equation:

$$p_z^\nu = \frac{1}{2(p_T^l)^2} \left[p_z^l X \pm E_l \sqrt{X^2 - 4(p_T^l)^2 (E_T^{\text{miss}})^2} \right], \quad (6.8)$$

with $X = m_W^2 + 2p_x^l E_x^{\text{miss}} + 2p_y^l E_y^{\text{miss}}$, where p_T^l is the lepton's transverse momentum, $p_{x,y,z}^l$ are the x , y and z components of the lepton's 4-momentum and $E_{x,y}^{\text{miss}}$ are the x and y components of the missing transverse momentum. If p_z^ν has an imaginary solution, the E_T^{miss} is shifted such that the discriminant is equal to zero. The mass of the top-quark is then reconstructed by selecting the combination of b -tagged jets and solution to the Equation 6.8 which minimises m_{top} .

For the full Run 2 analysis few new variables have been added to the BDT with respect to the $H \rightarrow b\bar{b}$ observation publication as studies have shown significant improvements in the training sensitivity:

- the binned MV2c10 b -tagging discriminants for the leading and sub-leading b -tagged jets constructed by grouping the scores into two bins corresponding to efficiencies of 0%–60% and 60%–70%, identical with the bins defined for the pseudo-continuous b -tagging working point (see Figure 4.9). These variables were found to improve the sensitivity of the 0-lepton (1-lepton) channel by around 7% (10%) thanks to the high discrimination power against the bc component of the $t\bar{t}$ and single top Wt -channel backgrounds;
- the track-based soft E_T^{miss} term, which is the vectorial sum of the p_T of all tracks in the events that are not associated to any reconstructed object. This variable was found to improve the sensitivity of the 0-lepton channel by 2% - 3% (region dependent);
- the polarization angle $\cos\theta(l^-, Z)$, which is the angle between the negatively charged lepton in the Z -boson rest frame and the Z -boson flight direction in the rest frame. It makes use of the difference in polarization between the signal and the Z +jets background suggested in [192] and was shown to improve the sensitivity of the 2-lepton channel by $\sim 7\%$.

The BDT parameters (also referred to as *hyper-parameters*), have been optimised in order to exploit the complete set of training events. Coarse scans of values for the learning rate, the number of trees and the depth of the trees have been performed. The final hyper-parameter configuration, reported in Table 6.8, was chosen in order to maximise the sensitivity without introducing significant level of overtraining.

The number of MC events available for the training can have a direct impact on the BDT performance. In order to control this effect, the training dataset is split into two sub-sets, named *folds*, according to the event number, even or odd. Each of the eight BDTs is

Table 6.8: Hyperparameters used for the BDT trainings in the context of the full Run 2 analysis. Exceptions for the 1-lepton VH and diboson trainings are given in brackets. The “Shrinkage” or learning rate controls the weight of the individual trees; the smaller the learning rate, the highest the number of necessary trees to be grown, which can significantly improve the accuracy of the prediction. A pruning procedure could be applied to remove statistically insignificant nodes and reduce the overtraining of the tree. However, no improvement was found in using this option.

TMVA Setting	Value	Definition
BoostType	Gradient Boosting	Boost procedure
Shrinkage	0.5	Learning rate
SeparationType	Gini index	Node separation gain
PruneMethod	No Pruning	Pruning method
NTrees	200 (600 for 1-lepton VH)	Number of trees
MaxDepth	4 (2 for 1-lepton diboson)	Maximum tree depth
nCuts	100	Number of equally spaced cuts tested per variable per node
nEventsMin	5%	Minimum number of events in a node (% of total events)

trained on the even (odd) fold and tested on the odd (even) fold. The final discriminant is then built by summing the MV discriminants of the even and odd folds since no difference in the physics is expected between them.

BDT Transformation A re-binning of the BDT_{VH} and BDT_{VZ} outputs is performed in order to achieve the best signal sensitivity whilst maintaining the MC statistical uncertainty at a reasonable level. The following function is considered:

$$Z(I[k, l]) = Z(z_s, n_s(I[k, l]), N_s, z_b, n_b(I[k, l]), N_b), \quad (6.9)$$

where $I[k, l]$ is an interval of the histogram containing the bins between bin k and bin l , N_s (N_b) are the total number of signal (background) events, $n_s(I[k, l])$ ($n_b(I[k, l])$) is the total number of signal (background) events in the interval $I[k, l]$, and z_s (z_b) are parameters used to tune the algorithm. The Z function used in the $VH, H \rightarrow b\bar{b}$ analysis is defined as:

$$Z = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b}. \quad (6.10)$$

The following procedure is adopted to rebin the BDT_{VH} and BDT_{VZ} outputs:

1. Starting at the maximum BDT score boundary, the bins are combined one after the other increasing the range $I[k, last]$ and the value of Z of equation 6.10 is calculated at each step;
2. Once $Z(I[k_0, last]) > 1$ and the MC statistical uncertainty in the range is below 20%, all bins in the interval $I[k_0, last]$ are merged into a single bin;
3. Steps 1-2 are repeated, starting from the last bin on the right not included in the previous rebinning until k_0 is the first bin.

Sensitivity studies have shown that the values $z_b = 5$ and $z_s = 10$, giving 15 bins in the $p_T^V < 250$ GeV region, and $z_b = 3$ and $z_s = 5$ giving 8 bins in the $p_T^V > 250$ GeV region provide the best signal over background ratio at high values of the BDT_{VH} score. Due to the limited number of events in the diboson samples, a value of 5 has been chosen for both z_b and z_s in the $p_T^V < 250$ GeV region, yielding 10 bins, and values of 3 and 2, respectively, in the $p_T^V > 250$ GeV which give 5 bins in total.

6.6 Statistical Interpretation

The $VH, H \rightarrow b\bar{b}$ fitting procedure is based on the profile-likelihood ratio formalism which is outlined in this section.

6.6.1 Significance testing

The goal of a significance test is to make a statement about how well the observed data is in agreement with a given hypothesis [193]. The hypothesis under consideration is called the **null hypothesis** H_0 and can be described by the probability density function (p.d.f.) of the random variable x given the hypothesis H_0 , $f(x | H_0)$. This is compared to an **alternative hypothesis** H_1 described by the p.d.f. of x given the hypothesis H_1 , $f(x | H_1)$. Assuming a dataset of n observed values of $x = (x_1, \dots, x_n)$, the agreement of the observed dataset with a certain hypothesis is tested using a **test statistic** $t(x)$. Each of the hypotheses will have a p.d.f. for the test statistic, $g(t | H_0)$ if H_0 is true and $g(t | H_1)$ if H_1 is true. A cut on the value of t will define a **critical region** such that if the measured value of t falls within it, the null hypothesis is rejected otherwise it is not. By applying this cut, one defines an **error of the first kind**, when the H_0 hypothesis is rejected but it is true:

$$\alpha = \int_{t_{cut}}^{\infty} g(t | H_0) dt, \quad (6.11)$$

and an **error of the second kind**, corresponding to the probability of rejecting H_1 although it is true:

$$\beta = \int_{-\infty}^{t_{cut}} g(t | H_1) dt. \quad (6.12)$$

An illustration of the concept can be found in Figure 6.9.

The quantity $1 - \beta$ is called the *power* of the test relative to H_1 . The *Neyman-Pearson lemma* states that the critical region giving the best power for a given error α is given by:

$$r = \frac{L(t | H_1)}{L(t | H_0)}. \quad (6.13)$$

The ratio r is called **likelihood ratio** for the H_0 and H_1 hypotheses. If one thinks of H_0

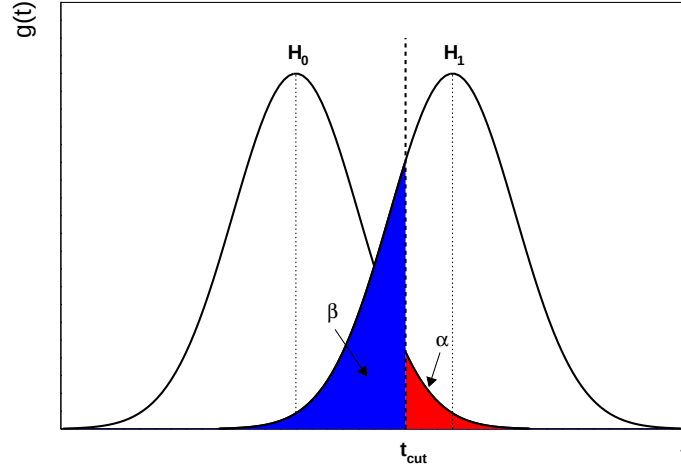


Figure 6.9: Illustration of the hypothesis test. The cut t_{cut} defines the two errors of first (α) and second (β) kind; if $t > t_{\text{cut}}$, H_0 is rejected with the probability α but it is true, while $t < t_{\text{cut}}$ corresponds to reject H_1 when it is true.

as the background-only hypothesis and to H_1 as the signal + background hypothesis, the likelihood ratio corresponds to the test statistics which gives the highest probability to reject the background-only hypothesis when there is an actual signal present.

The level of compatibility of the observed data with the background-only hypothesis can be quantified by computing the ***p*-value** which is the probability that, by repeating the experiment under the assumption H_0 , one obtains a disagreement equal or larger than the one found. The *p*-value for H_0 is defined as:

$$p_0 = \int_{t_{\text{obs}}}^{\infty} g(t | H_0) dt. \quad (6.14)$$

One can regard the hypothesis as excluded if its *p*-value is observed below a specified threshold. In particle physics the *p*-value is often translated into a quantity called “*significance*”, defined such that a Gaussian distributed variable with Z standard deviations above its mean has an upper-tail probability equal to p [193]. That is:

$$Z = \Phi^{-1}(1 - p), \quad (6.15)$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian. In case of a search analysis aiming for a discovery, the background hypothesis has to be rejected with a significance of at least $Z = 5\sigma$ (that corresponds to a *p*-value threshold of $2.87 \cdot 10^{-7}$). For purposes of excluding a signal hypothesis, a threshold *p*-value of 0.05 (i.e., 95% confidence level) is often used, which corresponds to $Z = 1.64$. A representation of the *p*-value and the relation between the *p*-value and the significance Z is given in Figure 6.10.

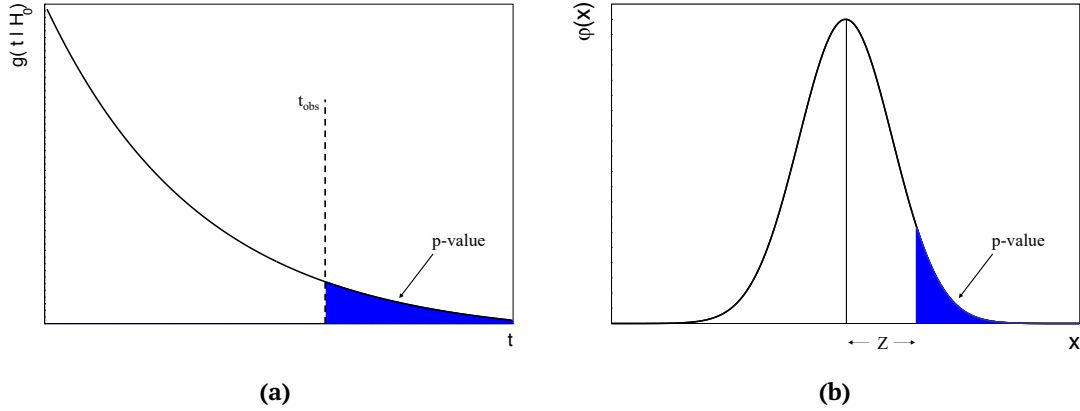


Figure 6.10: (a) Illustration of the relation between p -value obtained from an observed value of the test statistic t . (b) The standard Gaussian $\phi(x)$ showing the relation between the significance Z and the p -value.

6.6.2 Global profile likelihood procedure in the $VH, H \rightarrow b\bar{b}$ analysis

The statistical analysis of the data collected in the $VH, H \rightarrow b\bar{b}$ analysis uses a binned likelihood function to extract the normalisation factor of the signal process. A likelihood function is the probability to observe the measured data given a certain hypothesis. In the $VH, H \rightarrow b\bar{b}$ analysis, the hypothesis is represented by the presence or not of the $VH, H \rightarrow b\bar{b}$ signal which is expressed by the *parameter of interest* (POI) of the *signal strength*:

$$\mu = \frac{\sigma_{VH, H \rightarrow b\bar{b}}^{\text{exp}}}{\sigma_{VH, H \rightarrow b\bar{b}}^{\text{SM}}}, \quad (6.16)$$

where $\sigma_{VH, H \rightarrow b\bar{b}}^{\text{exp}}$ is the measured experimental cross section of the $VH, H \rightarrow b\bar{b}$ process and $\sigma_{VH, H \rightarrow b\bar{b}}^{\text{SM}}$ is its theoretical prediction in the SM. A value of $\mu = 0$ corresponds to the background-only hypothesis while $\mu = 1$ coincides with the production of a SM Higgs boson in the $H \rightarrow b\bar{b}$ decay channel. The best estimate of the signal strength μ is determined by maximising the likelihood or, equivalently, minimising its negative logarithm. The $\pm 1\sigma$ uncertainty on μ is defined by determining the points in which the negative logarithm of the likelihood increases by 0.5 with respect to the minimum value. The data in each bin of each distribution entering the fit follow a Poisson probability distribution around their expected number of events and the likelihood function for all the bins and regions can be written as:

$$L(\mu) = \prod_j^{\text{reg}} \prod_i^{\text{bins}_j} \frac{(\mu s_{ij} + b_{ij})^{n_{ij}}}{n_{ij}!} e^{-(\mu s_{ij} + b_{ij})}, \quad (6.17)$$

where n_{ij} is the number of data events and s_{ij} (b_{ij}) is the number of expected signal (background) events in the i^{th} bin of the fitted distribution in region j .

Systematic uncertainties, which encode the imperfect knowledge on the signal and

background predictions (from Section 6.7 to Section 6.10) and of the description of the physical objects (see Section 6.11), are incorporated in the likelihood through a set of parameters θ called *nuisance parameters* (NPs). The NPs parametrise the effect of each source of uncertainty on the template distributions of the signal and backgrounds. Varying the values of the NPs, both the shape and the normalisation of the templates are modified giving the fit model more freedom to find the best agreement between data and MC prediction and determining the uncertainty on the NPs.

Each systematic uncertainty is added to the likelihood through a function $f(\theta_k|\tilde{\theta}_k)$, which encodes the central value $\tilde{\theta}_k$ of the associated NP and its uncertainty σ_k . In this way, the binned likelihood defined in 6.17 gets modified as follows:

$$L(\mu, \theta) = \prod_j^{\text{reg}} \prod_i^{\text{bins}_j} \frac{(\mu s_{ij}(\theta) + b_{ij}(\theta))^{n_{ij}}}{n_{ij}!} e^{-(\mu s_{ij}(\theta) + b_{ij}(\theta))} \times \prod_k^{\text{n. systs}} f(\theta_k|\tilde{\theta}_k). \quad (6.18)$$

Typically, a Gaussian term is chosen as function $f(\theta_k|\tilde{\theta}_k)$, such that:

$$Gauss(\theta|\tilde{\theta}) = \prod_{k=1}^{\text{n. systs}} \frac{1}{\sqrt{2\pi}\sigma_k} e^{\frac{-(\theta_k - \tilde{\theta}_k)^2}{2\sigma_k^2}}, \quad (6.19)$$

where $\tilde{\theta}_k$ is the central value of the Gaussian and, by convention, is set to $\tilde{\theta}_k = 1$ for the normalisation NPs and to $\tilde{\theta}_k = 0$ for all the others. The prior uncertainty σ_k on the NP (can be a theoretical uncertainty or the uncertainty associated to the jet energy scale) is typically derived from auxiliary measurements and acts as a penalty against large deviation from the a priori central value $\tilde{\theta}_k$. Although the Gaussian distribution is most commonly assumed as prior, other functional forms such as the log-normal or the gamma function could be more suitable in other cases.

The nuisance parameters are fully determined by the fit. In case the data is not particularly sensitive to a given systematic, the Gaussian constraining term ensures that its central value stays at 0 and its uncertainty is dominated by its prior. On the other hand, the fit could shift (“pull”) the central value of the nuisance parameter to help the fit find a better data-MC agreement. A pull is usually accompanied by a reduction (“constraint”) of the uncertainty of the NP. The opposite case, when a NP appears more weakly constrained than the initial prior, is not a physical result and needs to be further investigated as it can indicate problems in the definition of the fit parameters or a consequence of numerical limitations of the fit.

Finally, the statistical uncertainty of the Monte Carlo templates of both signal and backgrounds is also incorporated in the fit. Following a simplified version of the approach used by Barlow and Beeston [194], a single nuisance parameter is introduced for each bin of the fitted distributions allowing the total MC prediction in that bin to vary within its statistical uncertainty.

From Equation 6.13, the most powerful test statistic to exclude the background-only

hypothesis can be written as:

$$q_\mu = -2 \ln \lambda(\mu) = -2 \ln \left(\frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} \right), \quad (6.20)$$

and it is referred to as “*profile likelihood ratio*”. The values $\hat{\mu}$ and $\hat{\theta}$ are those that maximise the likelihood (unconditional maximum likelihood) and $\hat{\theta}$ are the nuisance parameters values that maximise the likelihood for a given μ (conditional maximum-likelihood). A value of q_μ near 0 implies good agreement between data and the hypothesized value of μ while for higher values of q_μ their incompatibility increases. To quantify the level of the disagreement one can compute the p -value:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu | \mu) dq_\mu, \quad (6.21)$$

with $q_{\mu,obs}$, the value of the test statistic q_μ observed in data and $f(q_\mu | \mu)$, the p.d.f. of q_μ under the assumption of the signal strength μ . Wilks theorem [195] assures that the p.d.f. $f(q_\mu | \mu)$ asymptotically tends towards a χ^2 distribution with a number of degrees of freedom equal to the difference between the dimensionality of the numerator and the denominator in Equation 6.20. Thus, in the case with $\mu = 0$, $q_{\mu=0}$ has a χ^2 distribution with *one* degree of freedom:

$$f(q_0 | 0) = \frac{1}{2} \delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q}} e^{-\frac{q}{2}}. \quad (6.22)$$

This means that, following the demonstration in Ref. [193] and combining the result with Equation 6.15, the significance can be written as:

$$Z_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0}. \quad (6.23)$$

A p -value of 1.35×10^{-2} corresponds to a 3σ deviation from the background-only model and it is used as a benchmark to claim “*evidence*” of the signal. A p -value of 2.87×10^{-7} corresponds to a 5σ deviation from the background-only model, and is generally used as the benchmark deviation required for “*discovery*” of a signal in high energy physics. In case no hint of a signal is found in this procedure, the CL_s method [193, 196] is used to derive exclusion intervals.

6.6.3 Experimental sensitivity and the Asimov dataset

Before doing the actual measurement one wants to determine at which confidence level the background-only hypothesis is expected to be excluded. This is an important step in order to design the analysis and quantify the effect of possible improvements. To evaluate the **expected** exclusion of the background-only hypothesis one would need to use a very large number of MC events and look at the median significance obtained from these measurements. However, this is very expensive in terms of computational time

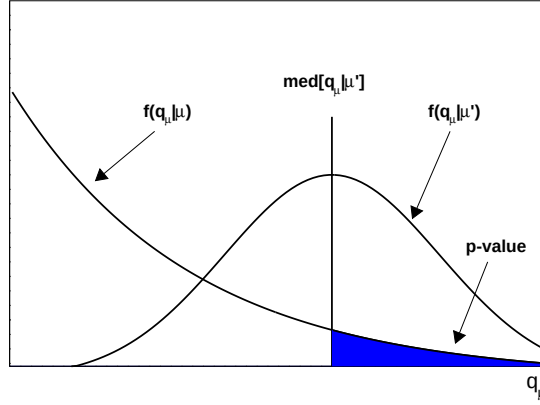


Figure 6.11: Illustration of the p -value corresponding to the median of the test statistic q_μ assuming an alternative hypothesis with μ' .

and complexity. Thus, an alternative method is to use approximate formulas which hold in the limit of very high statistics samples, using an extension of the Wilk's theorem formulated by Wald [197]. Generalizing Equation 6.21, the expected p -value is defined as:

$$p_\mu = \int_{q_{\mu,exp}}^{\infty} f(q_\mu | \mu) dq_\mu, \quad (6.24)$$

where $f(q_\mu | \mu)$ is the p.d.f. of the test statistic q_μ and $q_{\mu,exp}$ is the *median* of the p.d.f. for an alternative signal strength μ' where $\mu' \neq 0$:

$$q_{\mu,exp} = med[f(q_\mu | \mu')]. \quad (6.25)$$

This is summarised in Figure 6.11.

The p.d.f. of $f(q_\mu | \mu')$ can be found using the result by Wald which states that, in case of a single parameter of interest:

$$q_\mu = -2 \ln \lambda(\mu) = -2 \left(\ln(L(\mu, \hat{\theta})) - \ln(L(\hat{\mu}, \hat{\theta})) \right) = \left(\frac{\mu - \hat{\mu}}{\sigma} \right)^2 + O\left(\frac{1}{\sqrt{N}}\right), \quad (6.26)$$

where N represents the data sample size and $\hat{\mu}$ follows a Gaussian distribution with mean μ' and standard deviation σ . Neglecting the $O(\frac{1}{\sqrt{N}})$, one can show that q_μ follows a *non-central* χ^2 distribution for one degree of freedom.

With this assumption, one can replace the MC toy experiments with a single representative **Asimov** dataset such that, when it is used to evaluate the estimators for all the parameters of the likelihood, one obtains the expected parameter values. The estimators of the parameters that maximise the generic likelihood in Equation 6.18 can be found by setting the derivative of q_μ in Equation 6.26 with respect to all the parameters equal to zero. But this condition is satisfied only if all the parameters of the Asimov dataset (μ, θ)

are identical to their expectation values. This means that the Asimov not only follows the probability density function for a specific μ' but, when computing its parameters, gives the median of the test statistic by construction.

The remaining point is the calculation of the uncertainty on μ which can not be calculated analytically, as it depends on the nuisance parameters θ . Considering Equation 6.26 for the Asimov dataset, which for a signal strength μ' gives $\hat{\mu} = \mu'$, one finds:

$$q_{\mu,A} = -2 \ln \lambda_A(\mu) \approx \frac{(\mu - \mu')^2}{\sigma_A^2}, \quad (6.27)$$

and inverting this equation, the variance σ_A^2 is:

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}. \quad (6.28)$$

For the interesting case of the discovery where the μ under test is $\mu = 0$, the previous Equation reduces to:

$$\sigma_A^2 = \frac{\mu'^2}{q_{\mu,A}}. \quad (6.29)$$

The median significance is then obtained evaluating Z as a function of the median test statistic. Combining Equation 6.15 and Equation 6.25, one can show that the significance Z is a monotonic function of q_μ so that the median Z can be evaluated directly from the median of q_μ . Following Equation 6.23 the median Z for a $\mu' \neq 0$ is:

$$\text{med}[Z_0 | \mu'] = \sqrt{q_{0,A}}. \quad (6.30)$$

6.6.4 Correlations

When constructing the likelihood it is possible that two NPs, θ_i and θ_j , have a similar effect on a particular region or MC template and, therefore, become correlated. In the case of a high statistics sample, the expected values of the likelihood's parameter estimators are equal to their true value, meaning that the estimators are *unbiased*. According to the Cramer-Rao inequality [198], the variance of an unbiased estimator $\hat{\theta}$ obeys the following inequality:

$$\text{Var}[\hat{\theta}] \geq \frac{1}{E \left[-\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right]} \quad (6.31)$$

where the denominator is called *Fisher information* matrix or *Hessian* matrix. For unbiased estimators, it is possible to replace the mean of the second derivative with the second derivative evaluated at the estimator central value, so that the Hessian can be rewritten as:

$$\mathcal{H} = E \left[-\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right] = -\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\hat{\theta}} \quad (6.32)$$

and extensively:

$$\mathcal{H} = - \begin{pmatrix} \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \ln L}{\partial \theta_1 \partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 \ln L}{\partial \theta_n \partial \theta_2} & \dots & \frac{\partial^2 \ln L}{\partial \theta_n \partial \theta_n} \end{pmatrix}. \quad (6.33)$$

6.6.5 The fit model of the $VH, H \rightarrow b\bar{b}$ analysis

The fit model is defined by the variable used as discriminant in each fitted region, the list of systematic uncertainties that are added as nuisance parameters, and the way they are correlated across the regions for the signal and background templates.

The $VH, H \rightarrow b\bar{b}$ analysis, as well as the $VZ, Z \rightarrow b\bar{b}$ and the di-jet mass analyses, is split in three lepton channels, depending on the number of leptons in the final state. For each lepton channel, three p_T^V regions are defined: medium for $75 < p_T^V < 150$ GeV only in the 2-lepton channel, high for $150 < p_T^V < 250$ GeV and extreme for $p_T^V > 250$ GeV. Each p_T^V region has one signal region, one Low- ΔR CR, and one High- ΔR CR for each jet category, 2-jet and 3-jet. In the signal regions, the BDT_{VH} (BDT_{VZ}) and the m_{bb} distribution are used as fit discriminants for the $VH, H \rightarrow b\bar{b}$ ($VZ, Z \rightarrow b\bar{b}$) and the di-jet mass analyses, respectively. The Low- ΔR CR and High- ΔR CR have a single bin, therefore, only the yield of the MC templates enter the fit. However, each p_T^V bin and jet category has two one-bin CR which is equivalent to fit p_T^V with the CRs as bins. Studies have been performed by the author of this thesis to demonstrate the advantages of choosing p_T^V (E_T^{miss} in the 0-lepton channel) as fit discriminant and details will be given in the next paragraph of this section.

Template distributions for the $VH, H \rightarrow b\bar{b}$ ($VZ, Z \rightarrow b\bar{b}$) signal and all backgrounds are extracted from MC simulation in all the regions in which the analysis is categorized, with the exception of the multi-jet background in the 1-lepton channel, and the $t\bar{t}$ and single-top backgrounds in the 2-lepton channel, whose templates are extracted through a data-driven approach described in Section 6.10.

A summary of the regions and the respective fit quantity, the BDT distribution or the yields, is reported in Table 6.9.

The full list of systematics uncertainties entering the fit model and the procedure used to derive them are extensively treated in the next section. In general, two types of parameters are considered: *unconstrained* normalisation factors, which change the normalisation of the nominal templates, and *constrained* nuisance parameters (acceptance and shape uncertainties, per bin systematic to account for the MC statistical uncertainty), that have a prior on the NP distribution added as Gaussian (or log-normal) term inside the likelihood (see Equation 6.18).

p_T^V variable as fit discriminant in the control regions In the $H \rightarrow b\bar{b}$ observation analysis, the shape uncertainty on the W +jets background was found to have the largest impact on the total uncertainty on the signal strength. This was evaluated using the so-called **ranking method**, which consists of re-evaluating μ after fixing the NP relative

Table 6.9: Summary of the regions included in the fit with the respective fitted quantity.

Channel	p_T^V region	Signal or control regions					
		Low- ΔR CR		Signal region		High- ΔR CR	
		2-jets	3-jets	2-jets	3-jets	2-jets	3-jets
0-lepton	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	yields		BDT		yields	
1-lepton	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	yields		BDT		yields	
2-lepton	$75 \text{ GeV} < p_T^V < 150 \text{ GeV}$ $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$ $p_T^V > 250 \text{ GeV}$	yields		BDT		yields	

to the systematic under test to its best fit value $\hat{\theta}$ and varying it by $\pm 1\sigma$, with σ its fitted uncertainty. All other parameters are left free floating to account for correlations. The difference between the nominal μ and the new μ value, $\Delta\mu$, is used as an estimate of the impact of the systematic uncertainty. The higher the difference, the larger the impact of the systematic.

Figure 6.12 shows the nuisance parameters ranking of the observation analysis where it can be observed that the shape uncertainty on the W +jets p_T^V distribution is the first ranked systematic.

To reduce the impact of this systematic on the total uncertainty on the signal strength, studies were performed in the 1-lepton channel, having the largest W +jets contamination, adopting the same integrated luminosity, event selection and categorization used for the observation paper:

- proton-proton collision dataset of 79.8 fb^{-1} ;
- no split at $p_T^V = 250 \text{ GeV}$ but a single $p_T^V > 150 \text{ GeV}$ region;
- a single W +hf CR, with a purity of 80% (75%) in the 2-jet (3-jet) categories, built applying a cut on $m_{bb} < 75 \text{ GeV}$, to remove the contamination from signal events, and a cut on $m_{top} > 225 \text{ GeV}$ to remove a large fraction of the $t\bar{t}$ background.

Two p_T^V shape systematics, one for each jet category (2-jet and 3-jet), were derived in the 1-lepton channel using a first order polynomial fit to the ratio between the nominal and the alternative generator. These were treated as uncorrelated across all analysis regions due to the different order of calculation of the nominal and the alternative generator; NLO up to two extra partons for SHERPA 2.2.1, and LO for 0-4 extra partons

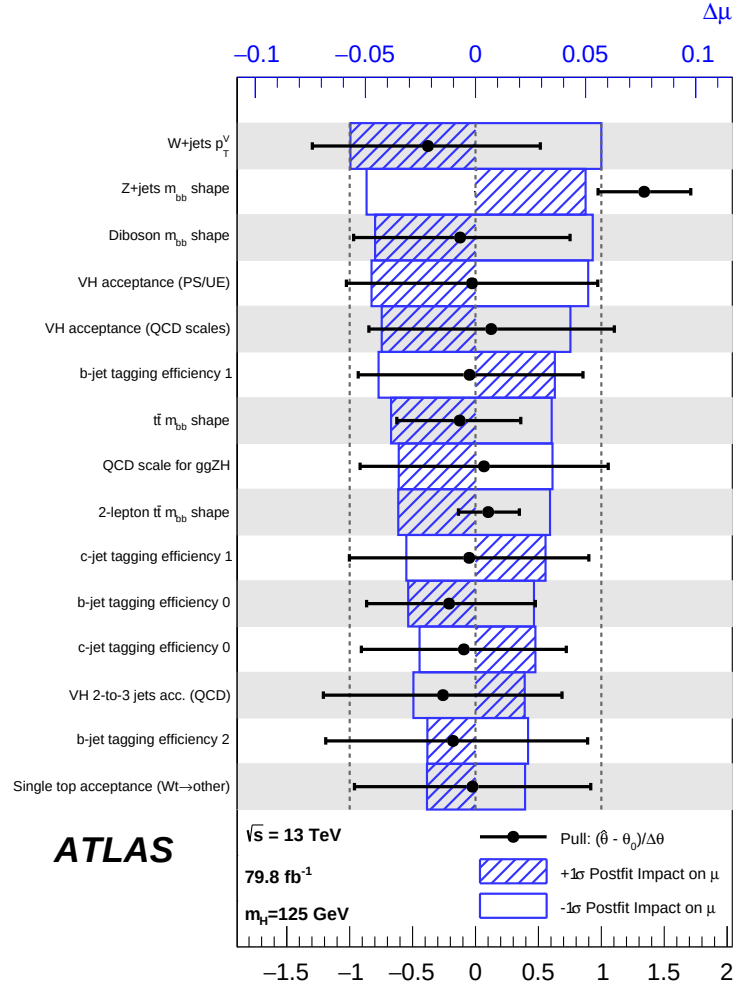


Figure 6.12: Impact of systematic uncertainties on the best fit value of the signal strength for the observation analysis. The systematic uncertainties are listed in decreasing order of their impact on μ . The boxes show the impact of the nuisance parameter on the total uncertainty, $\Delta\mu$, and the hatched and open areas correspond to the up and down variations, respectively. The black dots correspond to the deviation of the nuisance parameters from their nominal values (pulls) and the associated error bars correspond to their post-fit uncertainties. It can be seen that the nuisance parameter associated to the W +jets p_T^V shape has the largest impact on the total uncertainty [13].

for MADGRAPH 5. This implies that the uncertainty derived in the 3-jet category only compares the LO calculations of the two generators. Moreover, the 3-jet category has much more statistical power with respect to the 2-jet one, meaning that it has more constraining power on the normalisation of the W +jets background.

Figure 6.13 shows the comparison between the W +hf CR in the 1-lepton, 2-tag, 2-jet and 3-jet regions when using a one-bin only BDT_{VH} output as in the observation analysis, and the binned p_T^V distribution.

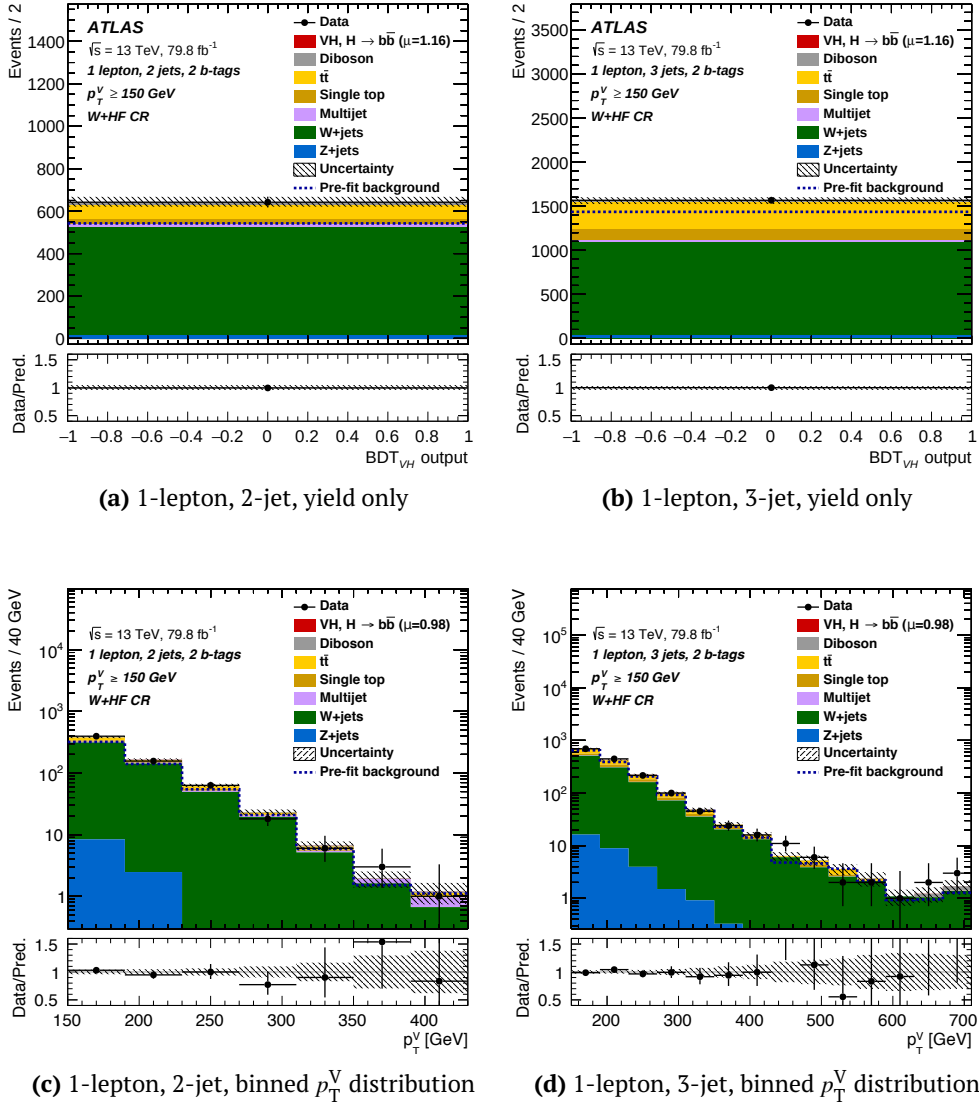


Figure 6.13: Figures (a) and (b) show the W +hf CR in the BDT_{VH} output for the 2-jet and 3-jet categories, respectively [13]. Figures (c) and (d) show the W +hf CR with the p_T^V distribution as discriminant in the 2-jet and 3-jet categories, respectively. It has to be noticed that the W +hf CR in the BDT_{VH} output is fitted using the signal strength μ obtained from a combined fit to the three channels, while the W +hf CR in the p_T^V shape has been fitted with the μ obtained from a standalone 1-lepton fit.

An improvement of 8.6% in the expected significance in the 1-lepton channel was estimated when using the binned p_T^V distribution in the W +hf CR instead of the simple

yield. Moreover, the impact of the W +jets p_T^V shape systematic on $\Delta\mu$, evaluated by the nuisance parameter ranking method, dropped from ± 0.16 to ± 0.06 and only the p_T^V shape in the 2-jet region appears as the third ranked systematic, as shown in Figure 6.14. To evaluate the impact of the W +jets modelling on the total uncertainty on the signal strength, the so-called **breakdown method** was used, which consists of evaluating the uncertainty of μ without a systematic (or a group of systematics) uncertainty and subtracting the resulting error quadratically from the full error obtained with all the systematics. With this method, a significant drop in the impact of the W +jets modelling on $\Delta\mu$ was observed, from 0.142 to 0.082.

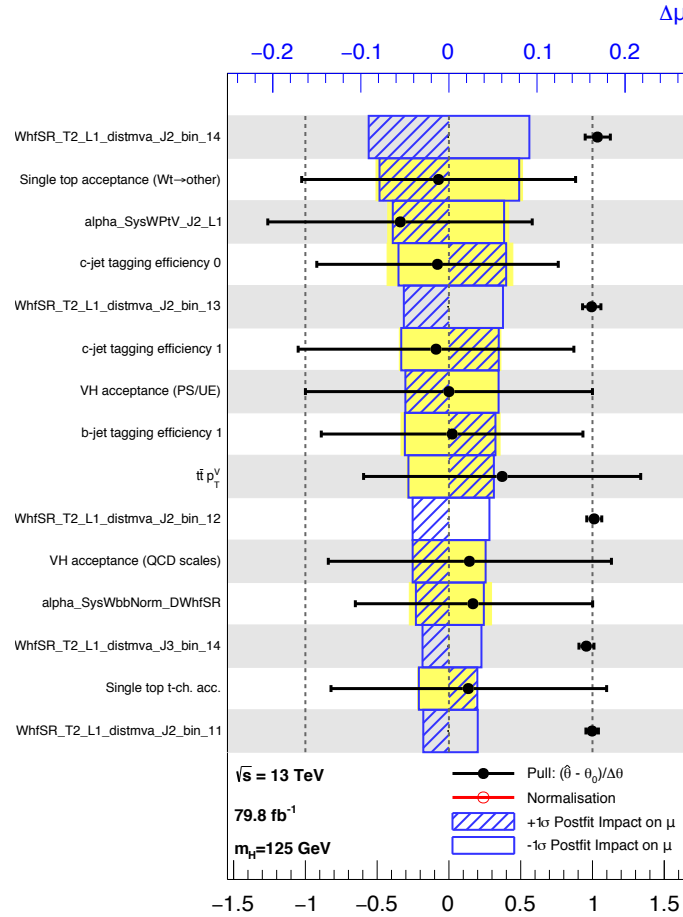


Figure 6.14: Impact of systematic uncertainties on the best fit value of the signal strength when using the binned p_T^V distribution in the W +hf control region. The systematic uncertainties are listed in decreasing order of their impact on μ . The boxes show the impact of the nuisance parameter on the total uncertainty $\Delta\mu$ and the hatched and open areas correspond to the up and down variations, respectively. The yellow bands indicate the impact of the nuisance parameters before the fit was performed. The black dots correspond to the deviation of the nuisance parameters from their nominal values (pulls) and the associated error bars correspond to their post-fit uncertainties. It can be seen that only the nuisance parameter associated to the W +jets p_T^V shape in the 2-jet region appears in the ranking plot at the third position. Also, its contribution on the total uncertainty $\Delta\mu$ has dropped to ± 0.06 .

6.7 Modelling Uncertainties

As seen in Section 6.2, the Monte Carlo generators used to simulate a specific process can differ for the modelling of the matrix element, the parton shower or the underlying event. The choice of a certain Monte Carlo generator instead of another can lead to differences in the final result. Therefore, an associated systematic uncertainty on the MC template of each background process is obtained by comparing the chosen nominal prediction to the alternative models, in order to quantify the variance of the differential cross-section, or the kinematic shapes. The modelling uncertainties can be divided into normalisation, acceptance and shape uncertainties.

6.7.1 Normalisation Uncertainties

The normalisation uncertainties are global values that change the normalisation of the MC template for a specific process in all regions of the analysis. The normalisation factors (NF) associated to the major backgrounds of this analysis, V +jets and $t\bar{t}$, are left free floating in the fit and are controlled by *unconstrained* NPs. Each NF is typically derived from an analysis region enriched in a particular background sample, such as a CR, and it is extrapolated to the other analysis regions through acceptance uncertainties. On the other hand, the normalisation associated to smaller samples, such as diboson or single-top, is controlled by *constrained* NPs that implement the theoretical knowledge on the respective cross sections.

6.7.2 Acceptance Uncertainties

The acceptance uncertainties are necessary to account for possible migration of the events among different analysis regions due to the choice of the generator. These are calculated as double-ratios defined as:

$$\frac{\text{Acceptance}[\text{Region}_A(\text{nominal MC})]}{\text{Acceptance}[\text{Region}_B(\text{nominal MC})]} \bigg/ \frac{\text{Acceptance}[\text{Region}_A(\text{alternative MC})]}{\text{Acceptance}[\text{Region}_B(\text{alternative MC})]}, \quad (6.34)$$

where Region_A (Region_B) is the yield of the nominal or alternative MC in the region A (B). In the case the variation originates from a change of the renormalization and factorization scales, the envelope is taken. In case different PDF sets are used, the one giving the maximal variation is considered. When a completely different generator is used, the difference with the nominal is taken. The total uncertainty is then calculated as the sum in quadrature of all these contributions.

The NPs associated to the acceptance uncertainties are implemented in the fit through Gaussian (or log-normal in some cases) constraining terms, as defined in Equation 6.19.

6.7.3 Shape Uncertainties

These uncertainties account for the variation in the shape of the kinematic distributions of the signal and the background due to the different prediction of the MC generators.

The shape uncertainties for the Z +jets, single-top and diboson backgrounds are derived on the most important kinematic variables entering the analysis-BDT, p_T^V and m_{bb} , as they were found to sufficiently cover the differences between MC generators on the analysis-BDT discriminant. A fit to the ratio between the nominal and the alternative generator is performed in all analysis categories and if the shapes are found to have similar impact on the distribution, then the largest variation is taken to keep a conservative approach. For W +jets and $t\bar{t}$, a BDT-based reweighting technique is exploited to discriminate between the nominal and the alternative generators, making use of all the kinematic variables entering the final analysis-BDT discriminant. The reason way a BDT-based reweighting procedure is employed lies in the fact that, due to the limited computational resources, the alternative sample typically contains fewer events with respect to the nominal. This means that fitting the alternative generator through the analysis-BDT, which would be the “natural” approach to derive a shape uncertainty accounting for the differences between two generators, can give rise to statistical fluctuations which might be difficult to smooth out with some parametric fit function.

More details on the 1-dimensional and multi-dimensional reweighting procedures are given in the following paragraphs.

1-dimensional Parametrisation The phase space of the $VH, H \rightarrow b\bar{b}$ analysis can be described by a set of kinematic variables $\mathbf{x} = (x_1, \dots, x_p)$ of dimension p and the MC predictions of the nominal and alternative generators can be described by the probability density functions $p(\mathbf{x} | \text{MC}_{\text{nom}})$ and $p(\mathbf{x} | \text{MC}_{\text{alt}})$, respectively. In a 1-dimensional parametrization, the p -dimensional phase space is reduced to a single dimension, $\mathbb{R}^p \rightarrow \mathbb{R}$, represented in this analysis by the m_{bb} or the p_T^V . The density ratio is given by:

$$r(\mathbf{x} \rightarrow m_{bb}/p_T^V) = \frac{p(\mathbf{x} \rightarrow m_{bb}/p_T^V | \text{MC}_{\text{alt}})}{p(\mathbf{x} \rightarrow m_{bb}/p_T^V | \text{MC}_{\text{nom}})}. \quad (6.35)$$

This is shown in Figure 6.15; the input space for each MC prediction is projected into a single observable $O = p_T^V$ or m_{bb} and the ratio of the two distributions is used to construct a shape uncertainty. This is then used to reweight all events produced by the nominal MC generator. The density ratio is typically derived from binned distributions of the observable, however, due to the limited available number of events for the alternative MC samples, it is preferred to perform a parametric fit to the unit-normalised⁶ density ratio in order to smooth out the statistical fluctuations.

It would be natural to extend the 1-dimensional reweighting to a multi-dimensional reweighting in which each dimension is represented by a kinematic variable. However,

⁶The unit-normalised ratio allows to remove all the differences between nominal and alternative prediction in terms of normalisation.

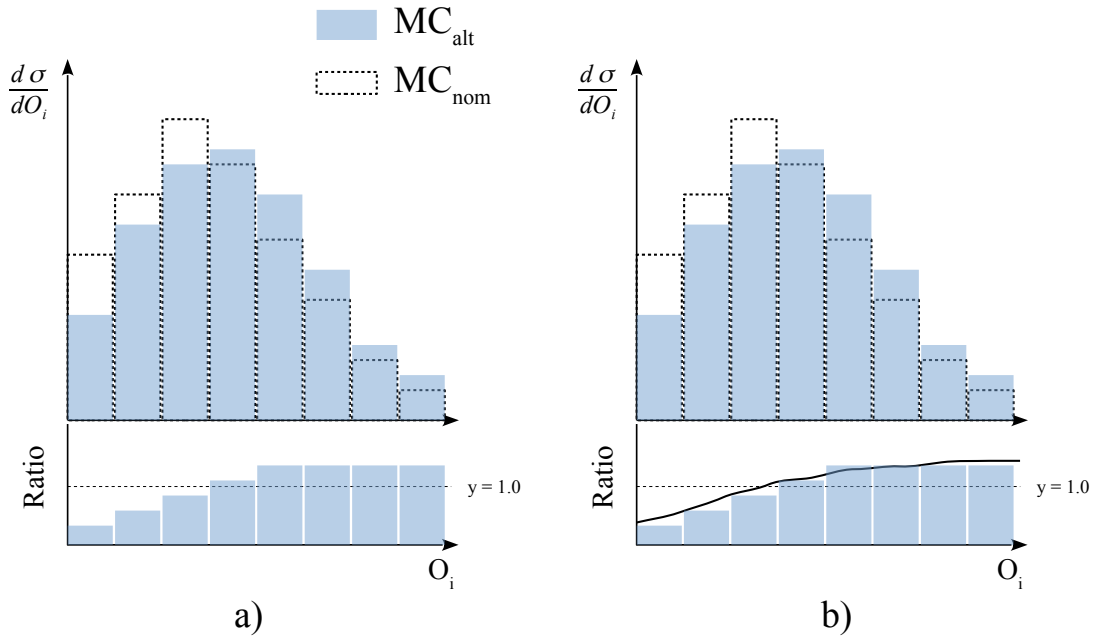


Figure 6.15: Illustration of the systematic uncertainty derived from the comparison between nominal and alternative differential cross-section in the observable O ; (a) shows the two MC predictions binned along with the ratio and (b) shows how a parametric fit (black solid line) is used to smooth the binned ratio. This ratio is then used to reweight all the kinematic variables describing the phase space of the analysis.

due to the limited number of events in the MC samples, it would not be possible to populate each cell of the multi-dimensional representation with a sufficient number of events without being subjected to statistical fluctuations.

This criticism is solved by the “BDTr” p -dimensional parametrization, described in the next paragraph, which gives an approximation of a multi-dimensional phase space.

BDTr p -Dimensional Parametrisation The main problem behind the 1-dimensional parametrization is its inability to capture the correlations among all the kinematic variables describing the analysis phase space. But even if it was able to capture these correlations, it is not guaranteed that the density ratio obtained in a given variable would be able to cover the differences between the generators in another variable. This information is lost during the $\mathbb{R}^p \rightarrow \mathbb{R}$ dimensional reduction. An example of this is shown in Figure 6.16 where the W +jets distributions of the m_{bb} and p_T^V variables are shown in the 1-lepton channel; the red histogram corresponds to the nominal prediction by SHERPA 2.2.1, while the blue histogram corresponds to the alternative one by MADGRAPH 5. Both distributions are normalised to one, therefore, they are shown in arbitrary units “a.u”. The ratio between the two distributions is performed with the alternative prediction as numerator and the nominal prediction as denominator. The shape uncertainty is derived fitting a functional form to the unit-normalised ratio in blue in the bottom panel, and then converting the fit function into a binned distribution to be compatible with the format used in the fit framework. The green and violet histograms in the bottom

panels in Figure 6.16 show the final shape uncertainties on the m_{bb} and p_T^V distributions, respectively. It can be seen that the p_T^V uncertainty covers the discrepancy between SHERPA 2.2.1 and MADGRAPH 5 on the p_T^V distribution but not on the m_{bb} one and vice versa. This is expected as the variables were found to be sufficiently uncorrelated in studies conducted in Run 1 [191].

An example of how the m_{bb} and p_T^V shape uncertainties are applied to one of the distributions entering the analysis-BDT, m_{top} , is shown in Figure 6.16 (c). In the top panel, the m_{top} distribution for the W +jets background is shown in the 1-lepton channel for the nominal (red) and alternative (blue) predictions. In the bottom panel, the unit-normalised ratio between the alternative and the nominal distributions is shown with the superimposed m_{bb} (green) and p_T^V (violet) shape systematics derived with the procedure explained above. It can be seen that neither the m_{bb} shape uncertainty nor the p_T^V one are able to cover the differences between the nominal and the alternative MC predictions. The brown binned ratio shows the attempt to derive an additional systematic uncertainty on m_{top} which was found to have no effect on the final BDT discriminant and was therefore discarded.

To overcome this criticism and take advantage of the information from all kinematic variables which describe the analysis phase space, the full p -dimensional density ratio in Equation 6.35 can be calculated by a MVA technique, as demonstrated in Ref. [199]. This makes use of a one dimensional Neural Network or BDT classifier response as a proxy of the p -dimensional density ratio. It can be demonstrated that if the response of the classifier $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ with given predictions MC_{nom} and MC_{alt} is a monotonic function with the density ratio in Equation 6.35, then the density ratio is invariant under the transformation $\mathbf{x} \rightarrow u = s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$:

$$r(\mathbf{x}) \rightarrow r(u) = \frac{p(u | \text{MC}_{\text{alt}})}{p(u | \text{MC}_{\text{nom}})}. \quad (6.36)$$

Therefore, this parametrization allows the exact same procedure as the 1-dimensional approach, but instead of using an observable O (with $O = m_{bb}$ or p_T^V in the 1-dimensional procedure) one can use the classifier response $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ directly. However, it is not guaranteed that after the multi-dimensional reweighting the nominal phase space will look exactly like the alternative one, as this procedure only approximates a p -dimensional phase space into a single dimension inevitably losing some information in some corner of the phase space. For this reason, this procedure has to be verified by the so-called “validation tests” in which the mapped nominal sample is compared to the alternative one and their agreement is evaluated through the $\chi^2/n_{\text{d.o.f}}$ metric.

For simplicity, in the $VH, H \rightarrow b\bar{b}$ analysis, this methodology exploits a BDT to derive the function $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ and, for this reason, the procedure is defined as “**BDT reweighting approach**” (**BDTr**). It can be summarized in four main steps:

1. **Train** a BDT classifier on a set of p observables $\vec{x} = (O_1, O_2, \dots, O_p)$ to separate the nominal prediction, which is treated as signal, from the alternative MC prediction,

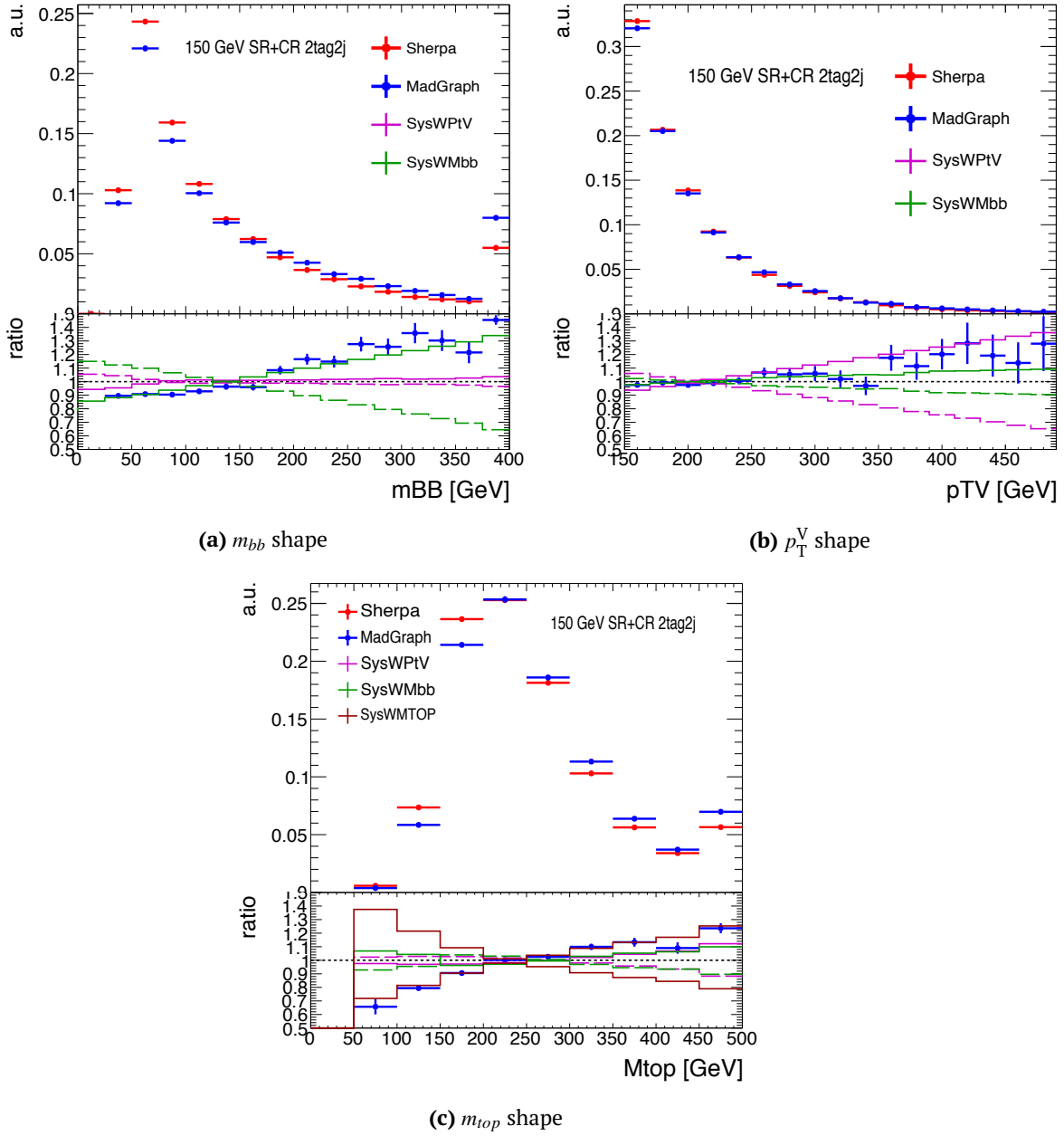


Figure 6.16: Shape uncertainties for W +jets in the 1-lepton channel derived on m_{bb} (a) and p_T^V (b) and applied to one of the key kinematic variables of this channel, the top mass m_{top} (c). In the top panel, there are the unit-normalised distributions of the variable predicted by the nominal (red) and alternative (blue) generators. The label a.u. refers to “arbitrary units”. In the bottom panel, the ratio is performed with the alternative prediction as numerator and the nominal one as denominator. The violet and green lines in the bottom panel correspond to the shape uncertainties on m_{bb} and p_T^V , respectively. These have been derived fitting a functional form (a polynomial of the first order in this case) to the unit-normalised ratio of the alternative and the nominal predictions, and then converting the fit function into a histogram. Figure (c) shows an example of how the m_{bb} and p_T^V shape uncertainties are applied to the distribution of m_{top} . It can be seen that in this case, both uncertainties are not able to cover the difference between the nominal and alternative predictions of the m_{top} distribution.

which is treated as background;

2. **Evaluate** the classifier response in order to get the output function $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$;
3. Derive the **density ratio** as in Equation 6.36;
4. **Reweight** the nominal MC distribution of all observables $O_1, O_2, \dots, O_p$ by the density ratio derived from step 3.

An example of the approach is illustrated in Figure 6.17 for a $p = 3$ -dimensional phase space. A Neural Network (a BDT in the case of the $VH, H \rightarrow b\bar{b}$ analysis) is trained with the distributions of three input observables O_1, O_2 and O_3 in order to separate the nominal MC prediction (MC_{nom}) from the alternative one (MC_{alt}) giving a single scalar function $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ as output. A systematic uncertainty is then derived through a functional fit to the unit-normalised ratio of the two MC predictions in $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$.

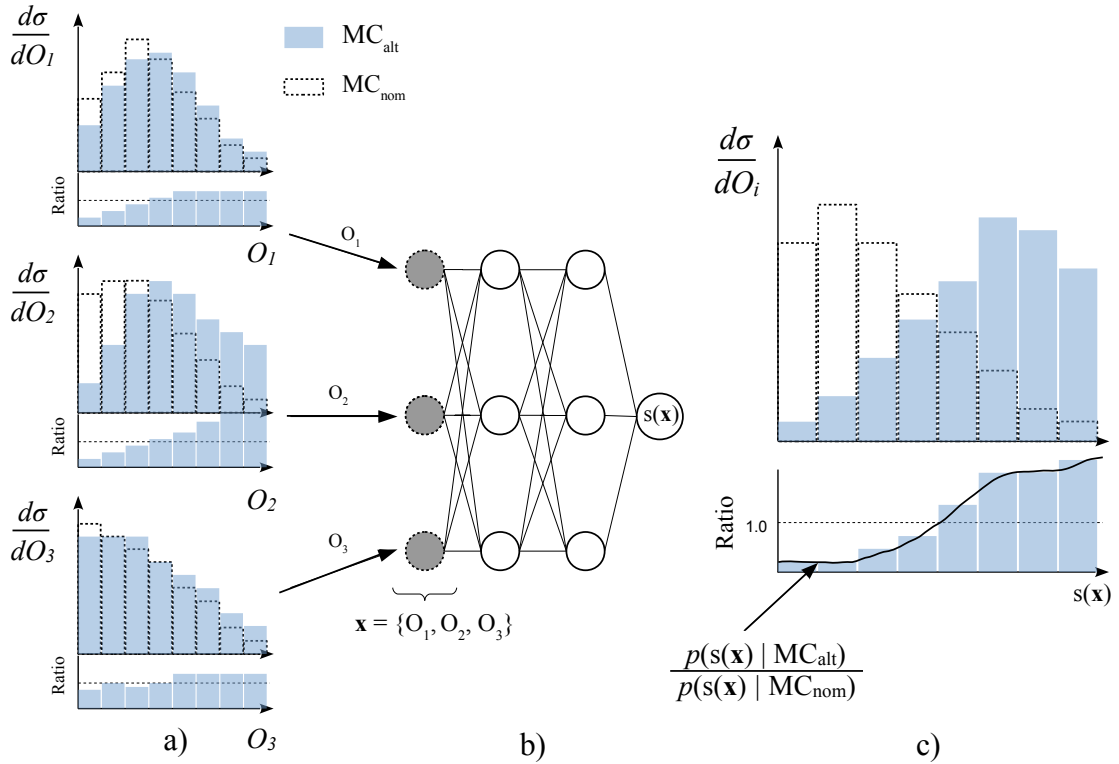


Figure 6.17: The distribution of three observables O_1, O_2 and O_3 for both nominal (MC_{nom}) and alternative (MC_{alt}) predictions (a) are inputs to the training of a MV analysis, a BDT in the case of the $VH, H \rightarrow b\bar{b}$ one, which gives a scalar function $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ as output (b). The systematic uncertainty is derived through a fit to the unit-normalised ratio between the Monte Carlo predictions in $s(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ (c) and used to reweight the nominal MC events.

Hybrid $p - n$ dimensional parametrization As seen in Section 6.4, p_{T}^{V} was chosen as fit discriminant in the CRs. This was justified by the fact that the analysis is split in p_{T}^{V} bins, each of them having two one-bin CRs, which is equivalent to fit p_{T}^{V} with

a coarse binning in CR. Starting from this concept, it was decided to separate the p_T^V shape uncertainty from the BDTr ‘mapping’ allowing the shape derived from the BDTr procedure to encode all differences in the remaining variables of the analysis-BDT. This is effectively a $\mathbb{R}^p \rightarrow \mathbb{R}^2$ reduction of the p -dimensional phase space into 2 dimensions. This approach was further supported by two observations regarding the modelling of the W +jets and $t\bar{t}$ backgrounds to which the BDTr method is applied. In the case of the W +jets background, the fit would have a single degree of freedom to correct for data-MC mismodelling in all 18 analysis regions in which this background has a significant contribution; in fact, the difference between SHERPA 2.2.1 and MADGRAPH 5 is the only generator comparison with a difference large enough to be accounted for⁷. To give the fit more freedom to accommodate for data-MC disagreement, it is convenient to define a new shape systematic associated to the p_T^V variable.

Concerning the $t\bar{t}$ background, it is convenient to factorize the p_T^V systematic to be able to cover a known mismodelling of the data at high p_T^V values; indeed, the nominal $t\bar{t}$ generator POWHEG + PYTHIA8 has a harder predicted p_T^V spectrum than data.

The procedure for this **hybrid BDTr approach** is illustrated in Figure 6.18. Prior to training the BDT classifier to separate the two MC predictions, the input space of the nominal generator is weighted by the density ratio of a subset of the dimensions of the input space, $\mathbb{R}^n \subset \mathbb{R}^p$. In the case of the $VH, H \rightarrow b\bar{b}$ analysis, $n = 1$ and it is represented by the p_T^V . The density ratio on this variable is derived as:

$$r(\mathbf{x} \rightarrow p_T^V) = \frac{p(\mathbf{x} \rightarrow p_T^V | \text{MC}_{\text{alt}})}{p(\mathbf{x} \rightarrow p_T^V | \text{MC}_{\text{nom}})}, \quad (6.37)$$

and it is used to map the nominal input space to look like the alternative one, as shown in Figure 6.18 (a) and (b). Afterwards, the p_T^V -reweighted input space of the nominal generator and the unaltered alternative one are used to train the BDT classifier to get the scalar function $s'(\mathbf{x}, \text{MC}_{\text{nom}}^{\text{weighted}}, \text{MC}_{\text{alt}})$. Here the input space $\mathbf{x}$ includes the p_T^V variable which, therefore, enters the training of the classifier. The classifier response is then used to derive a density ratio $r(s'(\mathbf{x}, \text{MC}_{\text{nom}}^{\text{weighted}}, \text{MC}_{\text{alt}}))$ which encodes all the remaining differences between the nominal and alternative input spaces but, by construction, is not able to learn anything about the p_T^V dependence. This procedure, which from now on will be addressed as “ **p_T^V -factorized BDTr approach**”, is summarised as follows:

1. Derive the p_T^V **density ratio** $r(p_T^V)$ from the comparison of its nominal and alternative MC prediction. This forms the p_T^V shape systematic;
2. **Reweight** the nominal MC input space by the density ratio $r(p_T^V)$;

⁷Previous studies performed by the author of this thesis have shown that both scales and PDF variations are largely covered by the difference between SHERPA 2.2.1 and MADGRAPH 5. Example distributions are reported in Figure B.73 and B.74 in Appendix B.6 for the largest contribution among the scale variations (μ_R and μ_F simultaneously doubled and halved) and for 1-lepton W +jets events having $p_T^V > 150$ GeV in 2-jet and 3-jet and for the sum of signal and control regions.

3. **Train** the BDT classifier to separate the p_T^V -reweighted nominal MC prediction from the unaltered alternative one;
4. **Evaluate** the classifier response for the two predictions, $s'(\mathbf{x}, \text{MC}_{\text{nom}}^{\text{weighted}}, \text{MC}_{\text{alt}})$ and derive the density ratio between the respective classifier scores ;
5. **Reweight** the original nominal MC prediction by the derived $r(s'(\mathbf{x}, \text{MC}_{\text{nom}}^{\text{weighted}}, \text{MC}_{\text{alt}}))$ density ratio, that corresponds to the *BDTr shape uncertainty*.

It was checked that the last step of this procedure has no effect on the p_T^V distribution, despite p_T^V being included in the training of the classifier. This is shown in Figure 6.19 for the p_T^V distribution of the W +jets background in the 1-lepton channel, 2-tag, 2-jet region, and for $p_T^V > 75$ GeV. The p_T^V and the BDTr shape uncertainties are considered as two separate and uncorrelated uncertainties in the fit model of the $VH, H \rightarrow b\bar{b}$ analysis.

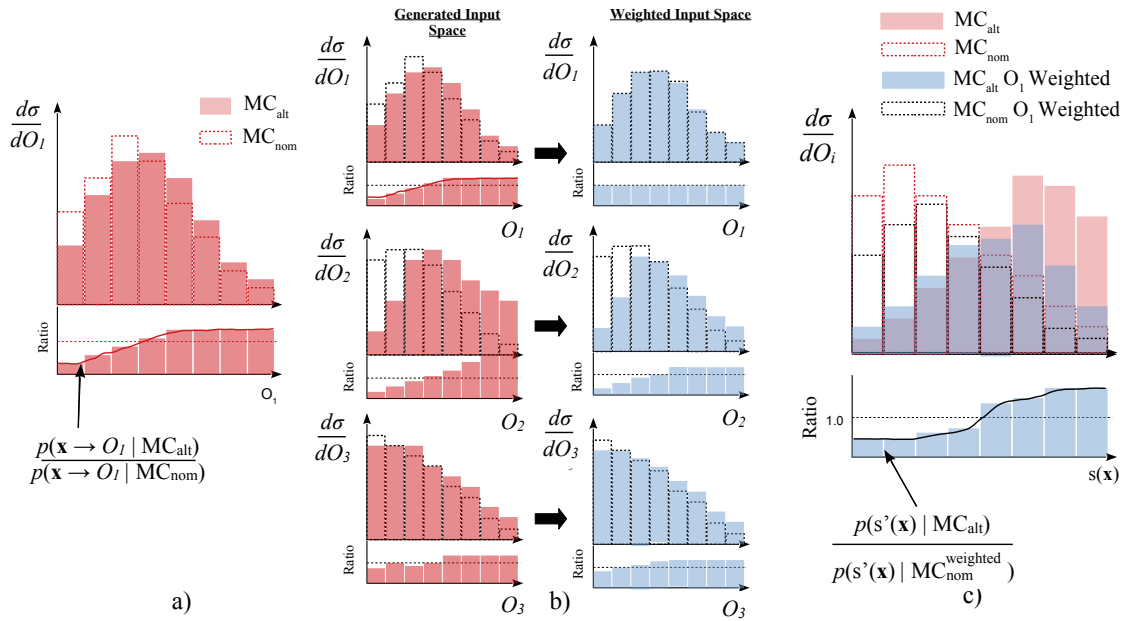


Figure 6.18: The density ratio is derived from the comparison between the nominal MC prediction (MC_{nom}) and the alternative one (MC_{alt}) in the observable O (p_T^V in the this analysis), as shown in (a). This is used to reweight all remaining observables of the nominal input space as in (b) to remove whatever dependency from the observable O . The re-weighted nominal input space and the unaltered alternative one are used to train a BDT classifier with output $s'(\mathbf{x}, \text{MC}_{\text{nom}}, \text{MC}_{\text{alt}})$ from which one can derive the BDTr shape systematic as density ratio between the reweighted nominal and unaltered alternative MC predictions. Figure (c) shows the resulting classifier; the filled red and dotted red distributions match those of Figure 6.17 (b) simply to demonstrate how the MVA response of the hybrid MVA differs from the non-hybrid one.

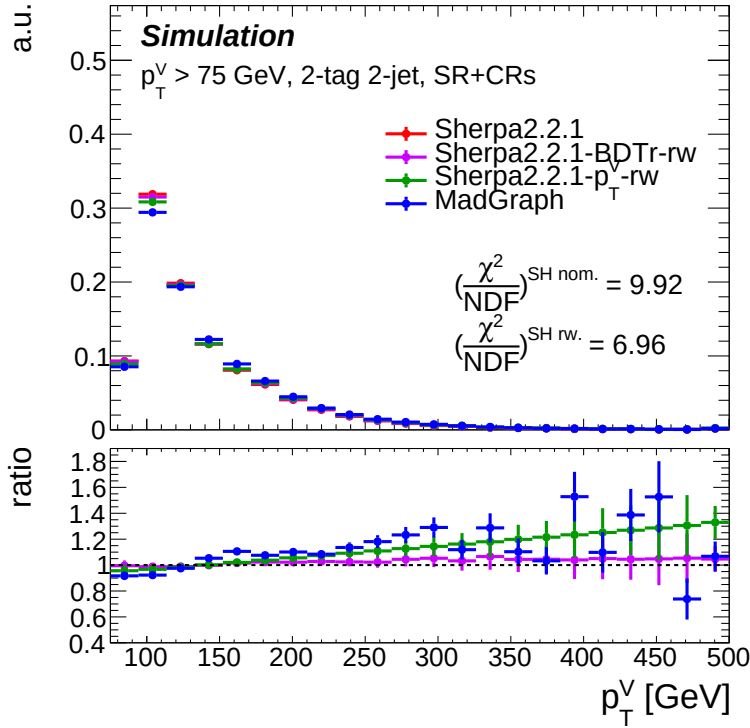


Figure 6.19: The p_T^V distribution for the W +jets background is shown in the 1-lepton channel, 2-tag, 2-jet region, and $p_T^V > 75 \text{ GeV}$. The red histogram corresponds to the nominal prediction by SHERPA 2.2.1, while the blue histogram represents the alternative prediction by MADGRAPH 5. The nominal SHERPA 2.2.1 prediction is first reweighted by the $r(p_T^V)$ density ratio and the outcome is the green histogram. Then, the nominal SHERPA 2.2.1 prediction, already reweighted by the p_T^V density ratio, is trained against the unaltered alternative prediction by MADGRAPH 5 and the ratio between the respective classifier scores is derived. This is used to reweight the nominal SHERPA 2.2.1 prediction (red) and the output is shown with the magenta histogram. It can be observed that the BDTr shape systematic derived with this procedure has no effect on the shape of the p_T^V distribution, as the training was not able to learn anything about it.

6.8 Signal Modelling

Following the recommendations of the LHC Higgs Cross Section Working Group [163, 164, 200–202], QCD scale uncertainties of 0.7% and 25% on the cross sections of the $qq \rightarrow VH$ (entering all lepton channels) and $gg \rightarrow ZH$ (entering only the 2-lepton channel) processes, respectively, are obtained by varying the μ_R and μ_F independently by a factor 1/3 and 3 with respect to their original value. A value of 1.7% is used as systematic uncertainty on the $H \rightarrow b\bar{b}$ branching ratio for all VH channels. Moreover, although the total VH cross section is calculated at NLO in EW, sizeable effects on the shape of the p_T^V distribution are expected from the EW corrections. Hence, a dedicated p_T^V shape uncertainty, one for each VH process, is extracted from the comparison between the NLO EW corrected p_T^V distribution and the nominal prediction. Further uncertainties are derived in the context of the *Simplified Template Cross Section*, illustrated in the next Section.

6.8.1 Simplified Template Cross Section

The cross section of the $VH, H \rightarrow b\bar{b}$ signal is calculated in the context of the *Simplified Template Cross Section* (STXS) Framework which has been developed to maximise the sensitivity of the cross section measurements while minimising the model dependence of the theoretical prediction. At the same time, it provides a common framework for the combination of measurements in different decay channels, as well as among different experiments. The STXS bins in which the analysis is divided represents mutually exclusive regions in which the cross section is measured. Different “stages” can be defined, corresponding to increasing granularity, as explained in Ref. [203]: in stage-0 the cross sections are measured for all Higgs boson production modes with the Higgs boson produced in the rapidity range $|y_H| < 2.5$, in stage 1.1 each production mode is further split according to the number of jets, the transverse momentum of the Higgs boson, or that of the vector boson V in the case of the $VH, H \rightarrow b\bar{b}$ analysis⁸. This is shown in Figure 6.20. For the full Run 2 $VH, H \rightarrow b\bar{b}$ analysis, *stage-1.2* [204, 205] is adopted which corresponds to stage-1.1 merged down to the five bins summarised in Table 6.10, together with the corresponding reconstructed categories. The three processes $qq \rightarrow WH$, $qq \rightarrow ZH$ and $gg \rightarrow ZH$, are divided into $p_T^{V,t}$ bins (where t stands for “truth”) using the truth level information and the corresponding reconstructed categories are shown.

Figure 6.21 shows the migration matrices between the STXS bins, defined at truth level, and the corresponding reconstructed categories used in the analysis and listed in Table 6.10. Thanks to the very good resolution in the reconstructed p_T^V , the reconstructed categories are dominated by their corresponding STXS bin.

However, migration uncertainties are estimated to account for event migrations across the reconstructed p_T^V and jet bins. These are derived from QCD scale variations μ_R and μ_F , whose value is doubled or halved with respect to the nominal one.

⁸Stage 1.0 is identical to stage 1.1 but without the p_T^V boundaries at 75 and 400 GeV.

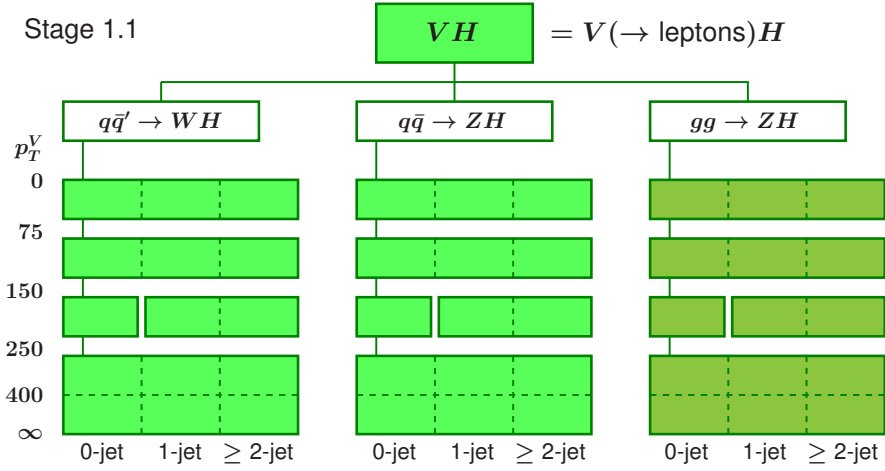


Figure 6.20: Stage 1.1 bins for the VH ($V \rightarrow \text{leptons}$) production mode. The total cross section is split in three main processes: $q\bar{q}' \rightarrow WH$, $q\bar{q} \rightarrow ZH$, and $gg \rightarrow ZH$, and each process is further split into p_T^V bins with boundaries 75, 150, 250, and 400 GeV. Each p_T^V bin is then split in 0-jet, 1-jet and ≥ 2 -jet categories [205].

Table 6.10: Simplified template cross-section regions and the corresponding reconstructed analysis regions. The current analysis is not sensitive to the regions WH , $p_T^{W,t} < 150$ GeV and ZH , $p_T^{Z,t} < 75$ GeV, and their cross-sections are fixed to the SM prediction within their theoretical uncertainties [146].

Process	STXS region	Reconstructed analysis regions		
	$p_T^{V,t}$ interval	Number of leptons	p_T^V interval	Number of jets
WH	150–250 GeV	1	150–250 GeV	2, 3
WH	> 250 GeV	1	> 250 GeV	2, 3
ZH	75–150 GeV	2	75–150 GeV	2, ≥ 3
ZH	150–250 GeV	0	150–250 GeV	2, 3
		2	150–250 GeV	2, ≥ 3
ZH	> 250 GeV	0	> 250 GeV	2, 3
		2	> 250 GeV	2, ≥ 3

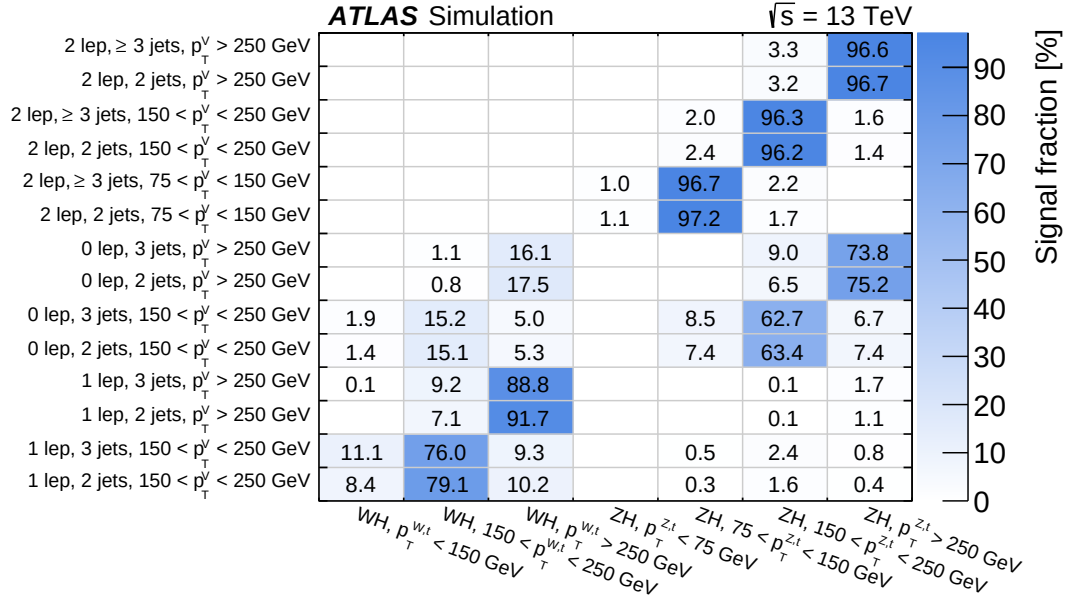


Figure 6.21: The predicted fraction of signal events (in percent) in the reconstructed event categories (y-axis) from each STXS truth signal region (x-axis). Entries with event yield below 0.1 or signal fractions below 0.1% are not shown [146].

Acceptance uncertainties across STXS bins are also considered and come from the comparison of the nominal samples with their internal PDF variations or with models with different underlying event, parton shower and hadronization (UEPS models).

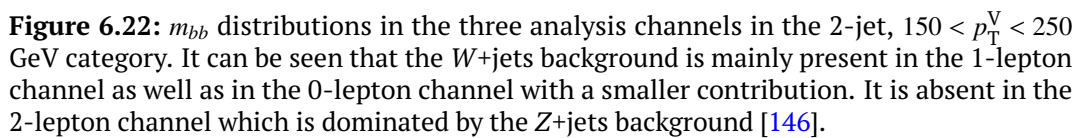
Shape uncertainties on the m_{bb} distribution are also considered. These are needed to account for the PS and UE variations and the QCD scale variations. The first are estimated from the comparison of the nominal POWHEG + MiNLO + PYTHIA8 with the alternative POWHEG + MiNLO + HERWIG7. The second is estimated by comparing the nominal POWHEG + MiNLO + PYTHIA8 samples to those generated with varied μ_R and μ_F .

6.9 Modelling of the W +jets background

The W +jets background is one of the most important backgrounds in the $VH, H \rightarrow b\bar{b}$ analysis. The most significant contribution is in the 1-lepton channel, however a non-negligible contamination can be found in the 0-lepton channel as well, as shown in the m_{bb} distributions in Figure 6.22 for the 2-jet category in the $150 < p_T^V < 250$ GeV bin. It is absent in the 2-lepton channel which is dominated by the Z +jets background.

The W +jets background, as well as the Z +jets one, can be divided into three main components based on the flavour composition of the event. To identify the flavour of the signal jets, the following procedure is performed:

1. if a b -hadron is found within a cone of $\Delta R = 0.3$ centred along the jet axis, the jet is labelled as a b -jet;
2. if a jet is not b -labelled and a c -hadron is found within a cone of $\Delta R = 0.3$ centred



along the jet axis, the jet is identified as a c -jet;

3. if a jet is not identified as a b -jet or a c -jet, the labelling hypothesis of τ -jet is tried;
4. if a jet is neither a b -jet, a c -jet, nor a τ -jet, then it is labelled as a l -jet.

The events are then classified according to the flavour pair: W + heavy flavours (W +hf), which includes $W + bb$, $W + bc$, $W + cc$ and $W + bl$, then $W + cl$ and $W + l$. Table 6.11 shows the relative composition of the four processes entering the W +hf definition in the 0-lepton and 1-lepton channels and for all the analysis regions, inclusively in p_T^V . It can be seen that the dominant contribution, after applying the event selection and the b -tagging requirement of 2 b -tagged jets, is represented by the $W + bb$ process which has the exact same final state as the $VH, H \rightarrow b\bar{b}$ signal.

Table 6.11: The fractional contribution of each W +hf sub-processes to the total W +hf yields in all analysis regions of the 1-lepton and 0-lepton channels.

1-lepton						
Process	2 jet			3 jet		
	Low- ΔR CR	SR	High- ΔR CR	Low- ΔR CR	SR	High- ΔR CR
$W + bb$	0.93	0.89	0.77	0.94	0.90	0.77
$W + bc$	0.03	0.07	0.15	0.03	0.05	0.13
$W + bl$	0.01	0.01	0.03	0.01	0.02	0.05
$W + cc$	0.03	0.03	0.05	0.03	0.03	0.04
0-lepton						
$W + bb$	0.94	0.78	0.90	0.94	0.79	0.91
$W + bc$	0.02	0.10	0.05	0.02	0.10	0.04
$W + bl$	0.01	0.09	0.03	0.01	0.07	0.02
$W + cc$	0.03	0.03	0.02	0.03	0.04	0.03

The $W + cl$ and $W + l$ components are strongly suppressed by the b -tagging requirement and compose less than 1% of the total W +jets background in all analysis regions.

6.9.1 Normalisation and acceptance uncertainties

All normalisation and acceptance uncertainties on the W +jets process were derived in the 0-lepton and the 1-lepton channels from comparisons of different MC generators and their parameter variations, and using fully reconstructed events.

To exploit the large number of events in the W +jets sample, a separate floating normalisation is adopted for the 2-jet and 3-jet categories. The main constraint on these normalisations comes from the Low- ΔR CR which is enriched in W +jets events, as shown in Figure 6.23 for the 2-jet and 3-jet categories in the 1-lepton channel.

Due to the very small contribution of the $W + cl$ and $W + l$ components to the total W +jets background, a single normalisation uncertainty is defined for each component.

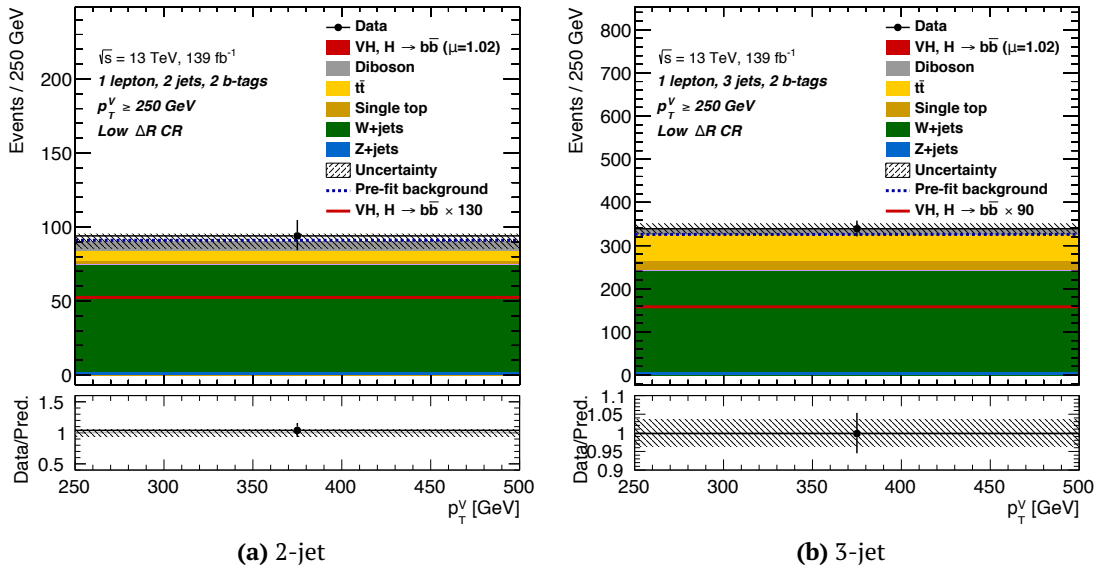


Figure 6.23: Background composition in the 1-lepton 2-tag, 2-jet (a) and 1-lepton 2-tag, 3-jet (b) Low- ΔR CR. It can be observed that the W +jets background is the dominant contribution.

The W +jets floating normalisations in the 2-jet and 3-jet regions are extrapolated from the control regions to the signal region of the 1-lepton channel and to the signal region of the 0-lepton channel via acceptance uncertainties, as explained in Section 6.7. A scheme explaining the procedure is reported in Figure 6.24.

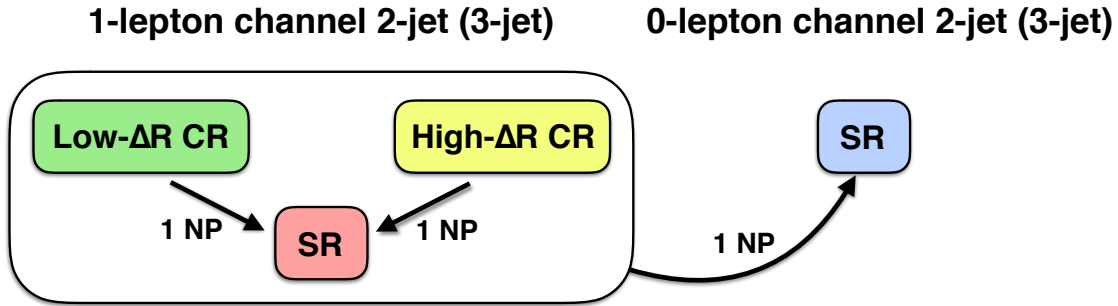


Figure 6.24: Scheme showing how the acceptance uncertainties are extrapolated to the analysis regions and the 0-lepton channel. An uncertainty controlling the migration effect across the analysis regions in the 1-lepton channel is calculated for each jet category in the Low- ΔR CR and High- ΔR CR and extrapolated to the SR. In addition, an uncertainty on the ratio of the yields between the 0-lepton and the 1-lepton channel is calculated for each jet category on the full Low- ΔR CR + SR + High- ΔR CR phase space of the 1-lepton channel and extrapolated to the 0-lepton SR.

The acceptance uncertainties are calculated as in Equation 6.34 and the final uncertainty is defined by summing in quadrature the following contributions:

- the envelope of the effect of varying the renormalization and the factorization

scales;

- the maximal variation between the nominal SHERPA 2.2.1 sample and the one obtained by choosing an alternative PDF-set;
- the difference between SHERPA 2.2.1 and MADGRAPH 5 predictions.

Two acceptance uncertainties are considered:

1-lepton $\rightarrow$ 0-lepton An uncertainty on the ratio of the yields between the 0-lepton and the 1-lepton channel is calculated separately in the 2-jet and 3-jet categories and on the full Low- ΔR CR + SR + High- ΔR CR. It is obtained as a double ratio (see Equation 6.34) by comparing the nominal SHERPA 2.2.1 sample with its internal scale and PDF variations, and with the alternative MC prediction by MADGRAPH 5. The full breakdown of variations entering the final uncertainty is reported in Table 6.12. The variations called PDF13000 and PDF25300 correspond to the up and down variations, respectively, obtained when changing the PDF-set. However, only the maximal variation of the two enters the final uncertainty; in this case the up variation given by PDF13000 is taken. A conservative uncertainty of 5% is applied to the 2-jet and 3-jet categories of the 0-lepton channel as the stringent constraint is coming from the 1-lepton channel, due to the dedicated W +hf control region. This uncertainty is dominated by the effect of changing MC prediction from SHERPA 2.2.1 to MADGRAPH 5.

Table 6.12: Systematic uncertainties on the ratio of yields between the 0-lepton and the 1-lepton channel calculated on the full Low- ΔR CR + SR + High- ΔR CR /SR phase space and separately in the 2-jet and 3-jet categories. The terms μ_R and μ_F are the variations induced by doubling or halving the renormalization or factorization scales one at the time. The term $\mu_R + \mu_F$ corresponds to the variation produced when doubling or halving both scale variations together. The term PDF MAX corresponds to the maximal variation between PDF13000 and PDF25300. In this case, PDF MAX is given by PDF13000 and it is the only PDF variation entering the final uncertainty. The term SH vs MG indicates the comparison between the two MC predictions where SH stands for SHERPA 2.2.1 while MG stands for MADGRAPH 5.

Systematic	2j: SR + Low- ΔR CR + High- ΔR CR	3j: SR + Low- ΔR CR + High- ΔR CR
μ_R	+0.0 / -0.0 %	-0.8 / +0.5%
μ_F	$\pm 0.2\%$	+0.2 / -0.1 %
$\mu_R + \mu_F$	-0.2 / +0.1 %	+0.5 / -0.8 %
PDF13000	$\mp 0.6\%$	$\mp 0.4\%$
PDF25300	$\mp 0.8\%$	$\mp 0.7\%$
PDF MAX	$\mp 0.6\%$	$\mp 0.4\%$
SH vs MG	-2.9 / -2.9 %	-0.9 / -0.9 %
Final uncertainty	$\pm 2.9\%$	$\pm 1.2\%$

[Low- ΔR CR, SR, High- ΔR CR] Migration This uncertainty controls the migration effect of the W +jets events from the Low- ΔR CR and the High- ΔR CR to the SR. As

explained in Section 6.7, the shape uncertainties of the W +jets background are estimated using the p_T^V -factorized BDTr approach which defines two shape uncertainties, one for p_T^V and one encoding all the differences between SHERPA 2.2.1 and MADGRAPH 5 on the remaining kinematic variables describing the analysis. The latter, called BDTr shape uncertainty, already induces yield migration effects from the CRs to the SR when modifying the shape of the nominal MC prediction.

However, to increase the degrees of freedom that the fit can manipulate to adjust the data-MC agreement, it was decided to consider only the shape effect of the BDTr shape uncertainty and to define a new NP, applied to the Low- ΔR CR and the High- ΔR CR, to control the yield migration to the SR. The prior on this NP is taken as the sum in quadrature of the difference between SHERPA 2.2.1 and MADGRAPH 5 plus the differences in the scale and PDF variations (scales and PDF are added for completeness as the effect is fully dominated by the SHERPA 2.2.1 versus MADGRAPH 5 comparison) as shown in Table 6.13.

Table 6.13: Systematic uncertainties on Low- ΔR CR /SR and High- ΔR CR /SR extrapolation for W +hf in the 1-lepton channel. The terms μ_R and μ_F are the variations induced by doubling or halving the renormalization and the factorization scales one at the time. The term $\mu_R + \mu_F$ corresponds to the variation produced when doubling or halving both scale variations together. The term PDF MAX corresponds to the maximal variation between PDF13000 and PDF25300. In this case, PDF MAX is given by PDF13000 and it is the only PDF variation entering the final uncertainty. The term SH vs MG indicates the comparison between the two MC predictions where SH stands for SHERPA 2.2.1 while MG stands for MADGRAPH 5. The prior on the NP which controls the yield migration is taken from the sum in quadrature of the difference between SHERPA 2.2.1 and MADGRAPH 5 and the contribution of the $\mu_R + \mu_F$ and PDF variations (for both 0-lepton and 1-lepton channels).

Systematic	2j: SR/Low- ΔR CR	2j: SR/High- ΔR CR	3j: SR/Low- ΔR CR	3j: SR/High- ΔR CR
μ_R	+2.7 / -2.2 %	-1.3 / +1.2 %	+4.5 / -3.7 %	-2.4 / +2.2 %
μ_F	-0.1 / +0.0 %	-0.4 / +0.3 %	-0.4 / +0.3 %	+0.0 / +0.0 %
$\mu_R + \mu_F$	-2.3 / +2.3 %	+1.6 / -1.4 %	-3.5 / +4.1 %	+2.2 / -2.3 %
PDF13000	± 0.2 %	∓ 0.1 %	± 0.1 %	± 0.0 %
PDF25300	∓ 0.0 %	∓ 0.2 %	± 0.0 %	∓ 0.1 %
PDF MAX	± 0.2 %	∓ 0.1 %	± 0.1 %	± 0.0 %
SH vs MG	+7.3 / +7.3 %	-14.8 / -14.8 %	-3.7 / -3.7 %	-6.6 / -6.6 %
Final uncertainty	± 7.7 %	± 14.9 %	± 5.6 %	± 7.0 %

This approach is justified by the fact that the yield migrations induced by the BDTr shape uncertainty are compatible with those derived from the double ratio between SHERPA 2.2.1 and MADGRAPH 5, as shown in Table 6.14.

Concerning the shape, the NP associated to the BDTr shape uncertainty is implemented so that only the change in shape is considered. The relative sign of the assigned prior on the Low- ΔR CR, SR and High- ΔR CR boundaries is taken from the net migration effect observed when applying the BDTr shape uncertainty. This is done so that the sign

matches the “direction” induced by the BDT shape uncertainty. Indeed, a monotonic trend is observed from Low- ΔR CR to High- ΔR CR (see Table 6.14).

More details on the choice of decorrelating the normalisation effect of the BDT shape systematic from the shape one will be given in Section 6.13.

Table 6.14: Migration effects from the Low- ΔR CR to the SR and from High- ΔR CR to the SR induced by the BDT shape uncertainty (BDT-rw) compared to the acceptance uncertainty obtained as a double ratio between the nominal SHERPA 2.2.1 (SH) and the alternative MADGRAPH 5 (MG). The prior on the NP which controls the yield migration is taken from the sum in quadrature of the difference between SHERPA 2.2.1 and MADGRAPH 5 and the contribution of the $\mu_R + \mu_F$ and PDF variations (see Table 6.13). The relative sign of the prior is taken from the monotonic net migration induced by the BDT shape uncertainty on the first row (for both 0-lepton and 1-lepton channels).

0-lepton				
Systematics	2j: SR/Low- ΔR CR	2j: SR/High- ΔR CR	3j: SR/Low- ΔR CR	3j: SR/High- ΔR CR
BDT-rw	-7.3%	+15.7%	-2.7%	+6.7%
SH vs MG	+11.1/+11.1%	-14.8/-14.8%	-0.3/-0.3%	-4.8%/-4.8%
1-lepton				
BDT-rw	-10.1%	+21.5%	-1.4%	+9.5%
SH vs MG	-7.3/+7.3%	-14.8/-14.8%	-3.7/-3.7%	-6.6%/-6.6%

p_T^V Migration This uncertainty takes into account the relative acceptance between the various p_T^V bins of the analysis. In this case, no acceptance uncertainty is added across the different p_T^V bins as the p_T^V shape uncertainty is sufficient to account for migration effects. The full breakdown of the migration effects induced by this systematic is reported in Table B.32 in Appendix B.1 for both the 0-lepton and 1-lepton channels.

6.9.2 Flavour composition uncertainties

The W +hf background is composed of four main processes: $W + bb$, $W + bc$, $W + bl$ and $W + cc$. Uncertainties on the relative composition of these processes are estimated separately in the 1-lepton and 0-lepton channels and are implemented as normalisation uncertainties on each of the smaller components (bc , bl , cc) relative to the more dominant bb flavour. The same value of the NP is applied in a correlated fashion in all jet categories and analysis regions. This is justified by the relatively small contribution of the non- bb components of the W +hf background after the 2 b -tagged jets requirement (see Table 6.11), which makes the impact of the choice of the correlation scheme in the fit negligible.

The full breakdown of the variations which contribute to the final uncertainty reported in Table 6.15 is summarized in Table B.33 (Table B.34) for the bc/bb extrapolation in the 1-lepton (0-lepton) channel, in Table B.35 (Table B.36) for the bl/bb extrapolation in the 1-lepton (0-lepton) channel and in Table B.37 (Table B.38) for the cc/bb extrapolation in the 1-lepton (0-lepton) channel, all in the Appendix B.2. It can be seen that the largest

contribution comes from the comparison of the SHERPA 2.2.1 and MADGRAPH 5 predictions. The final uncertainties on the flavour composition are summarized in Table 6.15 with the name of the corresponding NP implemented in the fit.

Table 6.15: The priors on the relative acceptance variations for W +hf. The first column details the flavour components across which the acceptance variation is being considered, the second column quotes the name of the corresponding nuisance parameter in the Profile Likelihood Fit, the third contains the uncertainty on the prior and the fourth column the processes and categories to which this nuisance parameter is applied.

Category	Name of NP	Prior	Applied to
$W + bc/W + bb$	SysWbcWbbRatio	15%	$W + bc$ events (0-lepton)
		30%	$W + bc$ events (1-lepton)
$W + bl/W + bb$	SysWblWbbRatio	10%	$W + bl$ events (0-lepton)
		30%	$W + bl$ events (1-lepton)
$W + cc/W + bb$	SysWccWbbRatio	26%	$W + cc$ events (0-lepton)
		23%	$W + cc$ events (1-lepton)

Variations of the flavour composition are also partly accounted for by the BDT shape uncertainty. In fact, this encodes the differences between the MC predictions of SHERPA 2.2.1 and MADGRAPH 5 for all the input variables entering the MVA (except for the p_T^V distribution which is factorized out), including the m_{bb} one. It can be seen from Figure 6.25 that the W +hf components have an intrinsically different m_{bb} distribution, in particular between processes containing gluon splitting diagrams ($W + bb$ and $W + cc$) and processes not containing them ($W + bc$ and $W + bl$)⁹.

A summary of the W +jets normalisation and acceptance uncertainties, together with the name of the associated NP and the regions where they are applied, is reported in Table 6.16.

6.9.3 BDT-reweighting shape uncertainties

As outlined in Section 6.7, the full Run 2 analysis utilises a BDT multidimensional mapping technique to derive the shape uncertainties for the W +jets background. On the other hand, the $H \rightarrow b\bar{b}$ observation analysis used a 1-dimensional parametrization for two variables. For completeness, the studies performed by the author of this thesis for the $H \rightarrow b\bar{b}$ observation publication are summarized in Appendix B.3.

The BDT-reweighting technique can be performed using a p -dimensional or a p - n -dimensional parametrization. The approach was first validated starting from the p -dimensional parametrization and using events at particle-level with truth tagging applied, in order to increase the number of events of the training samples. These studies can be found in Appendix B.4.

⁹This plot dates to studies performed for the derivation of the flavour composition uncertainties in the context of the $H \rightarrow b\bar{b}$ observation analysis [13].

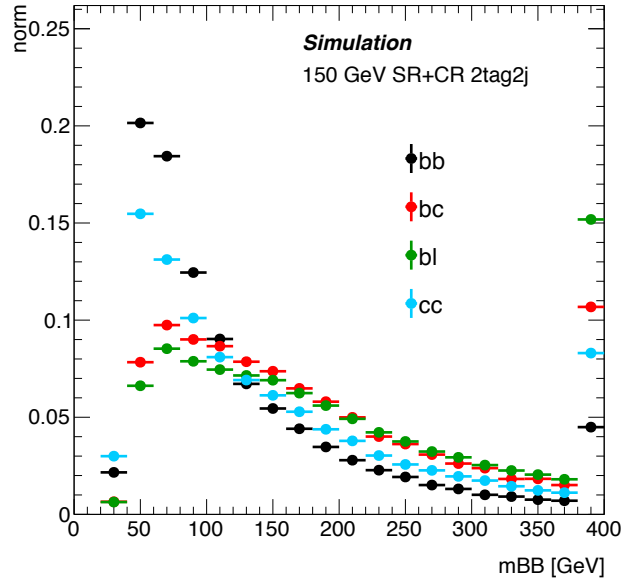


Figure 6.25: m_{bb} distribution in the 1-lepton channel for each heavy flavour category and for the 2-tag, 2-jet category. Differences are observed between processes containing gluon splitting diagrams ($W + bb$ and $W + cc$) and processes not containing them ($W + bc$ and $W + bl$). The last bin include also the events in the overflow region.

Table 6.16: Summary of the normalisation and acceptance uncertainties on the W +jets predictions in the various categories of the 0-lepton and 1-lepton channels. The second column corresponds to the region where the uncertainty is applied, the third one is the name of the NP associated to the systematic implemented in the fit framework. The normalisation NPs in the 2-jet and 3-jet categories have W_{bb} in the name only for historical reasons, but they refer to all W +hf events. In the fourth column, the effect of a $1 - \sigma$ variation is shown for systematics implemented with a prior in the fit model.

Process	Systematic	Name of NP	Prior in Region					
			1-Lepton					
			Low- ΔR CR 2-jet	SR 2-jet	High- ΔR CR 2-jet	Low- ΔR CR 3-jet	SR 3-jet	High- ΔR CR 3-jet
$W+l$	Wcl norm.	SysWclNorm	32%					
$W+cl$	Wl norm.	SysWlNorm	37%					
$W+hf$	2-jet $W+hf$ norm.	norm_Wbb_J2	Floating Normalisation (2-jet events only)					
$W+hf$	3-jet $W+hf$ norm.	norm_Wbb_J3	Floating Normalisation (3-jet events only)					
$W+hf$	$W+hf$ CR/SR extrap.	SysWbbCRSRextrap	-11.5%	–	+14.8%	-3.6%	–	+5.1%
0-Lepton								
$W+hf$	$W+hf$ 0-lep $\rightarrow$ 1-lep extrap.	SysWbbNorm_L0	5%					
$W+hf$	$W+hf$ CR/SR extrap	SysWbbCRSRextrap	-7.7%	–	+14.9%	-5.6%	–	+7.0%

However, as explained in Section 6.7.3, the hybrid p - n parametrization with $n = 2$ was chosen to increase the flexibility of the fit to accomodate for data/MC mismodelling, and the BDTr shape systematic was derived using a sample with a large number of fully reconstructed events with hybrid tagging applied, to be coherent with the W +jets (and $t\bar{t}$) MC templates used in the rest of the analysis.

In the following paragraphs, the hybrid p -2 parametrization of the BDTr shape uncertainty will be explained.

Derivation of the p_T^V shape uncertainty In both 0-lepton and 1-lepton channels, eight p_T^V shape uncertainties are derived according to the jet category (2-jet and 3-jet) and the flavour composition of the W +hf event ($W + bb$, $W + bc$, $W + bl$ and $W + cc$). Following the 1-dimensional procedure outlined in Section 6.7, a functional fit (first order polynomial in this case) is performed to the unit-normalised ratio between the SHERPA 2.2.1 and the MADGRAPH 5 predictions of the p_T^V distribution. In the 0-lepton channel, the E_T^{miss} distribution is used and the fit is performed in the $E_T^{\text{miss}} > 150$ GeV range. In the 1-lepton channel, the fit is performed in the $p_T^V > 75$ GeV range¹⁰.

Figure 6.26 and Figure 6.27 show the p_T^V shape systematics (called “SHERPA 2.2.1- p_T^V rw”) derived in the 1-lepton channel for each heavy flavour pair and for 2-jet and 3-jet events, respectively.

Given the small number of available W +jets events in the 0-lepton channel, the p_T^V shapes derived in the 1-lepton channel are also applied to the 0-lepton channel. The applicability of this approach is shown in Figure 6.28 for the 2-tag, 2-jet category and in Figure 6.29 for the 2-tag, 3-jet category, for each heavy flavour pair, and for the common $p_T^V > 150$ GeV range between the two lepton channels¹¹. The p_T^V shape uncertainties, derived in the 1-lepton channel, are therefore applied to both the 1-lepton and the 0-lepton channels. One NP for each jet category is implemented and treated as correlated accross the signal and control regions and the 0-lepton and 1-lepton channels.

Once the p_T^V shape uncertainty has been derived, this is used to reweight the full input space of the nominal MC prediction. This is characterized by the input variables entering the analysis-BDT that are reported in Table 6.7 for the three lepton channels. However, since the decision on adding the soft E_T^{miss} term (0-lepton channel), and the PCBT quantiles (0-lepton and 1-lepton channels) to the training of the analysis-BDT came after the studies performed for the derivation of the W +jets shape systematics, the BDT training used in the p_T^V -factorized BDTr approach does not include those variables.

¹⁰One of the planned improvements of the analysis in the 1-lepton channel was the addition of the so called “medium p_T^V ” region, $75 < p_T^V < 150$ GeV, as a new fitted region. However, the available bins (two CRs with a single bin and two SR in the discriminant distribution in both 2-jet and 3-jet categories) were found to be not sufficient to control the large number of events available in this region, causing strong constraints on many backgrounds that could not be easily (or necessarily) transported to the higher p_T^V regions. Due to the limited time during the publication process, no further study was performed in this direction and it was decided to drop the medium p_T^V region for this round of the analysis.

¹¹The p_T^V shape is derived from a unit-normalised ratio between SHERPA 2.2.1 and MADGRAPH 5 in the $p_T^V > 75$ GeV range, while the E_T^{miss} distribution is normalised to one in the $E_T^{\text{miss}} > 150$ GeV range.

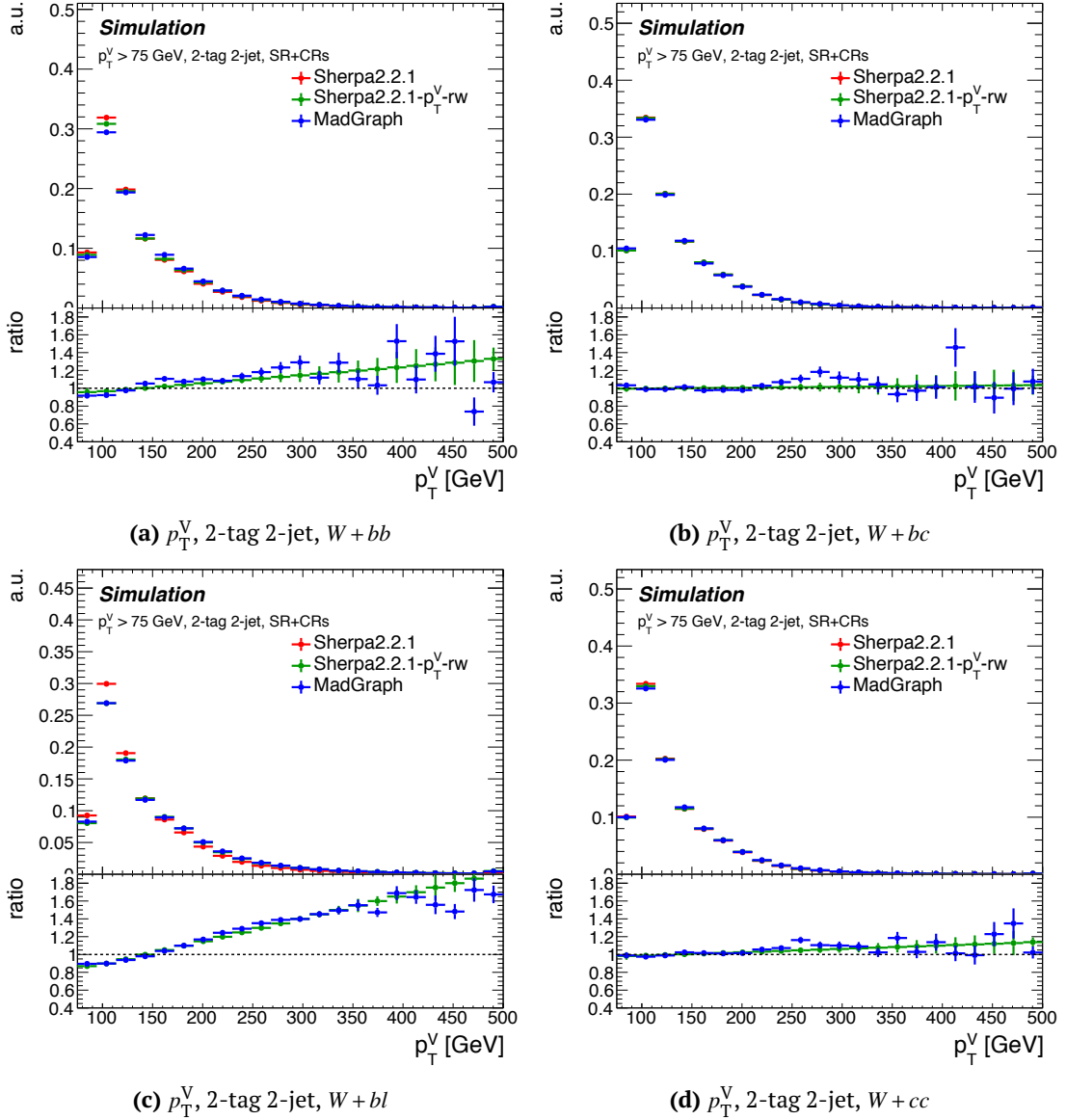


Figure 6.26: p_T^V shape uncertainties in the 1-lepton channel for each heavy flavour pair in the 2-jet category. In the top panel, the red and blue histograms correspond to the p_T^V distributions predicted by SHERPA 2.2.1 and MADGRAPH 5, respectively. In the bottom panel, the ratio is performed with the MADGRAPH 5 prediction at numerator and the SHERPA 2.2.1 one at denominator. The p_T^V binned shape uncertainty is plotted in green.

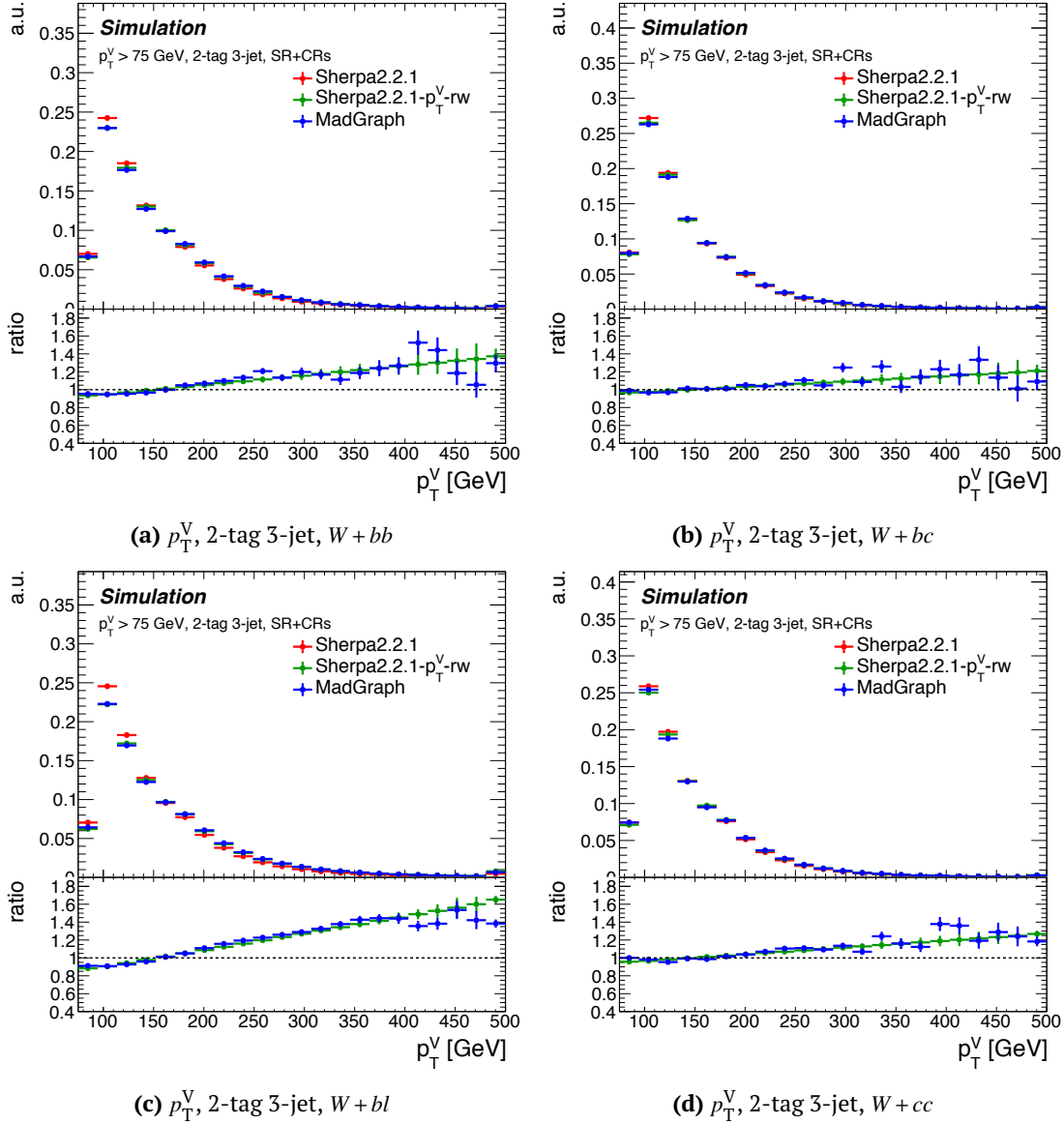


Figure 6.27: p_T^V shape uncertainties in the 1-lepton channel for each heavy flavour pair in the 3-jet category. In the top panel, the red and blue histograms correspond to the p_T^V distributions predicted by SHERPA 2.2.1 and MADGRAPH 5, respectively. In the bottom panel, the ratio is performed with the MADGRAPH 5 prediction at numerator and the SHERPA 2.2.1 one at denominator. The p_T^V binned shape uncertainty is plotted in green.

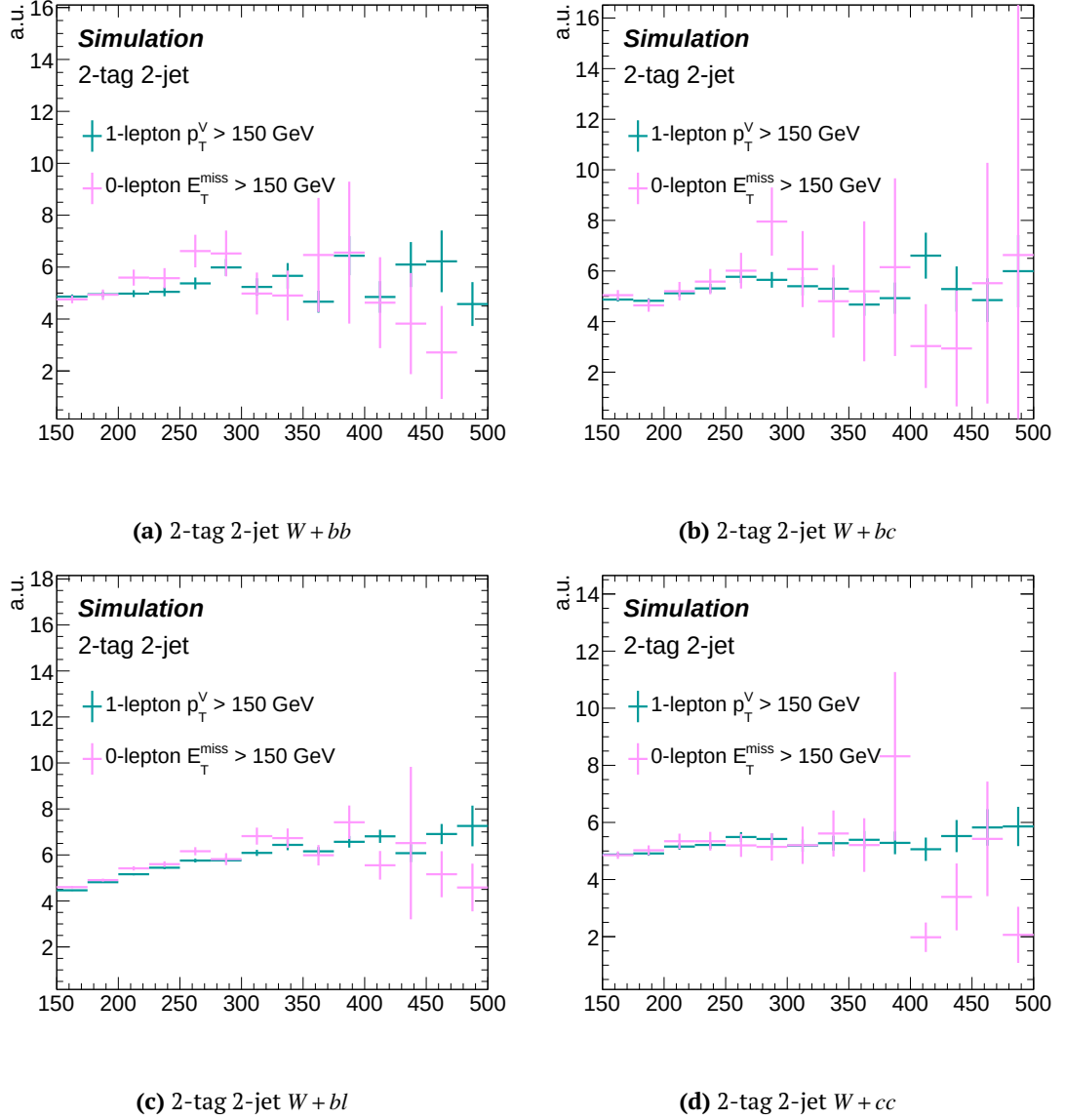


Figure 6.28: p_T^V and E_T^{miss} shape uncertainties derived in the 2-tag and 2-jet region of the 1-lepton and 0-lepton channels, respectively, for each heavy flavour pair.

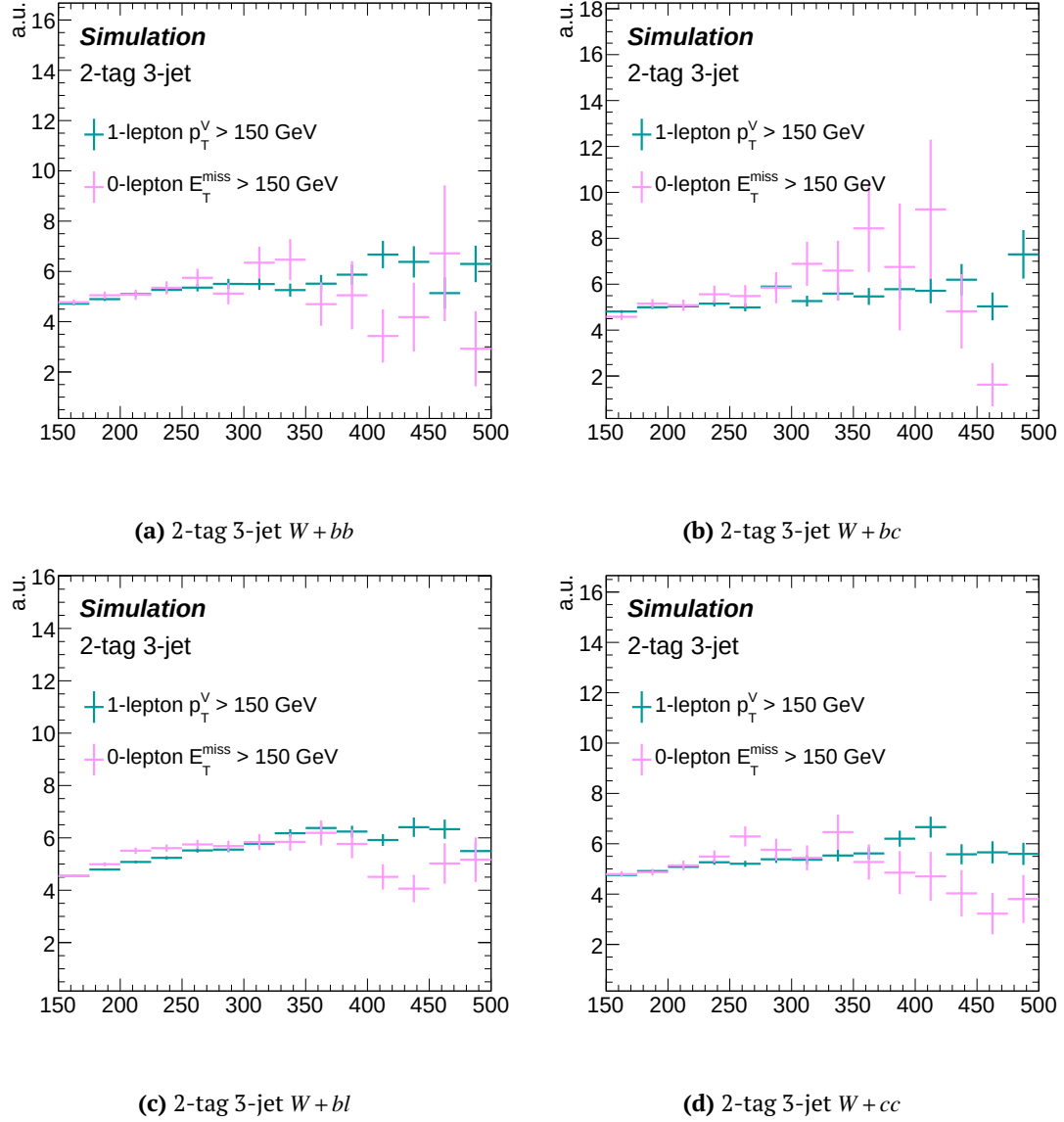


Figure 6.29: p_T^V and E_T^{miss} shape uncertainties derived in the 2-tag 3-jet region of the 1-lepton and 0-lepton channels, respectively, for each heavy flavour pair.

Hyper-parameter Optimization To achieve the best separation power in the BDT training procedure, a hyper-parameter optimization is performed in the 1-lepton channel as it has the largest number of MC events. The following studies have been performed using particle-level events, but similar conclusions can be drawn also using fully reconstructed events.

Seventeen trainings are performed on $W + bb$ events only, fixing the depth of the training to a value of 4 and varying the number of trees from 50 to 500 with a size of 50 and from 500 to 1200 with a size of 100. Eight trainings are performed fixing the number of trees to 200 and varying the depth from 4 to 11. The full set of $W + bb$ events entering each of these trainings is divided into two sub-sets, called A and B, in order to perform a 2-fold training. Then, the BDT output obtained on the testing sample, for each configuration of number of trees and depth, is scanned and the signal efficiencies and background rejection are calculated for different values of the threshold. Plotting the signal efficiency as a function of the background rejection for each BDT output value, one obtains the *Receiver Operating Characteristic* (ROC) curve, whose “area under the curve” is used to quantify the ability of the multiclassifier to resolve the signal (SHERPA 2.2.1) events from the background (MADGRAPH 5) events. If the AUC of the ROC is equal to 0.5, the discrimination power of the trained BDT is null. Figure 6.30 shows the AUC for each BDT output of the seventeen trainings as a function of the number of trees used in the trainings in both the 2-jet and 3-jet categories.

The label AonA identifies the AUC for a setup in which both training and testing are performed on the same set of events, while the label AonB shows the AUC for a setup in which the training is performed on the first half of events and the testing is performed on the second half of events. The AUC for the AonB setup shows that from a number of trees equal to 200, the discrimination power of the BDT does not increase anymore (black dashed line) and the BDT starts to be trained on statistical artefacts of the training samples, as shown by the AUC for the AonA setup (solid black line). For this reason, a value of 200 trees has been chosen.

Figure 6.31 shows the AUC for each BDT output of the 8 trainings as a function of the depth used in the trainings in both the 2-jet and 3-jet categories. It can be seen that the AUC for the AonB setup is almost constant while the one for the AonA setup rises at increasing values of the depth. For this reason, a value of 5 has been chosen, as higher values of the depth would mean overtraining the BDT.

Derivation of the BDT shape systematic Before training, the nominal SHERPA 2.2.1 events in both the 0-lepton and 1-lepton channels are weighted in each jet category and for each heavy flavour pair by the p_T^V density ratio obtained from the comparison between the nominal and the alternative p_T^V distribution in that region of the phase space.

Then, eight 2-fold trainings are performed in each lepton channel, one for each heavy flavour pair, and separately in the 2-jet and 3-jet categories. The nominal sample reweighted by the p_T^V density ratio is considered as signal, while the unaltered alternative sample is the background. The BDT scores for the signal and the background are

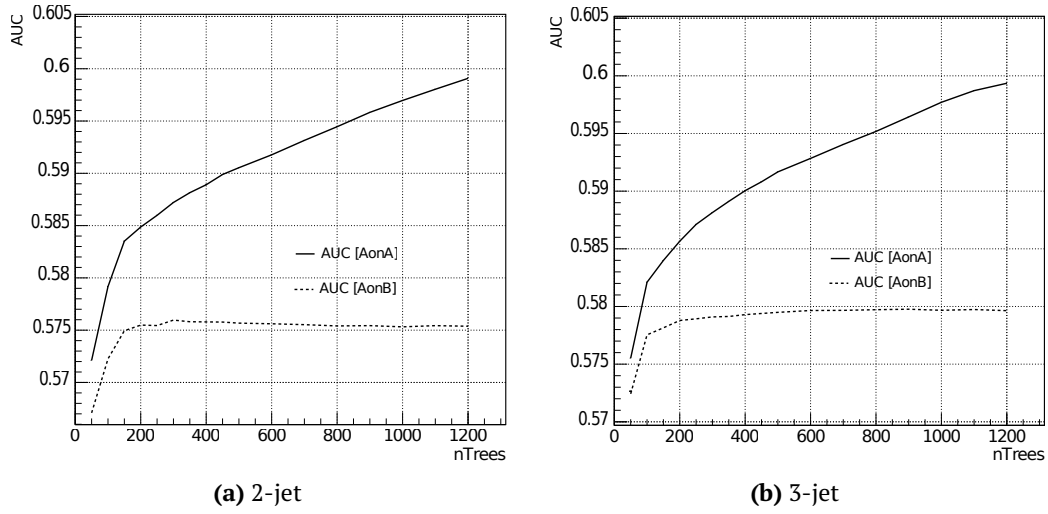


Figure 6.30: AUC for the two training setups AonA (continuous black line) and AonB (dashed black line) as a function of the number of trees nTrees. For a number of trees equal to 200, the discrimination power of the BDT classifier almost saturates. This means that adding more trees to the training does not increase its capability of distinguishing between the nominal (signal) sample and the alternative (background) sample.

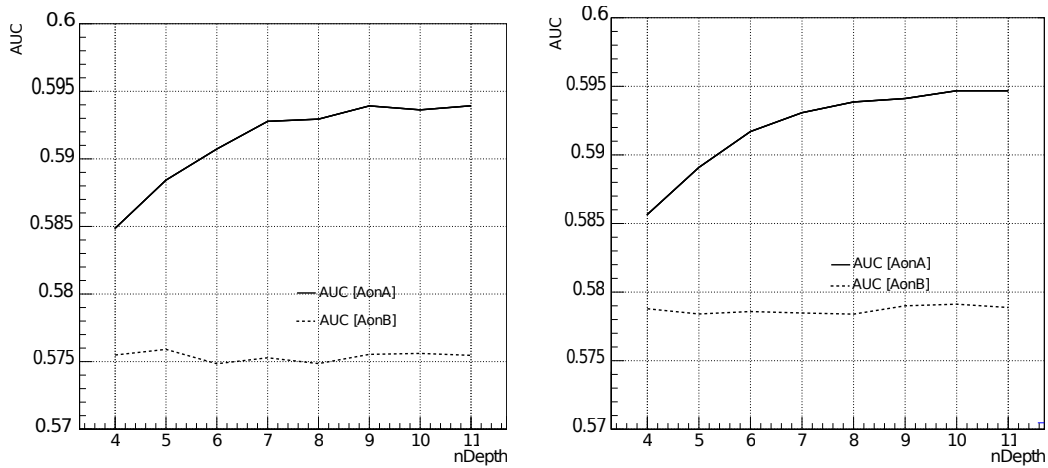


Figure 6.31: AUC for the two training setups AonA (continuous line) and AonB (dashed line) as a function of the depth of the BDT. The AUC for the AonB setup is almost constant while the one for AonA increases with higher values of the depth starting to overtrain the BDT. A value of depth equal to 5 has been chosen.

the outputs of each training and are used to derive the density ratio given in Equation 6.37. This corresponds to the BDT shape systematic of the $p-2$ reweighting procedure and has the form of a binned function. It is used to reweight the original nominal SHERPA 2.2.1 events and enters the profile likelihood fit as a nuisance parameter that acts in a correlated fashion in all analysis regions.

As explained in Section 6.7.3, it is not guaranteed that after the multi-dimensional reweighting the nominal input variable distributions, as well as the distribution of the analysis-BDT, will look exactly like the ones predicted by the alternative generator. In order to validate this procedure, the nominal distribution of each input variable (and of the analysis-BDT), weighted by the BDT density ratio, is compared to the alternative one and their agreement is evaluated through the $\chi^2/n_{\text{d.o.f}}$ metric. The closer the $\chi^2/n_{\text{d.o.f}}$ to one, the better the agreement between the two distributions.

Figure 6.32 and Figure 6.33 show the distributions of some of the variables entering the analysis-BDT for $W+bb$ events in the 1-lepton, 2-tag, 2-jet category. More plots for $W+bb$ events in the 1-lepton, 2-tag, 3-jet category can be found in Appendix B.5. The nominal distribution predicted by SHERPA 2.2.1 is shown in red, while the alternative one is shown in blue. The so-called “SHERPA 2.2.1- p_T^V -rw” distribution in green corresponds to the nominal SHERPA 2.2.1 distribution reweighted by the p_T^V density ratio, while the so-called “SHERPA 2.2.1-BDTr-rw” distribution in magenta corresponds to the nominal SHERPA 2.2.1 distribution reweighted by the BDT density ratio. In the bottom panel, the ratio is always performed with the MADGRAPH 5 distribution at the numerator.

The $\chi^2/n_{\text{d.o.f}}$ metric is first evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, then it is re-calculated by replacing the nominal SHERPA 2.2.1 distribution with the one weighted by the BDT density ratio. This gives a measure of how well the multi-dimensional procedure maps the nominal phase space to look like the alternative one. Good agreement between the SHERPA 2.2.1-BDTr-rw and the alternative MADGRAPH 5 distributions is found for all input variables. Similar agreement is found when considering $W+bc$, $W+bl$ and $W+cc$ events, as can be observed in Appendix B.5. However, 1-2% level of disagreement is found in some variables, but is of no worry as the most important variables entering the analysis-BDT, p_T^V and m_{bb} , and the analysis-BDT itself are fully covered by the BDT shape uncertainty. Moreover, the sizes of the $W+bc$, $W+bl$ and $W+cc$ components are small due to the strong b -tagging requirements against c -jets and l -jets.

In Figure 6.34 and Figure 6.35, the validation of the multi-dimensional reweighting is shown for some of the variables entering the analysis-BDT, considering $W+bb$ events in the 0-lepton, 2-tag, 2-jet category. More plots for the $W+bb$ events in the 0-lepton 2-tag, 3-jet category can be found in the Appendix B.5. The same validation test is performed also for $W+bc$, $W+bl$ and $W+cc$ events, and similar agreement is observed between the nominal distribution weighted by the BDT shape uncertainty and the alternative one.

Finally, the effect of the BDT shape uncertainty, as well as the p_T^V one, on the analysis-BDT is shown in Figure 6.36 for $W+bb$ events.

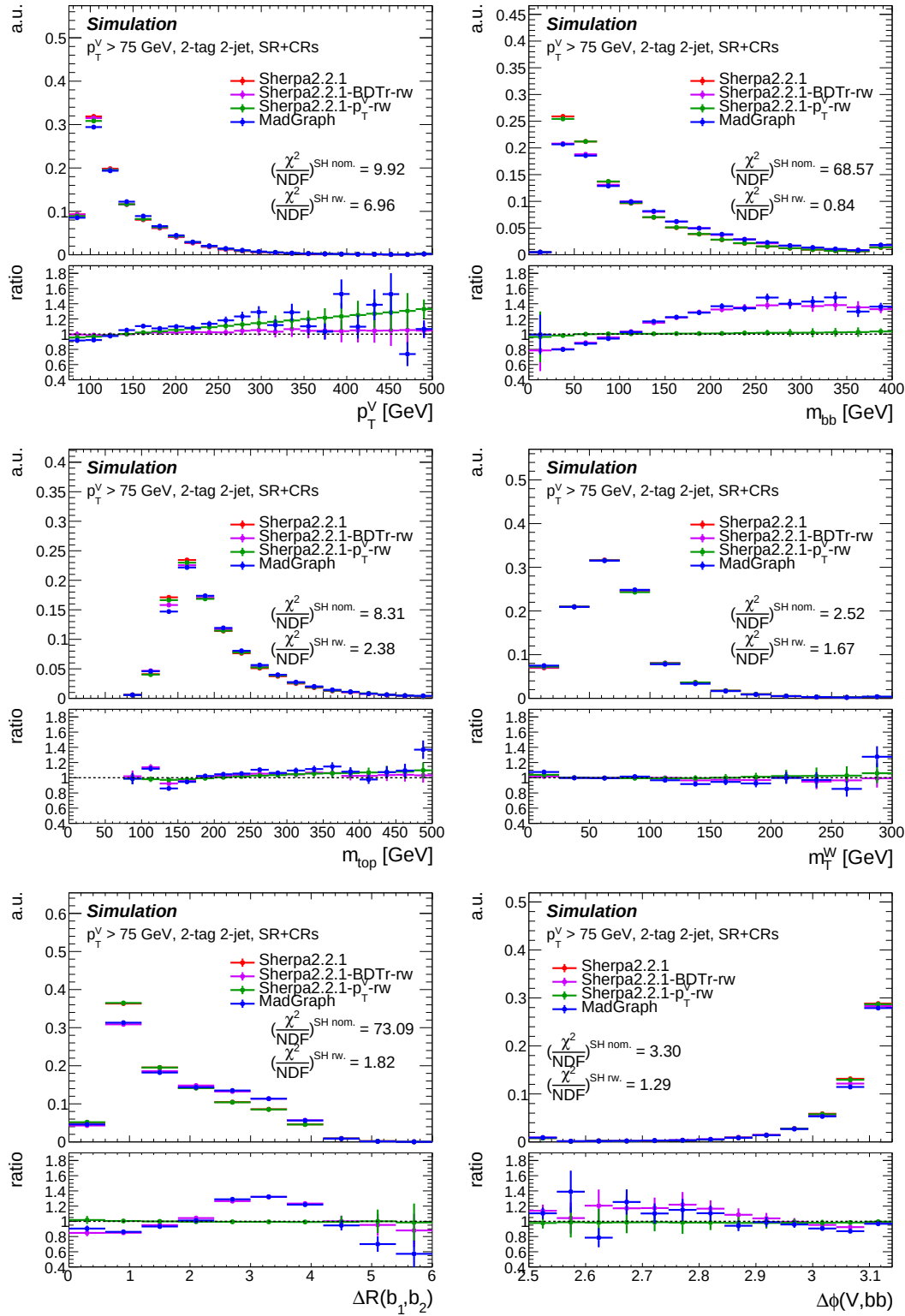


Figure 6.32: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT for $W + b\bar{b}$ events in the 1-lepton, 2-tag, 2-jet region. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

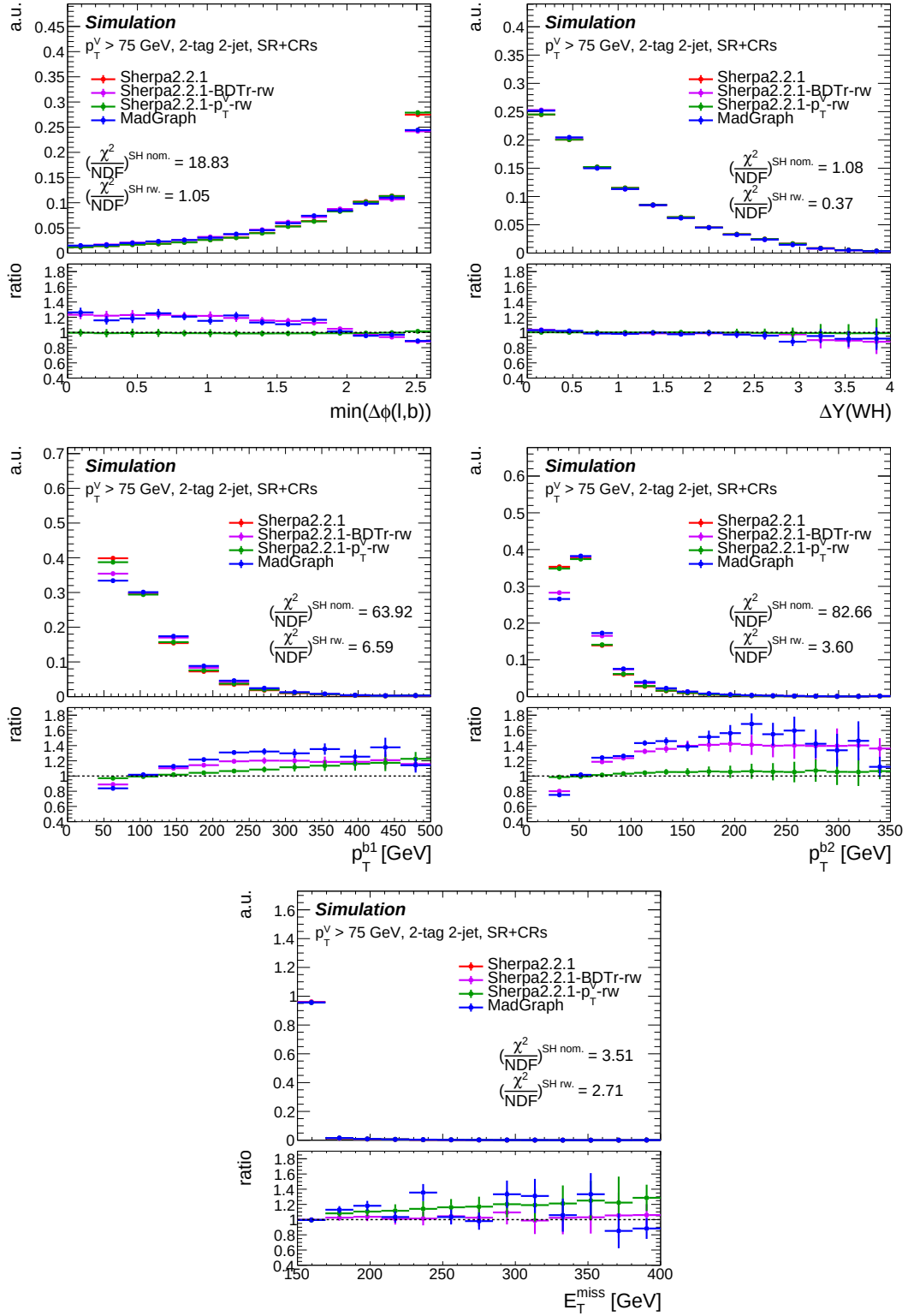


Figure 6.33: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT for $W+bb$ events in the 1-lepton, 2-tag, 2-jet region. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

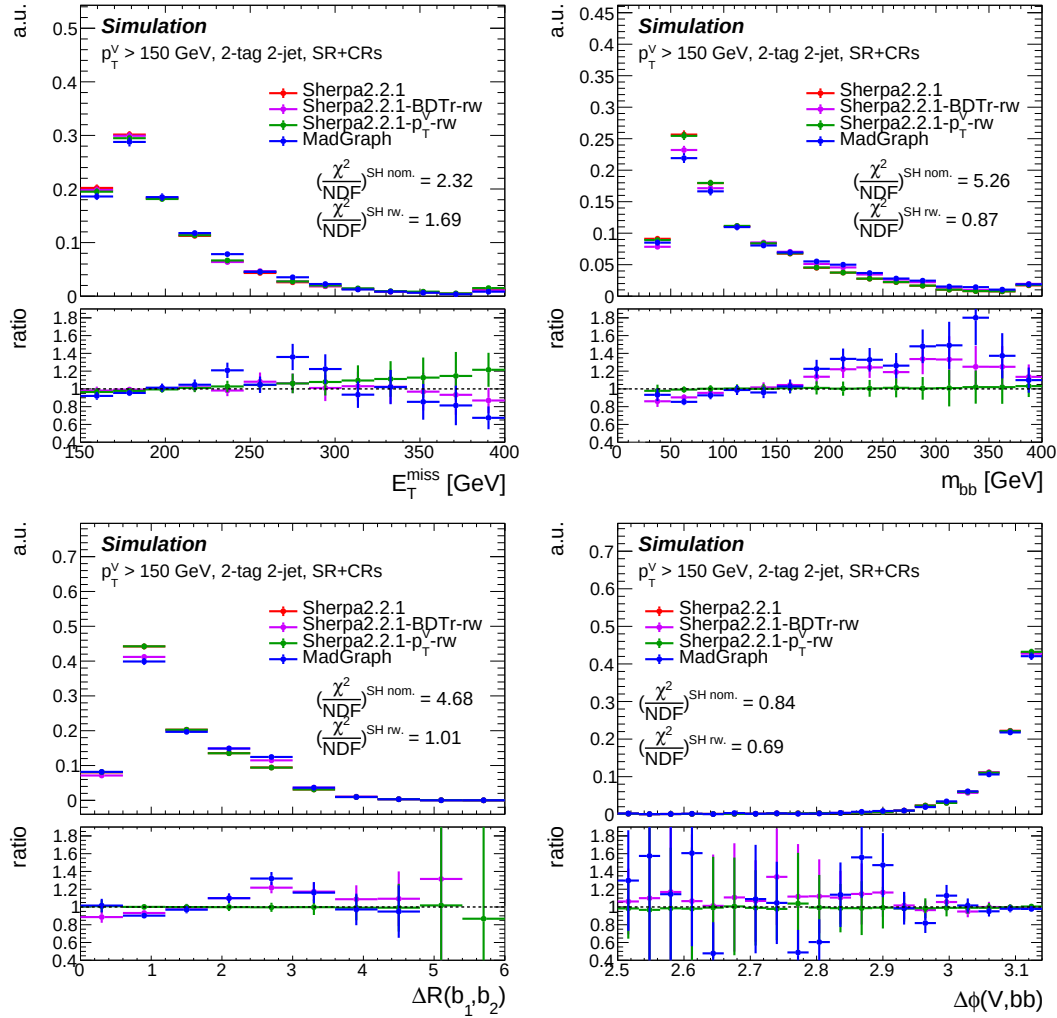


Figure 6.34: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT for $W + b\bar{b}$ events in the 0-lepton, 2-tag, 2-jet region. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

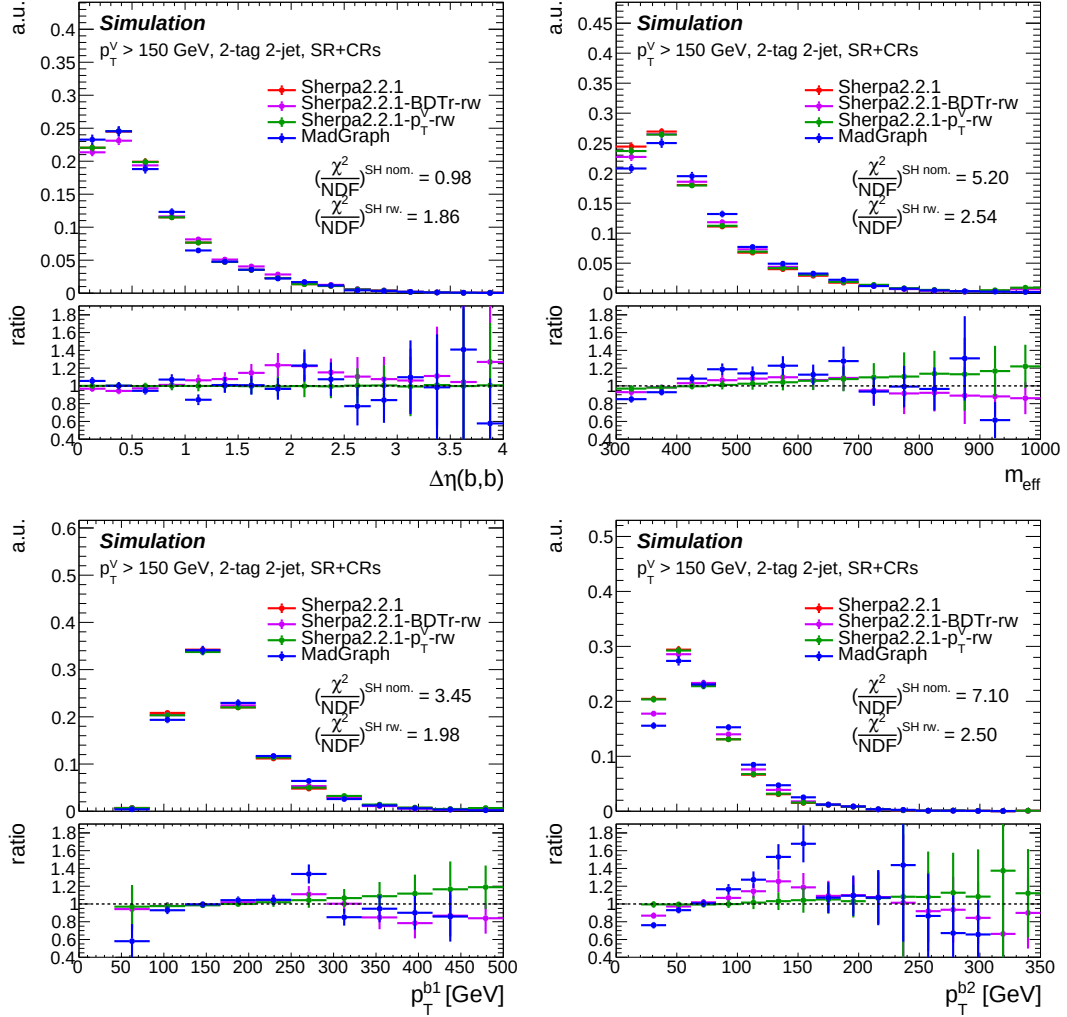


Figure 6.35: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT for $W+bb$ events in the 0-lepton, 2-tag, 2-jet region. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n.d.o.f$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

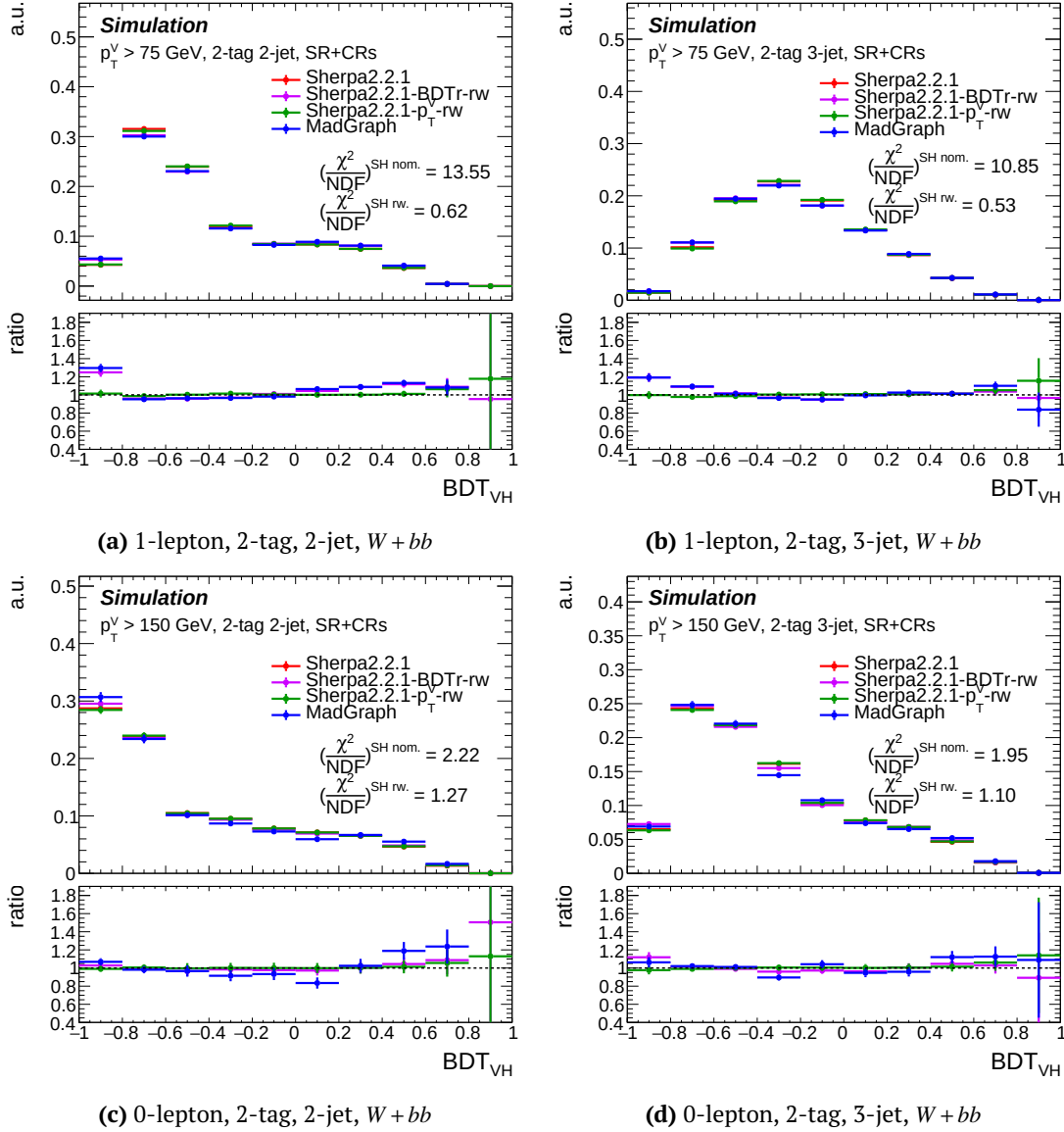


Figure 6.36: Validation of the $p-2$ reweighting procedure on the analysis-BDT considering $W + b\bar{b}$ events in the 1-lepton, 2-jet region (a), 1-lepton, 3-jet region (b), 0-lepton 2-jet region (c), and 0-lepton, 3-jet region (d). The nominal SHERPA 2.2.1 distribution is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDT r density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDT r density ratio and the alternative one.

As an additional validation test, the ROC curve was used to quantify the similarity between the nominal SHERPA 2.2.1 distribution weighted by the BDT density ratio and the unaltered alternative one. An AUC very close to 0.5 was found, representing the almost null discrimination power of the BDT.

When performing these validation tests, one needs to take in mind that the shape of each input variable can be modified in the fit by both the p_T^V and the BDT density ratio uncertainties, which are implemented as two separate nuisance parameters. Therefore, it makes more sense to evaluate the similarity between the weighted nominal distribution, obtained by combining the two shape uncertainties, and the alternative one. An example of this is given in Figure 6.37 for the m_{bb} , m_{top} , and the analysis-BDT distributions in the 1-lepton, 2-tag, 2-jet region and for $p_T^V > 150$ GeV. Despite the training of the multiclassifier being performed in the $p_T^V > 75$ GeV, the validation is shown in the $p_T^V > 150$ GeV to be consistent with the p_T^V range considered the 1-lepton channel.

It can be observed that, after being reweighted by both the p_T^V and BDT density ratio, the input variable distributions for the nominal sample are in very good agreement with those of the alternative sample. This demonstrates that the hybrid BDT-reweighting approach allows to parametrise the analysis phase space in a coherent and correlated fashion, without relying on just two kinematic variables, m_{bb} and p_T^V , but on all kinematic variables to describe the differences between the generators' predictions.

Small level of disagreement between the nominal kinematic distributions weighted by the $p - 2$ procedure and the alternative one can still be observed, an example of which is shown in Figure 6.38 in the case of the ΔR_{bb} distribution in the 1-lepton, 2-jet, SR+CRs, and $p_T^V > 150$ GeV. It can be observed that at high values of ΔR_{bb} , the weighted nominal distribution does not exactly map the alternative distribution.

To solve this disagreement, one could try to split the training in different regions, according to the p_T^V range and the jet category, to help the multiclassifier learning the generators' differences in all corners of the phase space. On the other hand, one could try to use more powerful mapping algorithms than the Boosted Decision Tree, as a neural network, for example.

Studies are still on-going to improve the BDT-reweighting technique with the aim of exploiting it also to model other major or minor backgrounds of the analysis, like Z +jets or single-top and diboson.

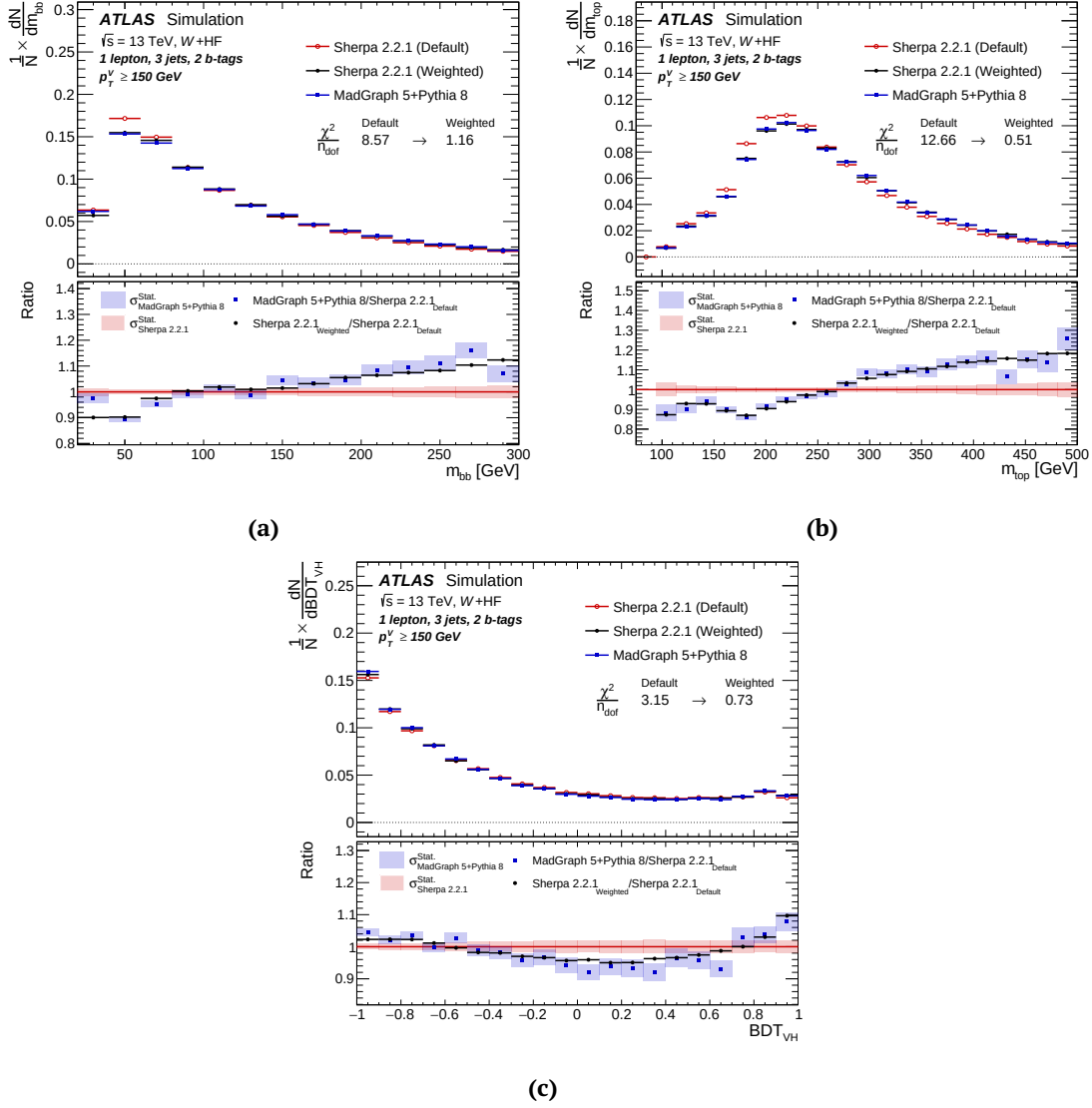


Figure 6.37: The nominal (red points), alternative (blue points) and reweighted nominal (black points) distributions of the m_{bb} (a), m_{top} (b), and the analysis-BDT (c) are shown to illustrate the validation of the $p-2$ reweighting procedure considering all $W+hf$ events in the 1-lepton, 2-tag, 2-jet region and for $p_T^V > 150$ GeV. The bottom panel shows the ratio between the alternative (blue points) and reweighted nominal (black points) distributions to the original nominal prediction, with the shaded regions representing the statistical uncertainty on the sample. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDT density ratio and the alternative one [187].

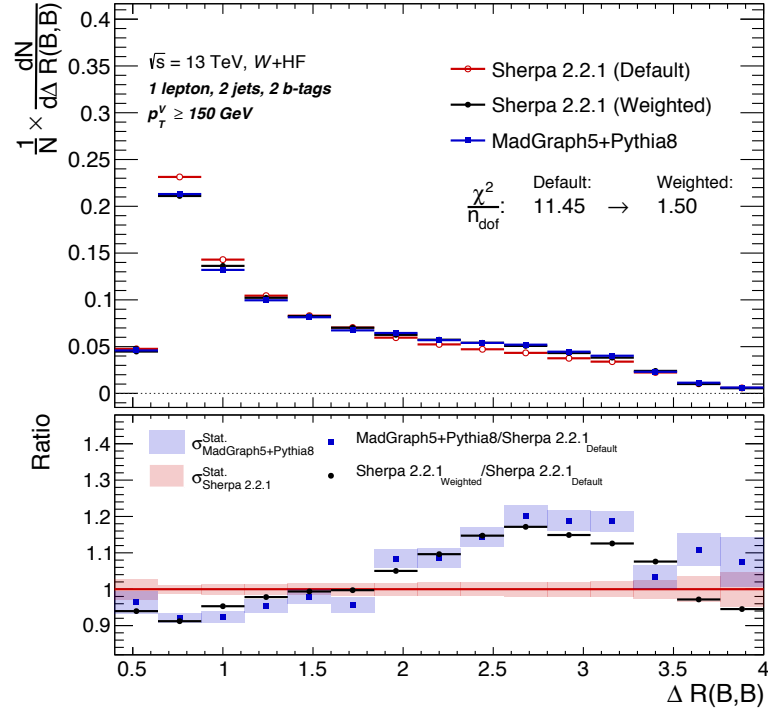


Figure 6.38: The nominal (red points), alternative (blue points) and reweighted nominal (black points) distributions of the ΔR_{bb} are shown to illustrate the validation of the $p-2$ reweighting procedure considering all W +hf events in the 1-lepton, 2-tag, 2-jet region and for $p_T^V > 150 \text{ GeV}$. The bottom panel shows the ratio between the alternative (blue points) and reweighted nominal (black points) distributions to the original nominal prediction, with the shaded regions representing the statistical uncertainty on the sample. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDT density ratio and the alternative one. Small level of disagreement between the reweighted nominal distribution and the alternative one is observed, especially at low and high values of ΔR_{bb} .

6.10 Modelling of other backgrounds

The modelling scheme for the remaining backgrounds of the $VH, H \rightarrow b\bar{b}$ analysis is explained in this Section. Tables 6.17 and 6.18 show a summary of the systematic uncertainties affecting the modelling of all background samples of this analysis and their implementation in the final fit scheme.

Z+jets background estimation The overall normalisation of the Z +hf background is left floating in the 2-jet and 3-jet categories and in the $75 < p_T^V < 150$ GeV and $p_T^V > 150$ GeV regions. The largest constraint comes from the High- ΔR CR and the SR in the 2-lepton channel, which are dominated by Z +jets events. For this reason, a relative acceptance uncertainty on the ratio of the event yields in the 0-lepton channel to that in the 2-lepton channel is considered. This is applied only to the 0-lepton channel across all regions; the Z +hf contribution in the 1-lepton channel is of the order of 1% in all analysis regions.

Uncertainties on the Z +hf flavour composition are derived as ratios of the yield of the smaller components bb, bc, bl, cc with respect to the main one bb . As in the case of the W +jets background, the contribution of the $Z + cl$ and $Z + l$ backgrounds is strongly suppressed by the b -tagging requirement and compose less than 1% of the total Z +jets background, therefore only a single normalisation uncertainty is considered.

Shape uncertainties are derived using a 1-dimensional approach, outlined in Section 6.7, on the m_{bb} and the p_T^V distributions in the 2-lepton channel, for $p_T^V > 75$ GeV, the sum of the SR, Low- ΔR CR and High- ΔR CR, and in the ≥ 2 -jets category. A data-driven method is employed and the region around the VH and VZ signals, $80 < m_{bb} < 140$ GeV, is removed from both the p_T^V and the m_{bb} distributions in order to veto the signal events when deriving the shape systematics. This phase space is enriched in Z +jets events but contains a non-negligible $t\bar{t}$ contamination and some smaller contributions from W +jets and diboson. Their expected contributions are subtracted from the data and the ratio between the data and the Z +jets MC prediction is fitted with a linear function to derive a shape systematic. The resulting functional forms are reported in Figure 6.39 for p_T^V and m_{bb} in the 2-lepton, 2-tag, ≥ 2 -jets region and inclusively in p_T^V .

The shape systematics on the m_{bb} and p_T^V distributions, derived in the 2-lepton channel, are also applied to the 0-lepton channel in a correlated fashion. They are implemented in the fit as nuisance parameters affecting the shape of the analysis-BDT discriminant.

Additional acceptance uncertainties, with priors calculated according to Equation 6.34, are assigned to the Low- ΔR CR and the High- ΔR CR to control the relative event migration to the SR. Decorrelating the yield migration from the shape effect induced by the uncertainty on m_{bb} has proven to give the fit more degrees of freedom to adjust the data-MC agreement. See Section 6.13 for a more detailed explanation.

The p_T^V shape uncertainty, on the other hand, is implemented as both a shape and a normalisation systematic that allows event migrations across the p_T^V bins.

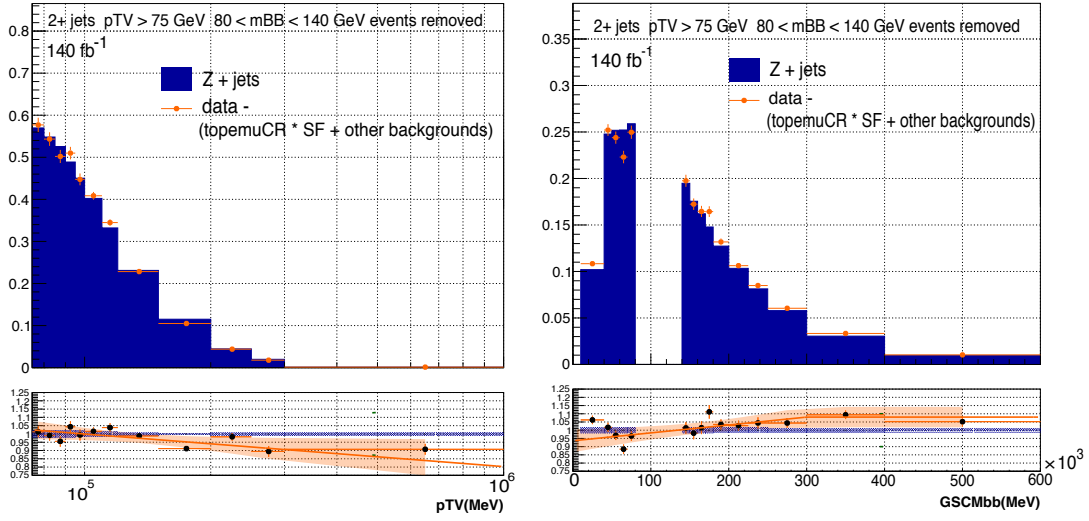


Figure 6.39: The p_T^Z and the m_{bb} distributions of the Z +jets data (orange dots) and MC template (filled in blue) in the 2-lepton, 2-tag, ≥ 2 jets category and for $p_T^V > 75$ GeV. The black dots with orange errors bars in the bottom panel show the ratio between the data and the MC template. The shape uncertainty is derived as a fit to the data/MC ratio and it is indicated by the continuous orange line with the shaded region being the 95% confidence level. (Courtesy of Thomas Paul Charman, Queen Mary University of London).

$t\bar{t}$ background estimation In the 0-lepton and 1-lepton channels the $t\bar{t}$ events present at least one non-reconstructed object, while the $t\bar{t}$ events in the 2-lepton channel are mostly fully reconstructed. Also, the $t\bar{t}$ events with two b -tagged jets in the 0-/1-lepton channel contain a non-negligible bc component. For this reason, two methodologies are adopted to estimate the $t\bar{t}$ background according to the channel: MC comparisons are used in the 0-lepton and 1-lepton channels, while a data-driven approach is utilized in the 2-lepton channel where a pure $e\mu - CR$ is built, as explained in Section 6.4.

In the 0-/1-lepton channels, two floating normalisations are defined separately in the 2-jet and 3-jet regions. Acceptance uncertainties, calculated as double ratios (see Equation 6.34), are defined to control the yield migration from the 2+3-jet region in the 1-lepton channel to the 2+3-jet region in the 0-lepton channel. The migration effects across p_T^V bins and across the control and signal regions are controlled by the p_T^V and the BDT shape uncertainties, respectively, which are implemented so that they cover both the shape and the normalisation variations on the final BDT discriminant. More details on the specific correlation scheme chosen for these two systematics are given in Section 6.13.

Given the presence of non- bb flavour components in the 0-/1-lepton channels, the $t\bar{t}$ processes are divided into three flavour categories, bb , bc and *others*, where *others* includes all the smaller components bl , cc , cl , ll . Flavour composition uncertainties are calculated by comparing the difference in the ratio of the bc or *other* component to the dominant bb component predicted by the nominal generator POWHEG + PYTHIA8 to the one predicted by the alternative POWHEG + HERWIG7 for the parton shower, MADGRAPH5_aMC@NLO + PYTHIA8.2 for the matrix element, and POWHEG + PYTHIA8 with modified values of the

renormalization and factorization scales.

Shape uncertainties are derived using the p_T^V -factorized BDTr approach. In each lepton channel and jet category, and inclusively in flavour composition, the p_T^V distribution of the nominal generator POWHEG + PYTHIA8 is compared to the alternative one predicted by POWHEG + HERWIG7 and by MADGRAPH5_aMC@NLO + PYTHIA8.2, in the $p_T^V > 75$ GeV range. Only the largest variation between the two in each lepton channel and jet region is taken as p_T^V shape uncertainty and used to reweight the nominal MC events.

Then, six different trainings are performed separately in the 0-lepton and 1-lepton channels, depending on the jet category (2-jet and 3-jet), and on the flavour composition of the event bb , bc , *others*. The nominal POWHEG + PYTHIA8 weighted by the p_T^V density ratio in each of the six regions is trained against the unaltered alternative predictions to evaluate two BDTr shape systematics, one for the variation of the matrix element and one for the parton shower.

In the 2-lepton channel, an $e\mu$ pair defines a 99% pure $t\bar{t}$ and single-top Wt control region. This is defined by using the nominal event selection but replacing the same-flavour lepton selection with a requirement of exactly one electron and one muon.

Due to the leptonic decays of the W bosons, the two top backgrounds in the 2-lepton channel are flavour symmetric, meaning that the events in the SR and the $e\mu$ -CR are kinematically the same. For this reason, the events in the $e\mu$ -CR are used as a template for the top backgrounds within the SR¹².

This data-driven procedure allows to eliminate the experimental uncertainties related to the detector, object reconstruction and b -tagging, as well as the dependency from the MC theoretical predictions. The dominant source of uncertainty in the data-driven estimation is given by the statistical uncertainty of the data sample.

Single-top background estimation The single-top background has a negligible contribution in the 0-lepton and 2-lepton channels, and all the uncertainties derived in the 1-lepton channel are applied also to the 0-/2-lepton channels in a correlated fashion. An overall normalisation uncertainty is defined for each single-top process (s -/ t -/ Wt -channels) and accounts for the renormalization and factorization scale variations, α_S uncertainty, and PDF variations. The sum in quadrature of these uncertainties give an overall uncertainty of 4.4%, 4.6% and 6.2% for the t -, s -, and Wt -channels, respectively.

The s -channel has a vanishing contribution in all analysis regions of the 1-lepton channel, therefore, systematic uncertainties are derived only for the t -channel and the Wt -channel.

Acceptance uncertainties are considered to account for event migrations accross the jet categories; the prior is calculated as the sum in quadrature of all variations, each of which is obtained as a double ratio of the nominal MC prediction and the alternative one.

¹²Any bias caused by lepton trigger, reconstruction, identification or acceptance is determined by comparing the ratio of simulated top background events in the SR to those in the $e\mu$ -CR. No significant bias on the shape or the normalisation is observed for any of the important kinematic variables, including the analysis-BDT discriminant.

Shape systematics are derived on the m_{bb} and the p_T^V distributions following the 1-dimensional procedure outlined in Section 6.7. In the Wt -channel the contribution to the modelling uncertainties was found to be greatly dependent on the process. In events in which the two b -tagged jets are genuine b -jets, the comparison between the *diagram removal* and *diagram subtraction* schemes is the dominant source of uncertainty. In events where only one or fewer of the b -tagged jets are actually b -jets, the dominant source of uncertainty is given by different parton shower or matrix element schemes. For the t -channel process, the main source of uncertainty is given by the choice of the parton shower scheme for the p_T^V shape and the increased or decreased radiation for the m_{bb} shape.

Diboson background estimation An overall normalisation uncertainty is defined accounting for the effect of the renormalization and factorization scales and the difference between the nominal prediction by SHERPA 2.2.1 and the alternative predictions given by POWHEG + PYTHIA8 and POWHEG + HERWIG+. These uncertainties are 25%, 26% and 20% for WW , WZ and ZZ , respectively. The WW background composes less than 0.1% of the total diboson background in all channels, so a single normalisation uncertainty is considered. Acceptance uncertainties accounting for scale and PS/UE variations are considered to control the event migration across the jet categories and from the 0-lepton to the 2-lepton channel in the case of ZZ events, and from the 0-lepton to the 1-lepton channel in the case of WZ events.

Shape uncertainties for the scale and PS/UE variations are derived on the m_{bb} and p_T^V distributions. The matrix element shape uncertainty is only considered for the m_{bb} in the WZ and ZZ processes, as it was found to have a negligible impact on the p_T^V shape.

Multi-jet data-driven estimation The multi-jet background was demonstrated to be negligible in the 0-lepton and 2-lepton channels, using the same techniques as detailed in Ref. [13]. In the 1-lepton channel, tighter selection criteria, based on calorimeter isolation for electrons and track isolation for muons, are applied to reduce the multi-jet contamination. Furthermore, a cut on $E_T^{\text{miss}} > 30$ GeV is applied in all p_T^V regions, while an additional cut on $m_T^W > 20$ GeV is applied in the medium $75 < p_T^V < 150$ GeV region, which has a higher multi-jet contamination compared to the higher p_T^V regions. To increase the number of events, the higher p_T^V regions are merged together into a single $p_T^V > 150$ GeV region. The residual contamination of the multi-jet background in the 1-lepton channel is extracted through a template fit in the 2-jet and 3-jet categories and separately for the electrons and muons as the sources of fake electrons and fake muons are different.

A multi-jet enriched control region, called **MJ-CR**, is defined by applying the nominal event selection, but inverting the tight lepton isolation requirement with the loose isolation cut still applied and subtracting all the non-multi-jet ($t\bar{t}$, single-top, W +jets, Z +jets and diboson) simulated backgrounds from the data. The transverse mass m_T^W of the W -candidate is chosen to derive the nominal multi-jet template fits as it has the largest discrimination power between multi-jet and non-multi-jet backgrounds. Furthermore, a

single b -tagged jet is required in order to increase the number of events in the MJ-CR.

A fit to the m_T^W distribution, is performed in each jet category and p_T^V region, separately for the e/μ channel, and in the full Low- ΔR CR, SR, and High- ΔR CR phase space, in order to extract the normalisation factors for the multi-jet background and for all the other simulated backgrounds. Figure 6.40 shows the post-fit distribution of m_T^W in the 2-tag, 2-jet signal region, and for $p_T^V > 150$ GeV, for the electron and muon sub-channels, as an example of the performance of the procedure. Good agreement is observed between data and MC prediction.

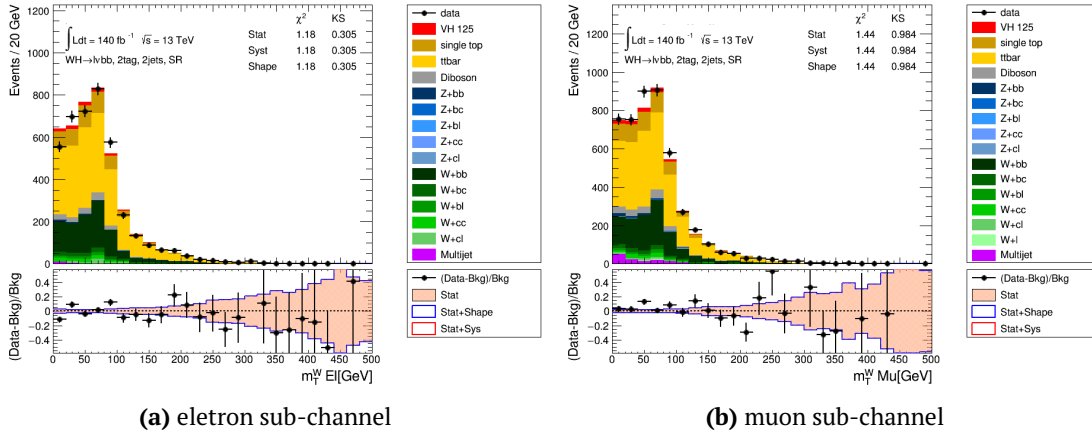


Figure 6.40: The m_T^W distribution in the $p_T^V > 150$ GeV signal region, 2-tag, 2-jet region in the e sub-channel (a) and μ sub-channel (b). The bottom panel shows the good agreement between data and MC prediction. (Courtesy of Konie Al Khoury, Université Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay).

Two uncertainty components are considered, those altering the normalisation and those affecting the shape of the analysis-BDT multi-jet template. They have the same sources of uncertainties which are derived by changing the definition of the multi-jet control region (more stringent isolation requirements or a different single-electron-trigger to probe potential trigger biases) but also varying the normalisation of the contamination from the $t\bar{t}$, single-top Wt processes and V +jets processes. Furthermore, the usage of another discriminant variable instead of m_T^W and the inclusion of a cut on $E_T^{\text{miss}} < 30$ GeV (which enhances the multi-jet contamination) are considered as sources of normalisation effects on the multi-jet templates. To evaluate the final normalisation uncertainty, all individual contributions are summed in quadrature.

Table 6.17: Summary of the systematic uncertainties for the modelling of the Z +jets, W +jets, $t\bar{t}$, single top and multi-jet backgrounds. ‘ME’ indicates a matrix element generator variation and ‘PS’ indicates a parton shower generator variation. The ‘M+S’ symbol is used when a shape uncertainty includes an event migration effect that allows the relative acceptance to change between regions, whilst ‘S’ indicates that the uncertainty only acts on the shape of the final BDT discriminant in the signal region. If an uncertainty is considered independently in different regions, these are detailed in parentheses. Where the size of an acceptance systematic uncertainty varies between regions, a range is displayed [146].

Z + jets	
Z + ll normalisation	18%
Z + cl normalisation	23%
Z + HF normalisation	Floating (2-jet, 3-jet) $\times$ (75 GeV < p_T^V < 150 GeV, p_T^V > 150 GeV)
Z + bc -to- Z + bb ratio	30 – 40%
Z + cc -to- Z + bb ratio	13 – 16%
Z + bl -to- Z + bb ratio	20 – 28%
SR-to-low ΔR CR ratio	3.8 – 9.9% (75 GeV < p_T^V < 150 GeV, p_T^V > 150 GeV)
SR-to-high ΔR CR	2.7 – 4.1% (75 GeV < p_T^V < 150 GeV, p_T^V > 150 GeV)
0-to-2 lepton ratio	7%
p_T^V	M+S (75 GeV < p_T^V < 150 GeV, p_T^V > 150 GeV)
m_{bb}	S (75 GeV < p_T^V < 150 GeV, p_T^V > 150 GeV)
W + jets	
W + ll normalisation	32%
W + cl normalisation	37%
W + HF normalisation	Floating (2-jet, 3-jet)
W + bc -to- W + bb ratio	15% (0-lepton) and 30% (1-lepton)
W + cc -to- W + bb ratio	10% (0-lepton) and 30% (1-lepton)
W + bl -to- W + bb ratio	26% (0-lepton) and 23% (1-lepton)
SR-to-CR ratio	3.6-15%
0-to-1 lepton ratio	5%
p_T^V	M+S (2-jet, 3-jet)
BDT _R	S
$t\bar{t}$ (0+1-lepton channels only)	
$t\bar{t}$ normalisation	Floating (2-jet, 3-jet)
0-to-1 lepton ratio	8%
$t\bar{t}$ (flavour composition) bc -to- bb ratio (ME)	7.6 – 8.2% (0-lepton), 1.3 – 3.8% (1-lepton)
$t\bar{t}$ (flavour composition) bc -to- bb ratio (PS)	2.1 – 3.2% (0-lepton), 1.5 – 7.1% (1-lepton)
$t\bar{t}$ (flavour composition) other-to- bb ratio (ME)	2.8 – 6.4% (0-lepton), 3.3 – 5.7% (1-lepton)
$t\bar{t}$ (flavour composition) other-to- bb ratio (PS)	5.6 – 13% (0-lepton), 0.3 – 2.1% (1-lepton)
p_T^V	M+S (2-jet, 3-jet)
BDT _R ME variation	M+S (2-jet, 3-jet)
BDT _R PS variation	M+S (0-lepton, 1-lepton)
Single top-quark	
Cross-section	4.6% (s -channel), 4.4% (t -channel), 6.2% (Wt)
Acceptance 2-jet	17% (t -channel), 55% ($Wt(bb)$), 24% ($Wt(\text{other})$)
Acceptance 3-jet	20% (t -channel), 51% ($Wt(bb)$), 21% ($Wt(\text{other})$)
m_{bb}	M+S (t -channel, $Wt(bb)$, $Wt(\text{other})$)
p_T^V	M+S (t -channel, $Wt(bb)$, $Wt(\text{other})$)
Multi-jet (1-lepton)	
Normalisation	30 – 200% (2-jet), 100% (3-jet)
BDT template	M+S

Table 6.18: Summary of the systematic uncertainties in the background modelling for the diboson production. ‘PS/UE’ indicates parton shower/underlying event. The ‘M+S’ symbol is used when a shape uncertainty includes a migration effect that allows relative acceptance changes between regions. If an uncertainty is considered independently in different regions, these are detailed in parentheses. Where the size of an acceptance systematic uncertainty varies between regions, a range is displayed [146].

ZZ	
Normalisation	20%
0-to-2 lepton ratio	6%
Acceptance from scale variations	10 – 18%
Acceptance from PS/UE variations for 2 or more jets	6%
Acceptance from PS/UE variations for 3 jets	7% (0-lepton), 3% (2-lepton)
m_{bb} from scale variations	M+S (correlated with WZ uncertainties)
p_T^V from scale variations	M+S (correlated with WZ uncertainties)
m_{bb} from PS/UE variations	M+S (correlated with WZ uncertainties)
p_T^V from PS/UE variations	M+S (correlated with WZ uncertainties)
m_{bb} from matrix-element variations	M+S (correlated with WZ uncertainties)
WZ	
Normalisation	26%
0-to-1 lepton ratio	11%
Acceptance from scale variations	13 – 21%
Acceptance from PS/UE variations for 2 or more jets	4%
Acceptance from PS/UE variations for 3 jets	11%
m_{bb} from scale variations	M+S (correlated with ZZ uncertainties)
p_T^V from scale variations	M+S (correlated with ZZ uncertainties)
m_{bb} from PS/UE variations	M+S (correlated with ZZ uncertainties)
p_T^V from PS/UE variations	M+S (correlated with ZZ uncertainties)
m_{bb} from matrix-element variations	M+S (correlated with ZZ uncertainties)
WW	
Normalisation	25%

6.11 Experimental Systematic Uncertainties

Several sources of experimental systematic uncertainties are considered in the $VH, H \rightarrow b\bar{b}$ analysis, as outlined in this Section.

Luminosity The uncertainty on the integrated luminosity is 2.1%, 2.6%, 2.4% and 2.0% for the data collection periods of 2015, 2016, 2017 and 2018, respectively. The uncertainty on the combined integrated luminosity for the full Run 2 data collection, 2015-2018, is 1.7%. It is derived following a methodology outlined in Ref. [206], and using the LUCID-2 detector for the baseline luminosity measurements [207].

Pile-up reweighting The average number of bunch crossing in the MC simulation is rescaled by 3% to match the data and an uncertainty, large as the correction, is included [208].

Leptons Uncertainties are considered on the lepton trigger, reconstruction and identification efficiencies, along with the uncertainty on their energy scale and resolution following the recommendations in Ref. [209] for electrons and in Ref. [108] for muons.

E_T^{miss} An uncertainty is assigned to the E_T^{miss} trigger correction factors which accounts for the statistical uncertainty in the measured correction factors when using W +jets, Z +jets or $t\bar{t}$ events. Furthermore, the uncertainties on the energy scale and resolution of the jets and leptons are propagated to the calculation of the E_T^{miss} , together with further uncertainties on the scale, resolution and reconstruction efficiency of the tracks used to compute the soft E_T^{miss} term [119].

Jets The main sources of uncertainties come from the jet energy scale (JES) decomposed into 30 independent components and the jet energy resolution (JER) with 8 independent components, following the recommendations in Refs. [210, 211].

b -tagging The dominant experimental uncertainty in this analysis comes from the b -tagging. Data-MC scale factors are derived in the five MV2c10 bins of the pseudo-continuous b -tagging (see Section 4.4) and applied in order to correct for mismodelling in the output of the b -tagging algorithm. These are derived separately for b -jets [137], c -jets [138] and l -jets [140] and uncertainties on them are estimated from multiple measurements and decomposed into 29, 18 and 10 uncorrelated components, respectively. An additional systematic is considered, related to the extrapolation in the case of high- p_T jets.

6.12 Smoothing and pruning of systematics

Limited number of events in the MC samples or small variations can lead to statistical fluctuations in the systematic uncertainties. Therefore, a “**smoothing**” algorithm is used to mitigate these effects. The relative effect of the systematic uncertainty on the analysis-BDT is iteratively “re-binned” from one extremum to the next until no local

extrema remain in the template. At each step, the number of bins that need to be merged is chosen as the one for which the difference between the merged and unmerged template is the smallest. The bins resulting from this merging iteration are sequentially merged, starting from the upper end of the distribution, until the statistical uncertainty in each of the merged bins of the nominal template is smaller than 5%.

A systematic uncertainty can have a negligible effect on the BDT discriminant. Therefore, a “**pruning**” procedure is applied for every sample and in each analysis region:

1. a normalisation uncertainty is neglected if it induces a variation smaller than 0.5% with respect to the nominal template of the BDT discriminant, or if its up and down variations have the same sign;
2. a shape uncertainty is neglected if it does not induce a variation greater than 0.5% with respect to the nominal template of the BDT discriminant in at least one bin after the overall normalisation is removed, or if only the up or down variation is non-zero and passes the previous steps;

Additional pruning criteria are applied to the background samples contributing to less than 1% of the total background:

1. in regions where at least one bin of the BDT distribution has a signal contribution $> 2\%$ of the total background, both shape and normalisation uncertainties on the sample are pruned if they generate a variation smaller than 2% of the signal yield;
2. in regions where no bin of the BDT distribution has a signal contribution $> 2\%$, normalisation uncertainties inducing a variation < 0.005 of the total background yield, and shape uncertainties inducing a variation in each bin smaller than 0.005 of the total background are pruned.

6.13 Fit model cross-checks

As explained in Section 6.6, the statistical interpretation of the results leads to the extraction of the observed significance which can indicate the discovery (or not) of the process. It is, therefore, mandatory that the analyser remains as unbiased as possible when designing the analysis strategy and the fit model. For this reason, a rigorous *blinding* procedure is applied, stating that it is not allowed to look at the data/MC agreement in the signal regions¹³ until one is confident with the full analysis chain.

To determine the stability of the background modelling and the accuracy of the treatment of the systematic uncertainties within the fit model, one can look at the data/MC agreement in the CRs and the sidebands of the SR, as well as to the impact of each nuisance parameter on the total uncertainty on the signal strength (with the ranking

¹³The signal regions enter in the fit model, however, the data points in these regions, as well as their agreement with the MC prediction, are hidden so that no bias is introduced in the analysis chain to correct, for example, for a data/MC mismodelling.

and breakdown methods). Once a fit model has been agreed on, one can proceed with the unblinding of the signal region.

This section summarises the studies that led to a particular treatment of some of the analysis nuisance parameters within the fit model, before the final unblinding of the signal region. Particular attention will be given to the $t\bar{t}$ shape systematics and to the extrapolation uncertainties across signal and control regions for the W +jets backgrounds. In fact, these represent the main contributions of the author in the context of the fit cross-checks.

Treatment of the parton shower BDTr shape systematic for the $t\bar{t}$ background A -1σ pull on the NP associated to the BDTr shape systematic (SysBDTr_ttbar_PS), which models the parton shower of the $t\bar{t}$ background, was observed in the 0-lepton channel and in the 1-lepton channel.

To understand the origin of the pull in the 1-lepton channel, a conditional fit was performed by decorrelating SysBDTr_ttbar_PS in all the jet and p_T^V categories and in the signal and control regions and comparing the resulting pulls with the one obtained when treating the NP as correlated everywhere. The values of the pulls obtained from these two fits are reported in Table 6.19 from which it can be observed that the pull on SysBDTr_ttbar_PS in the correlated case is coming exclusively from the 3-jet, $150 < p_T^V < 250$ GeV signal region (highlighted in red).

Table 6.19: Pulls of the NP associated to the parton shower BDTr shape systematic on the $t\bar{t}$ background when this is treated as correlated or uncorrelated accross all analysis regions, jet and p_T^V categories. It can be observed that the -1σ pull of the NP in the correlated case is driven by a pull in the 3-jet, $150 < p_T^V < 250$ GeV signal region.

Analysis region	Pull
All correlated	-0.86 ± 0.74
2-jet, High- ΔR CR , $p_T^V > 250$ GeV	$+0.15 \pm 0.99$
3-jet, High- ΔR CR , $150 < p_T^V < 250$ GeV	-0.15 ± 0.97
2-jet, Low- ΔR CR , $150 < p_T^V < 250$ GeV	-0.02 ± 0.99
2-jet, Low- ΔR CR , $p_T^V > 250$ GeV	-0.01 ± 0.99
3-jet, Low- ΔR CR , $150 < p_T^V < 250$ GeV	$+0.02 \pm 0.98$
3-jet, Low- ΔR CR , $p_T^V > 250$ GeV	-0.02 ± 0.99
2-jet, SR, $150 < p_T^V < 250$ GeV	-0.26 ± 0.91
2-jet, SR, $p_T^V > 250$ GeV	$+0.20 \pm 0.99$
3-jet, SR, $150 < p_T^V < 250$ GeV	-0.90 ± 0.79
3-jet, SR, $p_T^V > 250$ GeV	-0.07 ± 0.94

The propagation of this pull to the other analysis regions was shown to worsen the data-MC agreement especially in the 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR . This effect

was particularly evident in the ΔR_{bb} variable, therefore, the following studies focus on this variable.

Figure 6.41 (a) shows the post-fit distribution of ΔR_{bb} in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR obtained from a conditional fit in which the NP SysBDTr_ttbar_PS is treated as correlated in all analysis regions. A significant data-MC disagreement can be observed in the bottom panel at high values of ΔR_{bb} .

Figure 6.41 (b) shows the post-fit distribution in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR obtained when decorrelating SysBDTr_ttbar_PS in all analysis regions. A better data-MC agreement can be observed at high values of ΔR_{bb} , meaning that the propagation of the pull from the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, signal region to the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR is causing the data-MC mismodelling. This is further confirmed when looking at the change in the data-MC ratio between (a) and (b); if the NP is treated as correlated in all analysis regions, the -1σ pull in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, signal region is propagated to the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR, and the $t\bar{t}$ MC template gets modified by the down variation of the NP shown in blue in (c). The fact that a pull on SysBDTr_ttbar_PS, mainly coming from a single region, gets propagated to the other analysis regions when the NP is treated as correlated is a sign that some additional systematic uncertainty is missing, that would create the right degree of freedom in the region where the pull is, thus recovering the observed data-MC mismodelling.

Due to the lack of time, no further studies were performed on this issue and it was decided to simply decorrelate the SysBDTr_ttbar_PS in the 3-jet, $150 < p_T^V < 250$ GeV signal region of the 1-lepton channel from the rest of the 1-lepton analysis regions.

To study the origin of the pull on the SysBDTr_ttbar_PS nuisance parameter in the 0-lepton channel, a conditional fit was performed in all signal and control regions with the distribution of the BDT_{VH} discriminant in the signal regions reduced to a single bin in order to allow just the normalisation of the signal and backgrounds MC prediction to vary. In Table 6.20, the pull obtained in the default fit allowing both the normalisation and the shape of the MC predictions to change is compared to the one obtained in the fit with all signal regions having a single bin. The two pulls are observed to be compatible, meaning that the pull is originating from a tension in the normalisation of the MC predictions across the different analysis regions.

Since the origin of the pull was found to be different in the 0-lepton and 1-lepton channels, it was decided to treat the SysBDTr_ttbar_PS NP as uncorrelated in the two channels.

Treatment of the matrix element BDT and p_T^V shape systematics for the $t\bar{t}$ background A simultaneous fit in the three lepton channels was performed decorrelating the p_T^V and BDT shape systematics of the $t\bar{t}$ background in the jet categories. Table 6.21 shows the pulls of the NPs associated to the above systematics, SysBDTr_ttbar_ME for the BDT shape uncertainty modelling the matrix element and SysTTbarPtV_BDT describing the differences in the p_T^V shape of the $t\bar{t}$ template. A tension is observed in the

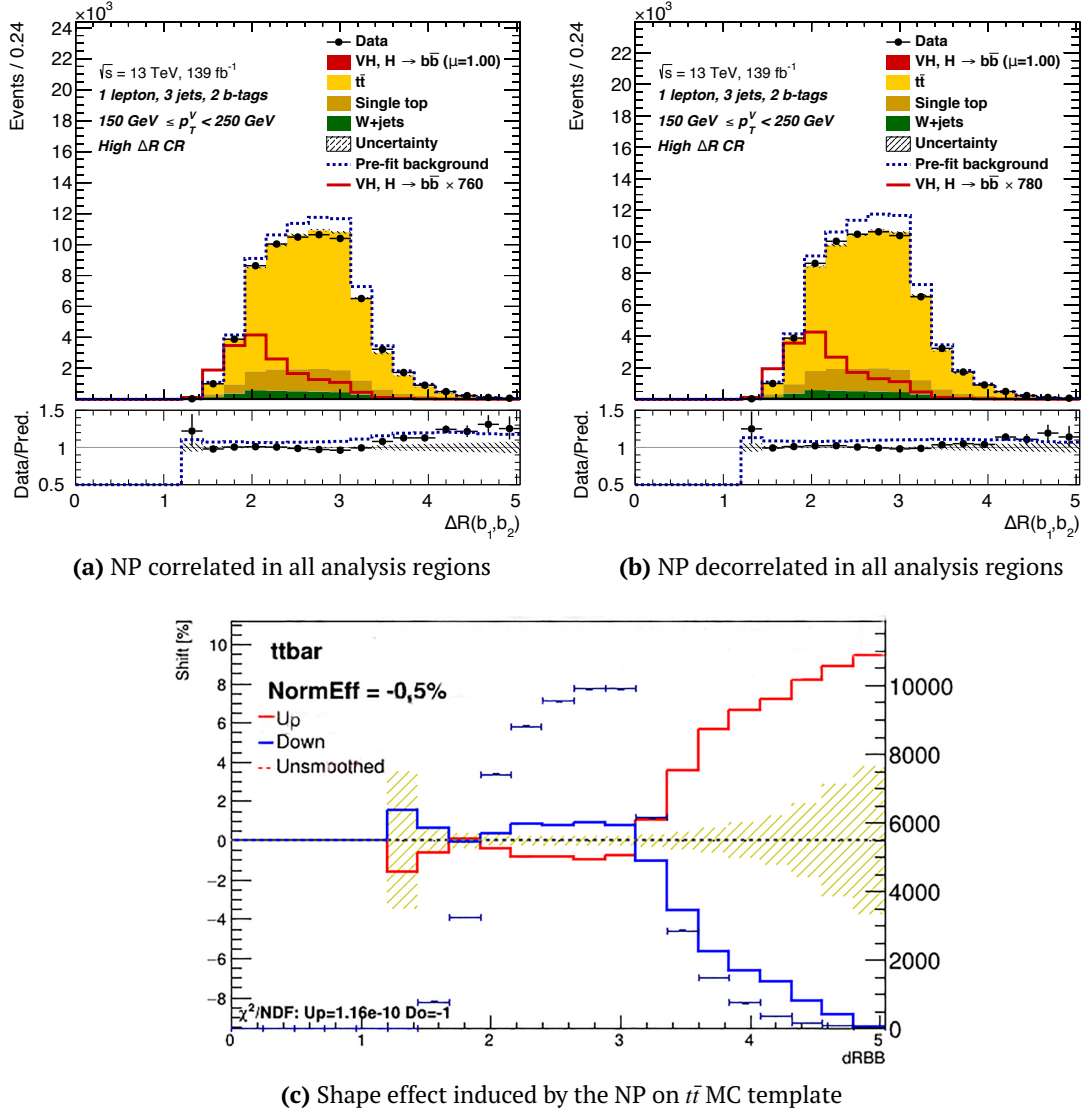


Figure 6.41: (a) Post-fit distributions of the ΔR_{bb} variable in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR, in the case SysBDTr_ $t\bar{t}$ bar_PS is treated as correlated across all analysis regions. (b) Post-fit distributions of the ΔR_{bb} variable in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR, when decorrelating the NP in all analysis regions. The change in the data-MC ratio in the bottom panel between (a) and (b) is compatible with the down variation of the SysBDTr_ $t\bar{t}$ bar_PS in the 1-lepton, 3-jet, $150 < p_T^V < 250$ GeV, High- ΔR CR as shown in blue in (c). The y-label “Shift[%]” indicates how much the systematic increases or decreases the nominal yield in a given bin.

Table 6.20: Pull on the NP associated to the parton shower BDT shape systematics when performing a default fit in all signal and control regions, allowing both the normalisation and the shape of the MC predictions to vary, compared to a fit performed assuming all signal regions having a single bin such that only the normalisation of the MC prediction is allowed to change. It can be observed that the pull on SysBDTr_ttbar_PS in the default fit is compatible with the pull obtained in the fit with all signal regions having a single bin. This means that the pull on the parton shower BDT nuisance parameter is mainly coming from the extrapolation of the $t\bar{t}$ normalisation across the analysis regions.

Systematic	Default fit	Fit with single-bin SRs
SysBDTr_ttbar_PS	-0.93 ± 0.84	-1.36 ± 0.86

direction of the pulls for different jet categories and for both NPs. The decorrelation of the SysTTbarPtV_BDT NP in the 2-jet and 3-jet categories gives very similar fit results (in terms of ranking and breakdown methods, as well as post-fit distributions) to the case in which the NP is correlated in the jet categories. In the case of SysBDTr_ttbar_ME, only the NP acting in the 2-jet category has an impact on the total uncertainty on the signal strength, while the NP in the 3-jet case is more constrained due to the higher number of events in the $t\bar{t}$ MC template.

Given that the SysBDTr_ttbar_ME NP acting in the 2-jet and 3-jet categories have a different impact on the total uncertainty on the signal strength and that in the case both SysTTbarPtV_BDT and SysBDTr_ttbar_ME are decorrelated in the jet categories, this is only 0.5% different from the one obtained in the case both NPs are correlated, it was decided to adopt the decorrelation scheme to give more degrees of freedom to the fit.

Table 6.21: Pulls related to the matrix element BDT and the p_T^V shape systematics when treated as correlated in all analysis regions, p_T^V and jet categories, or uncorrelated only in 2-jet and 3-jet. A tension is observed in the direction of the pulls for the different jet categories.

Name of NP	Correlated	Uncorrelated	
		2-jet	3-jet
SysBDTr_ttbar_ME	$+0.35 \pm 0.44$	-0.25 ± 0.64	0.57 ± 0.47
SysTTbarPTV_BDT	$+1.02 \pm 0.62$	-0.23 ± 0.86	0.99 ± 0.54

Flavour composition uncertainties for the $t\bar{t}$ background Given that the bc component of the $t\bar{t}$ background has a signal-like shape in the m_{bb} distribution, the di-jet mass analysis is expected to be particularly sensitive to a potential flavour composition uncertainty on the bc/bb ratio. This was verified by implementing two new flavour composition uncertainties in the fit model, accounting for the PS and the ME variations, to model the bc/bb and the $other/bb$ ratios. They were derived using Equation 6.34 and implemented

as shape variations on the $t\bar{t}$ MC nominal template. Both the MV and the di-jet mass analyses fits were re-run with these additional flavour composition uncertainties; the bc/bb systematic resulted to be the uncertainty having the second largest impact on the total uncertainty on the signal strength in the di-jet mass analysis, but was proven to be negligible in the MV one. On the other hand, the uncertainty on the $other/bb$ ratio was demonstrated to be negligible in both the MV and the di-jet mass analyses, but was kept in the fit model for completeness.

As an example, the $\pm 1\sigma$ impact on the nominal m_{bb} distribution of the $t\bar{t} bc$ component in the 0-lepton, 2-jet, $150 < p_T^V < 250$ GeV region, induced by the bc/bb uncertainty accounting for the ME variation is shown in Figure 6.42. The blue histogram with error bars shows the very signal-like distribution of the $t\bar{t} bc$ component, the blue line indicates the up variation of the bc/bb uncertainty and the red line represents the correspondent down variation. It can be observed that the shape systematic acts immediately underneath the signal which is the reason why in the di-jet mass analysis, this systematic gives such a high contribution to the total uncertainty on the signal strength.

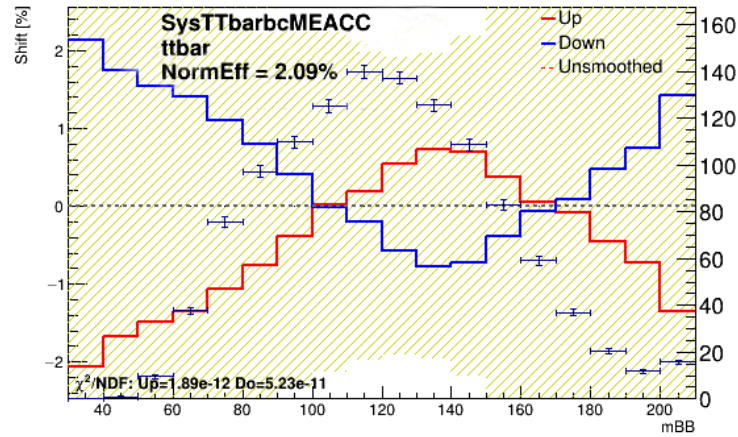


Figure 6.42: Impact of the bc/bb shape uncertainty on the nominal $t\bar{t} bc$ MC template. The blue histogram with error bars shows the nominal distribution, while the blue (red) lines shows the $+1\sigma$ (-1σ) impact of the shape systematic on the nominal template. The y-label “Shift[%]” indicates of how much the systematic increases or decreases the nominal yield in a given bin.

Modification of the pruning procedure A fit without any pruning criteria was performed in all three channels of the analysis to verify the impact of the pruning procedure on the fit outcome. The nuisance parameter associated to the first component of the b -tagging uncertainties for the l -jets was found to be very dependent on the pruning procedure; in fact, the pull from the NP’s central value was found to be largely reduced by the removal of the pruning procedure and, specifically, of the pruning criteria on small uncertainties applied to samples which contribute less than 1% of the total background. This means that, probably, the fit is biased in pulling this NP as too much information is removed in the pruning procedure. It was, thus, decided to reduce the pruning threshold

to 0.005 (see second option of the pruning procedure on samples contributing less than 1% of the total background in Section 6.12) in order not to prune away this NP.

Additional control regions to signal region uncertainties for the W +jets modelling Table 6.22 shows the pulls on the normalisation factors (norm_Wbb_J2 and norm_Wbb_J3) and the NP associated to the BDTr shape uncertainty (SysBDTr_W_SHtoMG5) of the W +hf background, obtained from a fit performed combining all lepton channels, in the context of the diboson cross-check analysis, and using the BDT_{VZ} as discriminant. The fit was performed in four different configurations:

- SR, High- ΔR CR and Low- ΔR CR are considered in the fit;
- only SR and Low- ΔR CR are considered in the fit;
- only the SR and the High- ΔR CR enter the fit;
- only SR is included in the fit.

Table 6.22: Pulls on the normalisation factors of the W +jets background separately in the 2-jet and 3-jet categories and pulls on the BDTr shape uncertainty which is supposed to cover for the Low- ΔR CR and High- ΔR CR to SR extrapolation uncertainties. When removing the Low- ΔR CR from the fitted regions, the values of the estimated normalisation factors decrease and, as a consequence, the pull on the BDTr shape uncertainty increases.

Region	norm_Wbb_J2	norm_Wbb_J3	SysBDTr_W_SHtoMG5
SR + Low- ΔR CR + High- ΔR CR	1.18 ± 0.10	1.15 ± 0.07	-0.06 ± 0.50
SR + Low- ΔR CR	1.16 ± 0.11	1.14 ± 0.09	0.07 ± 0.50
SR + High- ΔR CR	1.03 ± 0.13	0.88 ± 0.11	0.31 ± 0.64
SR	1.07 ± 0.13	0.86 ± 0.12	0.52 ± 0.65

When removing the Low- ΔR CR from the fitted regions, necessary to control the W +jets background, the estimated values of the W +hf normalisation factors decrease and, as a consequence, the pull of the BDTr shape uncertainty, implemented in the SR, tries to recover for this effect. This is a sign that an extrapolation uncertainty from the Low- ΔR CR to the SR is missing. This is expected to be covered by the NP associated to the BDTr shape uncertainty which is, however, only slightly pulled in the default configuration (SR+CRs). Therefore, it was decided to consider only the shape effect (and not the normalisation one) induced by the BDTr shape systematic and to introduce one new nuisance parameter SysWbbCRSRe_{extrap} treated as correlated in all analysis regions and in the 0-lepton and 1-lepton channels and controlling the extrapolation uncertainties from the Low- ΔR CR and High- ΔR CR to the SR. This new NP is derived as the sum in quadrature of the variations from the comparison between SHERPA 2.2.1 and MADGRAPH 5, the renormalization and factorization scales and the PDF variations,

following the procedure explained in Section 6.7. Its values and relative signs are reported in Table 6.23.

Table 6.23: Values of the NP priors controlling the extrapolation uncertainties from the Low- ΔR CR and High- ΔR CR to the SR on the $W+hf$ background in the 0-lepton and 1-lepton channels and for the 2-jet and 3jet categories.

Region	0-lepton		1-lepton	
	2-jet	3-jet	2-jet	3-jet
Low- ΔR CR to SR	-0.115	-0.036	-0.077	-0.056
High- ΔR CR to SR	0.148	0.051	0.149	0.070

Additional control regions to signal region uncertainties for the Z+jets modelling

The m_{bb} shape systematic for the Z+jets background is derived in the inclusive 2-lepton, SR+CRs, $p_T^V > 75$ GeV and ≥ 2 -jets phase space, but removing the $80 < m_{bb} < 140$ GeV region around the VH and VZ processes in order to veto the signal events. Then, a fit to the unit-normalised ratio between data and MC is performed in the sidebands. The NP associated to this shape systematic is applied in an uncorrelated fashion in the medium $75 < p_T^V < 150$ GeV and the high + extreme $p_T^V > 150$ GeV regions. However, when looking at data/MC comparison in the $75 < p_T^V < 150$ GeV region, a clear tension is observed, with the events in the SR following the up m_{bb} variation, and those in the Low- ΔR CR going in the opposite direction. This is shown in Figure 6.43 where the m_{bb} data and MC distributions of the Z+jets background are drawn in the $75 < p_T^V < 150$ GeV range, and in the Low- ΔR CR (a) and SR (b). The red and blue histograms correspond to the up and down variations of the m_{bb} shape uncertainty, respectively. Since this tension resulted to have a non-negligible effect on the $75 < p_T^V < 150$ GeV STXS bin relative to the ZH process¹⁴, it was decided to consider only the shape effect (and not the normalisation one) induced by the m_{bb} shape uncertainties accross different analysis regions.

As such, four new nuisance parameters were introduced to describe the extrapolation uncertainties from the Low- ΔR CR and High- ΔR CR to the SR: those applied in the high+extreme p_T^V regions and treated as correlated in all jet categories in the 0-lepton channel and 2-lepton channel, and those applied only in the medium p_T^V region of the 2-lepton channel and treated as correlated in the 2-jet and 3-jet categories. The extrapolation uncertainties have been derived as the sum in quadrature of the variations from the comparison between SHERPA 2.2.1 and MADGRAPH 5, the renormalization and factorization scales and the PDF variations, following the procedure explained in Section 6.7. Their values and relative sign are reported in Table 6.24.

¹⁴An Asimov dataset was created from scratch without including the Low- ΔR CR. Then, a fit was performed on this Asimov dataset but using real data including the Low- ΔR CR.

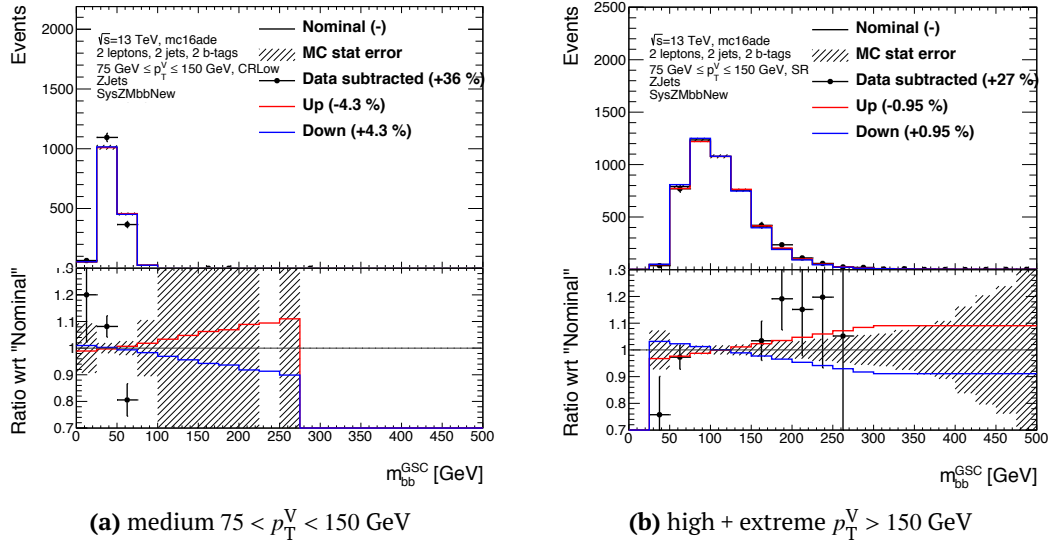


Figure 6.43: Figure (a) shows the m_{bb} distribution of the data and the MC Z+jets template in the $75 < p_T^V < 150$ GeV, Low- ΔR CR, while Figure (b) shows the distributions in the $75 < p_T^V < 150$ GeV, SR. In both figures, the red and blue histograms correspond to the up and down m_{bb} shape variations derived in the inclusive 2-lepton, SR+CRs, $p_T^V > 75$ GeV and ≥ 2 -jets phase space. It can be seen that the data in the first two bins of Figure (a) follow the trend of the down m_{bb} variation while the data in Figure (b) follow the up m_{bb} variation in the full m_{bb} range. (Courtesy of Yohei Noguchi, Kyoto University, Department of Physics.)

Table 6.24: Values of the NP priors controlling the extrapolation uncertainties from the Low- ΔR CR and High- ΔR CR to the SR on the Z+hf background in the 0-lepton and 2-lepton channels and for the 2-jet and 3jet categories.

Name of the NP	Region	0-lepton		2-lepton	
		2-jet	3-jet	2-jet	3-jet
SysZbbCRSRExtrap_CRLow	Low- ΔR CR to SR	-0.060	-0.066	-0.099	-0.038
SysZbbCRSRExtrap_CRHigh	High- ΔR CR to SR	0.038	0.039	0.027	0.041
SysZbbCRSRExtrap_75-150_L2_CRLow	Low- ΔR CR to SR	-	-	-0.099	-0.038
SysZbbCRSRExtrap_75-150_L2_CRHigh	High- ΔR CR to SR	-	-	0.027	0.041

6.14 Fit results

In this section, the results of the global likelihood fit to the data and to the Asimov dataset for the $VH, H \rightarrow b\bar{b}$, $VZ, Z \rightarrow b\bar{b}$ and di-jet mass analyses are presented.

6.14.1 Main $VH, H \rightarrow b\bar{b}$ analysis

Post-fit distributions and yields tables The post-fit normalisation factors of the unconstrained $t\bar{t}$, W +hf and Z +hf backgrounds in the global likelihood fit are shown in Table 6.25 and the post-fit signal and background yields are shown in Table 6.26 for the 0-lepton channel and 1-lepton channels and in Table 6.27 for the 2-lepton channel. Figures 6.44, 6.45 and 6.46 show the BDT_{VH} output distributions in the SRs in the 2-jet and 3-jet categories for the 0-lepton, 1-lepton and 2-lepton channels, respectively.

Table 6.25: Normalisation factors for the $t\bar{t}$, W +hf, and Z +hf backgrounds, obtained from an unconditional fit to the 13 TeV data for the main multivariate analysis. The errors represent the combined statistical and systematic uncertainties [146].

Process and Category	Normalisation factor
$t\bar{t}$ 2-jet	0.98 ± 0.09
$t\bar{t}$ 3-jet	0.93 ± 0.06
W +hf 2-jet	1.06 ± 0.11
W +hf 3-jet	1.15 ± 0.09
Z +hf 2-jet, $75 < p_T^V < 150$ GeV	1.28 ± 0.08
Z +hf 3-jet, $75 < p_T^V < 150$ GeV	1.17 ± 0.05
Z +hf 2-jet, $150 \text{ GeV} < p_T^V$	1.16 ± 0.07
Z +hf 3-jet, $150 \text{ GeV} < p_T^V$	1.09 ± 0.04

Figures 6.47 shows the p_T^V distributions in the SRs in the 2-jet and 3-jet categories for the 0-lepton (top), 1-lepton (middle) and 2-lepton (bottom) channels. Figures 6.48, 6.49 and 6.50 show the $m_{b\bar{b}}$ distributions in the SRs in the 2-jet and 3-jet for the 0-lepton, 1-lepton and 2-lepton channels, respectively.

Signal significance and signal strength μ The measured μ_{VH} signal strength is:

$$\mu_{VH}^{b\bar{b}} = 1.02_{-0.11}^{+0.12}(\text{stat.})_{-0.13}^{+0.14}(\text{syst.}) \quad (6.38)$$

which is 89% compatible with the Standard Model prediction of $\mu = 1$.

The expected significances from the Asimov dataset and the observed ones from the full Run 2 dataset of the $VH, H \rightarrow b\bar{b}$ signal in the three lepton channels separately and combined are reported in Table 6.28.

The expected significances from the Asimov dataset and the observed ones from the full Run 2 dataset for the WH and ZH processes in a combined fit with all three channels

Table 6.26: Signal, background and data yields for each signal region in the 0-lepton and 1-lepton channels after the full event selection. The signal and background yields are normalised to the results of the global likelihood fit. All systematic uncertainties are included in the indicated uncertainties [146].

Signal regions	0-lepton		0-lepton		1-lepton		1-lepton	
	$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$		$p_T^V > 250 \text{ GeV}$		$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$		$p_T^V > 250 \text{ GeV}$	
Sample	2-jet	3-jet	2-jet	3-jet	2-jet	3-jet	2-jet	3-jet
$Z + \text{jets}$	2846 ± 80	3830 ± 160	338 ± 13	533 ± 23	102 ± 5	207 ± 11	10 ± 1	25 ± 2
$W + \text{jets}$	634 ± 63	1500 ± 110	83 ± 9	220 ± 19	1850 ± 160	4080 ± 270	353 ± 33	935 ± 69
Single top	237 ± 35	770 ± 130	9 ± 2	36 ± 8	990 ± 160	3570 ± 600	70 ± 15	313 ± 66
$t\bar{t}$	1157 ± 76	5470 ± 220	39 ± 5	151 ± 16	4600 ± 210	21030 ± 620	188 ± 19	970 ± 58
Diboson	360 ± 55	333 ± 79	86 ± 13	70 ± 17	229 ± 57	264 ± 83	57 ± 15	68 ± 22
Multi-jet	–	–	–	–	130 ± 47	24 ± 14	16 ± 10	7 ± 8
Total bkg.	5234 ± 63	11910 ± 100	554 ± 15	1011 ± 21	7899 ± 82	29170 ± 170	694 ± 23	2318 ± 42
Signal ($\mu = 1.02$)	147 ± 24	130 ± 22	40 ± 6	33 ± 6	148 ± 24	125 ± 21	52 ± 9	44 ± 7
Data	5397	11875	578	1046	8044	29316	727	2378

Table 6.27: Signal, background and data yields for each signal region in the 2-lepton channel after the full event selection. The signal and background yields are normalised to the results of the global likelihood fit. The top background is derived from a data-driven approach from the $e\mu$ -CR data. All systematic uncertainties are included in the indicated uncertainties [146].

Signal regions	2-lepton		2-lepton		2-lepton	
	$75 \text{ GeV} < p_T^V < 150 \text{ GeV}$		$150 \text{ GeV} < p_T^V < 250 \text{ GeV}$		$p_T^V > 250 \text{ GeV}$	
Sample	2-jet	$\geq 3\text{-jet}$	2-jet	$\geq 3\text{-jet}$	2-jet	$\geq 3\text{-jet}$
$Z + \text{jets}$	5900 ± 100	11630 ± 160	716 ± 19	2499 ± 52	84 ± 3	537 ± 16
$W + \text{jets}$	1 ± 0	6 ± 0	< 1	2 ± 0	< 1	< 1
Top	3193 ± 57	8796 ± 87	52 ± 7	389 ± 19	1 ± 1	15 ± 4
Diboson	283 ± 47	443 ± 78	83 ± 14	169 ± 30	20 ± 4	52 ± 10
Total bkg.	9378 ± 86	20880 ± 130	851 ± 19	3058 ± 44	106 ± 4	605 ± 14
Signal ($\mu = 1.02$)	78 ± 14	106 ± 21	34 ± 6	59 ± 12	10 ± 2	18 ± 3
Data	9463	20927	881	3148	123	614

Table 6.28: Expected significances from the Asimov dataset and the observed ones from the full Run 2 dataset for a single signal strength μ_{VH} , or three corresponding to the three lepton channels.

Channel	Expected sig.	Observed sig.
MVA VH analysis		
0L	4.1	4.4
1L	4.1	4.0
2L	4.3	4.2
0L+1L+2L	6.7	6.7

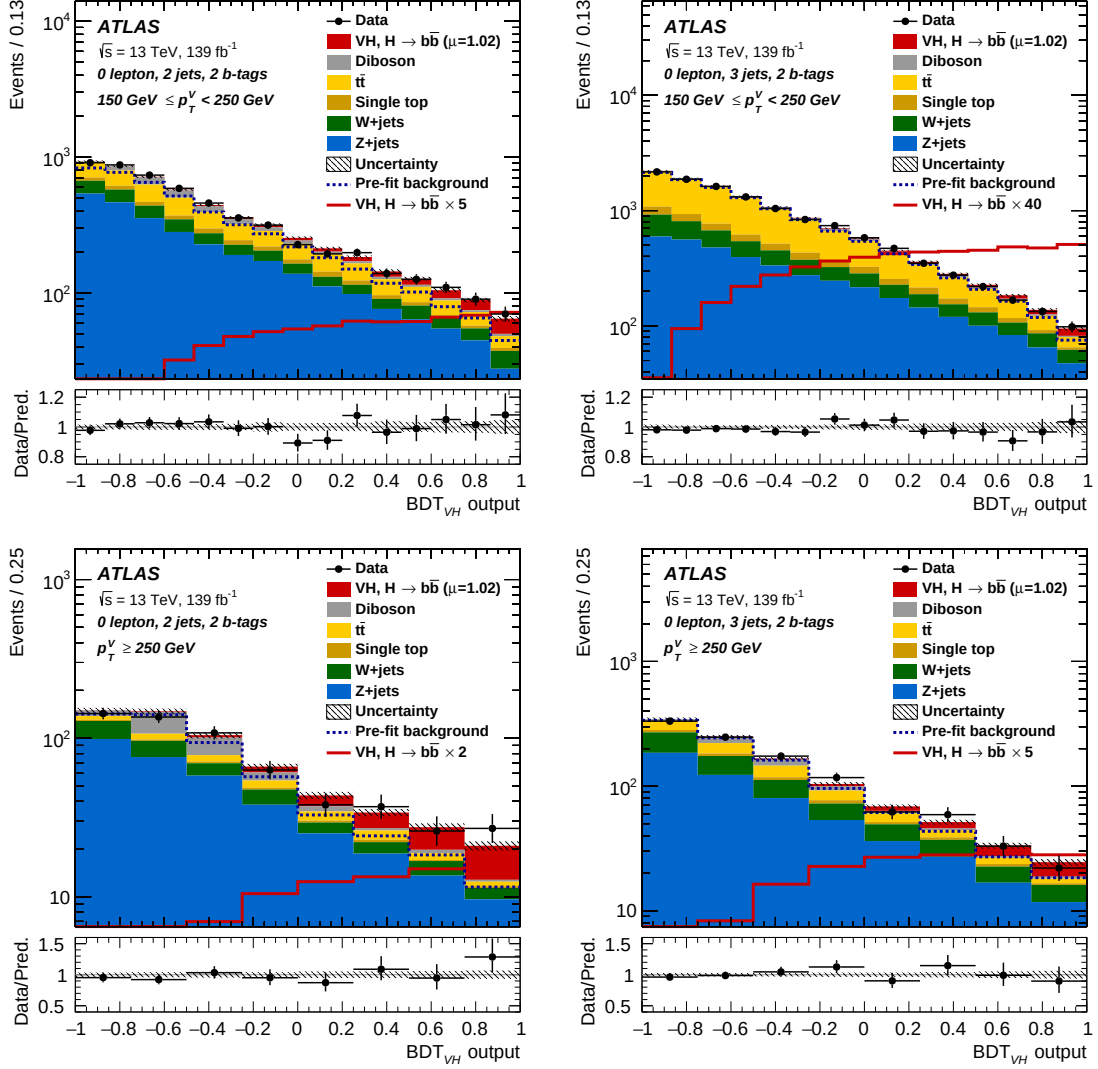


Figure 6.44: Post-fit distributions of the BDT_{VH} discriminant in the 0-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250 \text{ GeV}$ (top) and $p_T^V \geq 250 \text{ GeV}$ (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13 \text{ TeV}$ while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125 \text{ GeV}$) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

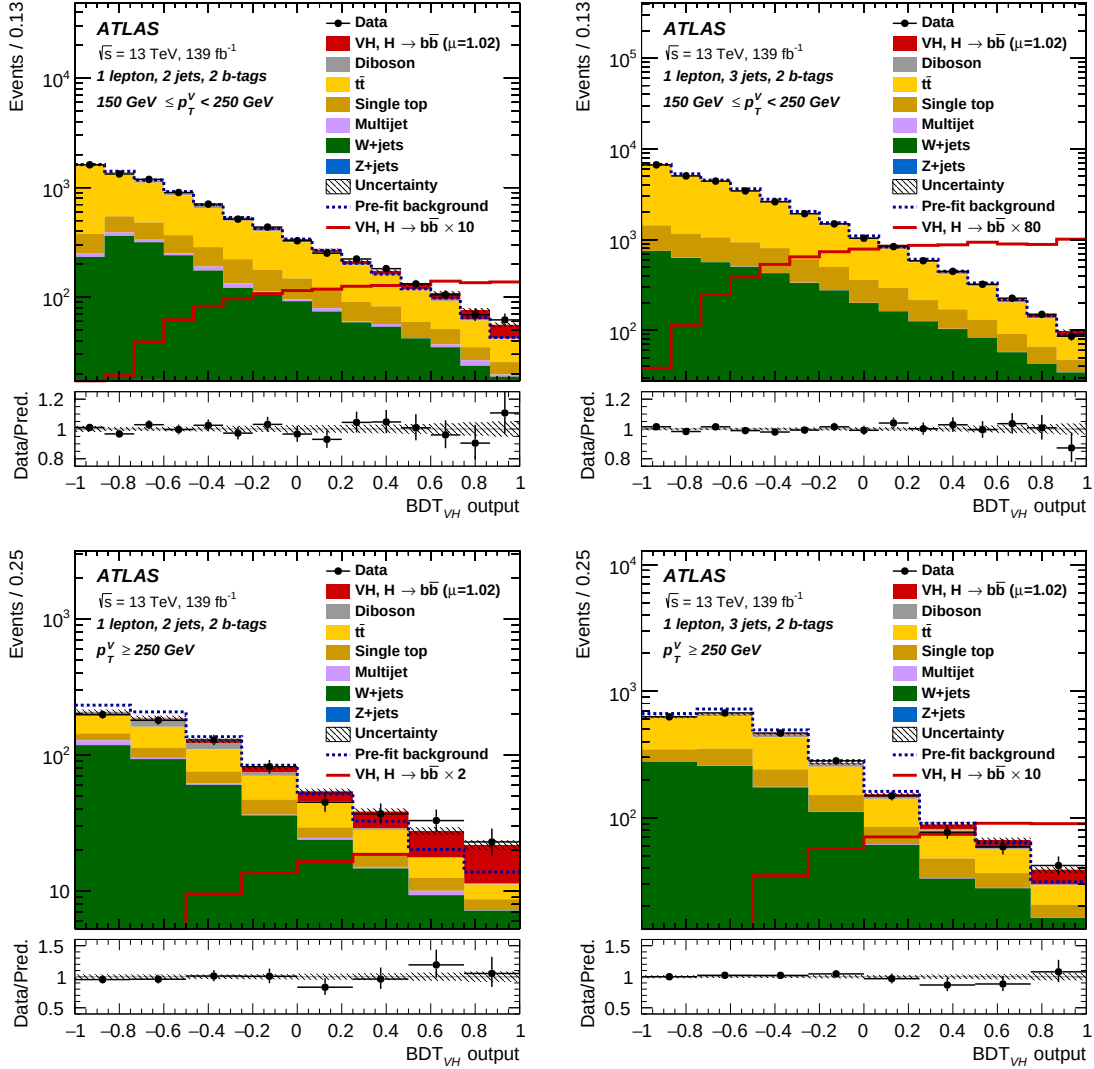


Figure 6.45: Post-fit distributions of the BDT_{VH} discriminant in the 1-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250$ GeV (top) and $p_T^V \geq 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

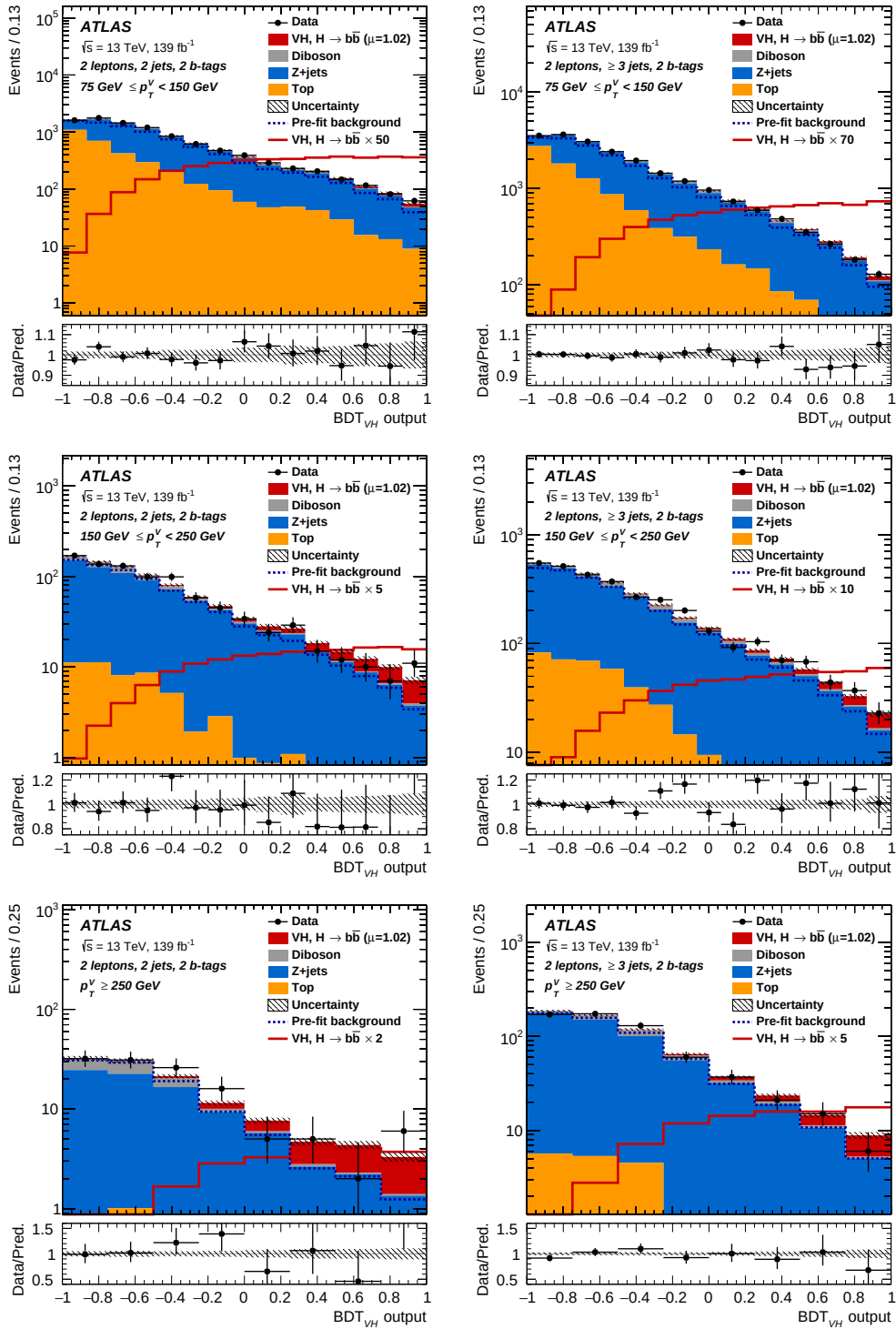


Figure 6.46: Post-fit distributions of the BDT_{VH} discriminant in the 2-lepton channel for 2-b-tag, 2-jet (on the left) and 3-jet (on the right) events and for the $75 < p_T^V < 150$ GeV (top), $150 < p_T^V < 250$ GeV (middle) and $p_T^V > 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

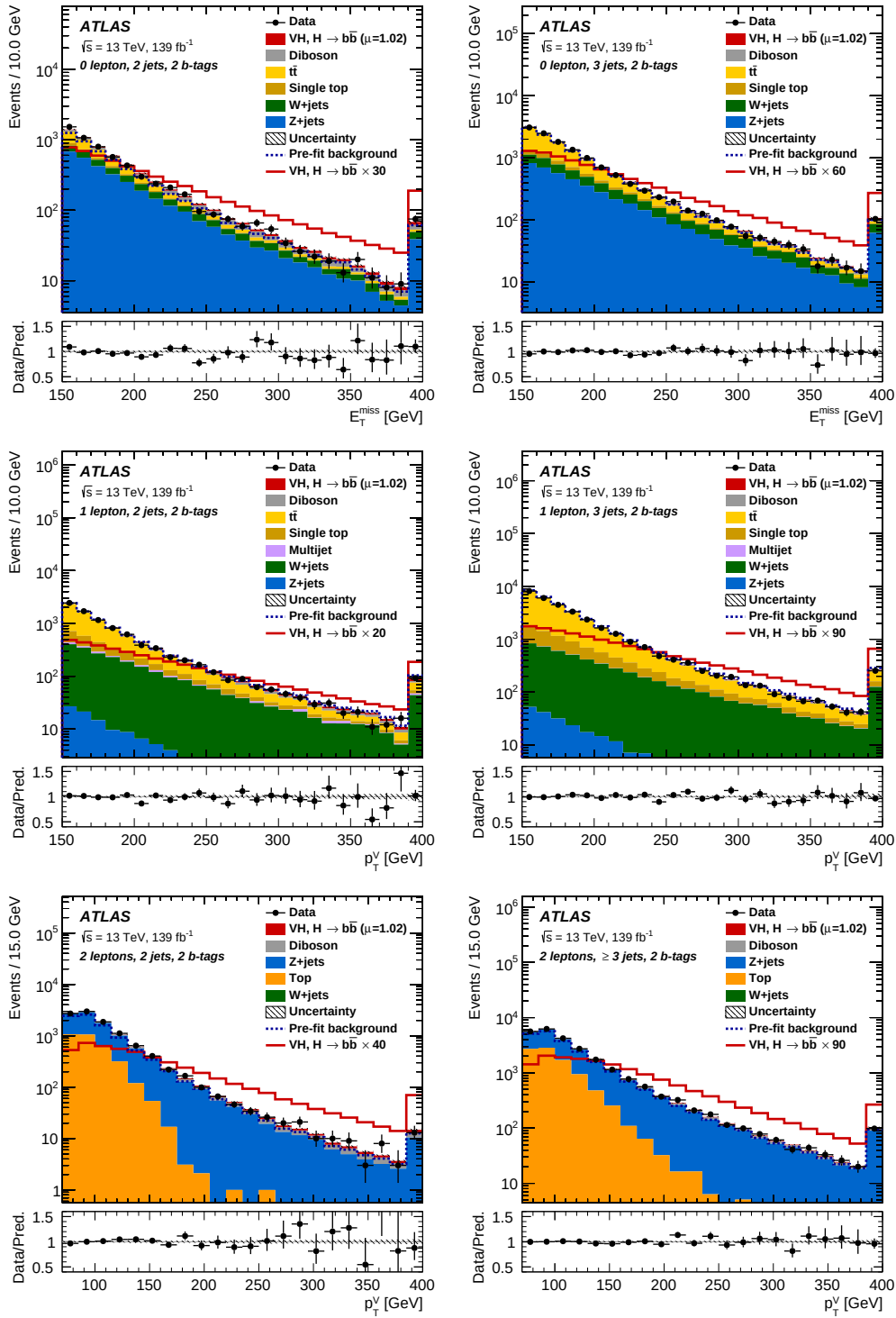


Figure 6.47: Post-fit distributions of the p_T^V variable in the 0-lepton (top), 1-lepton (middle) and 2-lepton (bottom) channels for 2-b-tag, 2-jet (on the left) and 3-jet (on the right) events and inclusively in $p_T^V > 150$ GeV for the 0-lepton and 1-lepton channels, and in $p_T^V > 75$ GeV for the 2-lepton channel. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13 \text{ TeV}$ while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125 \text{ GeV}$) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

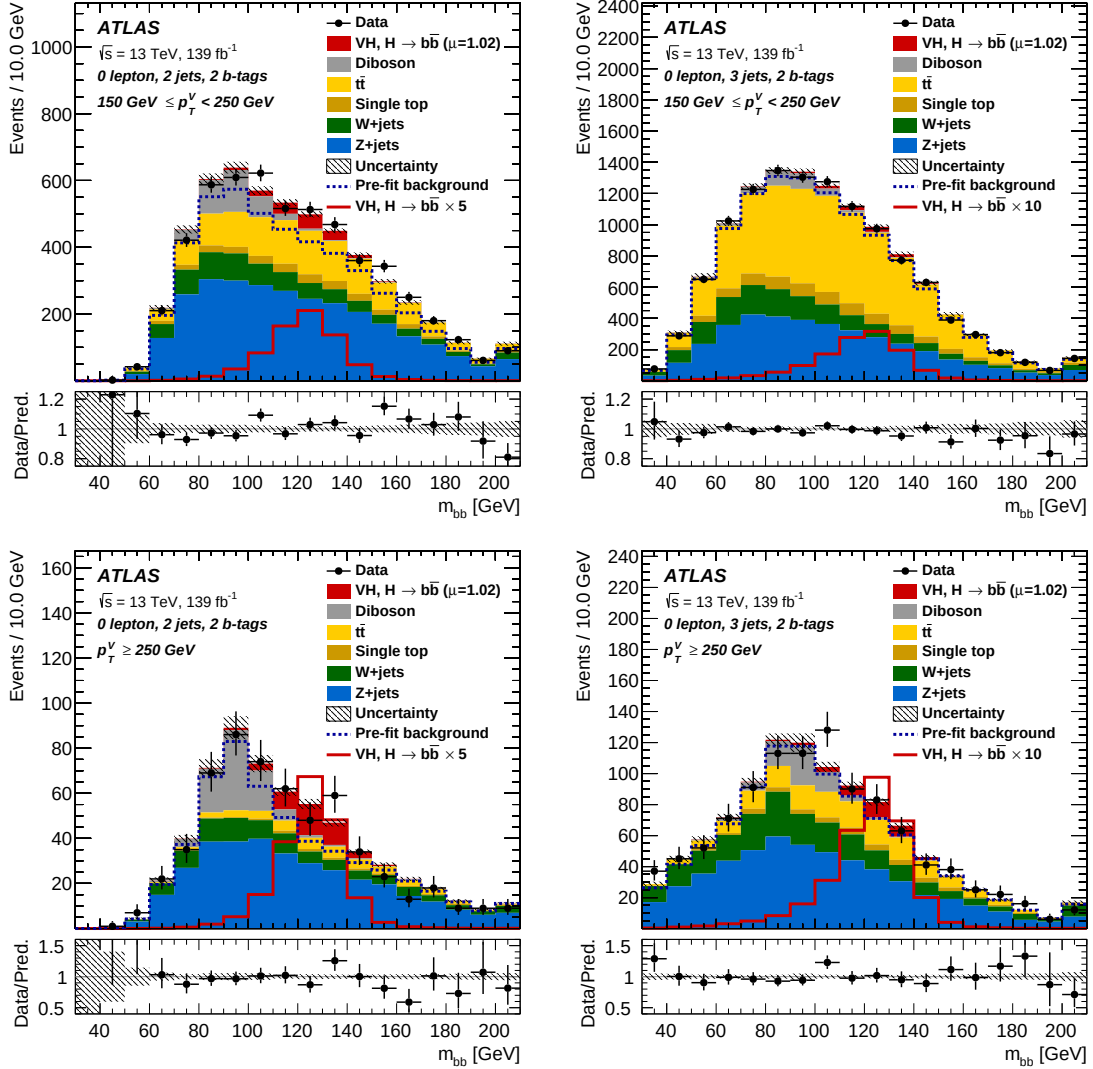


Figure 6.48: Post-fit distributions of the m_{bb} variable in the 0-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250$ GeV (top) and $p_T^V > 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

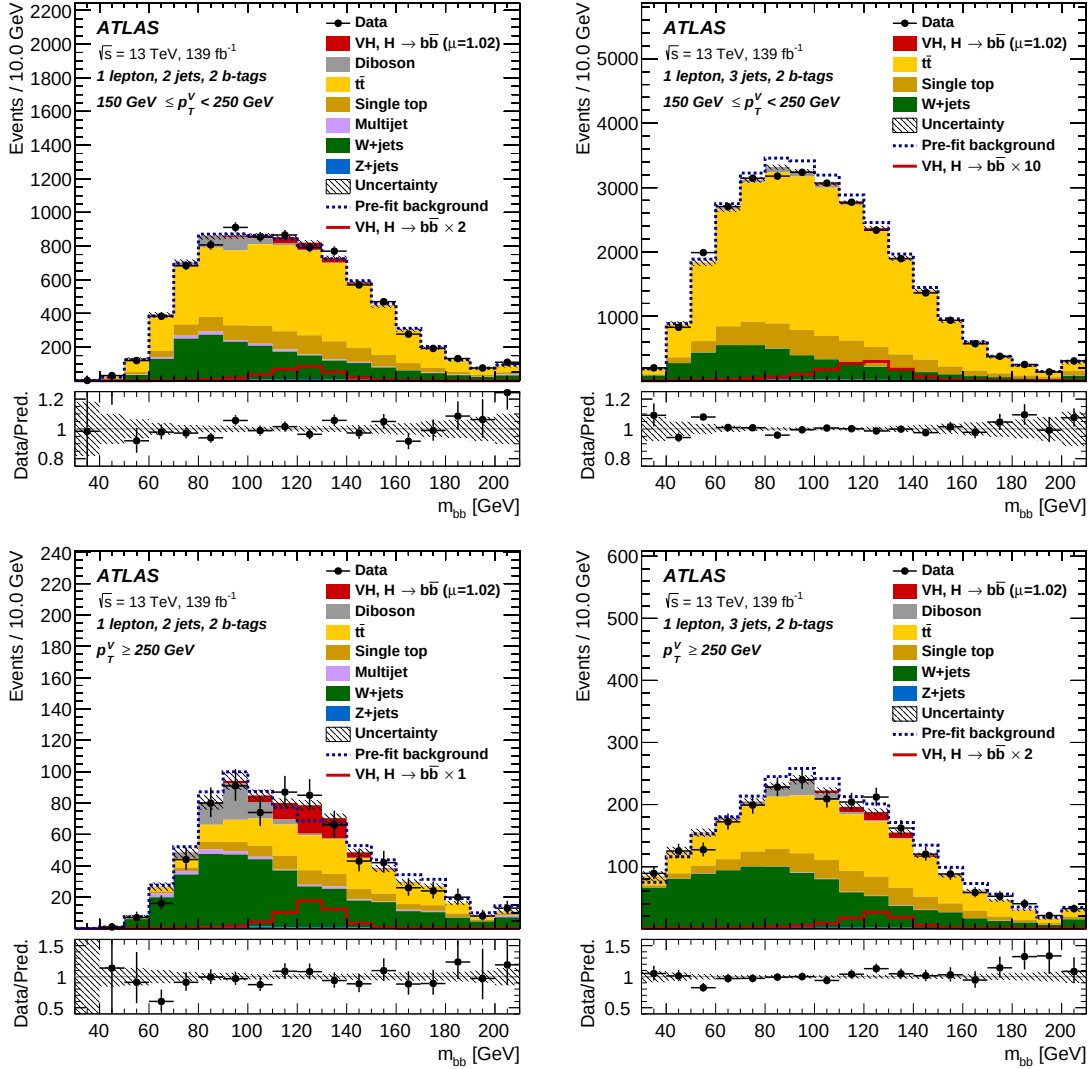


Figure 6.49: Post-fit distributions of the m_{bb} variable in the 1-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250$ GeV (top) and $p_T^V > 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

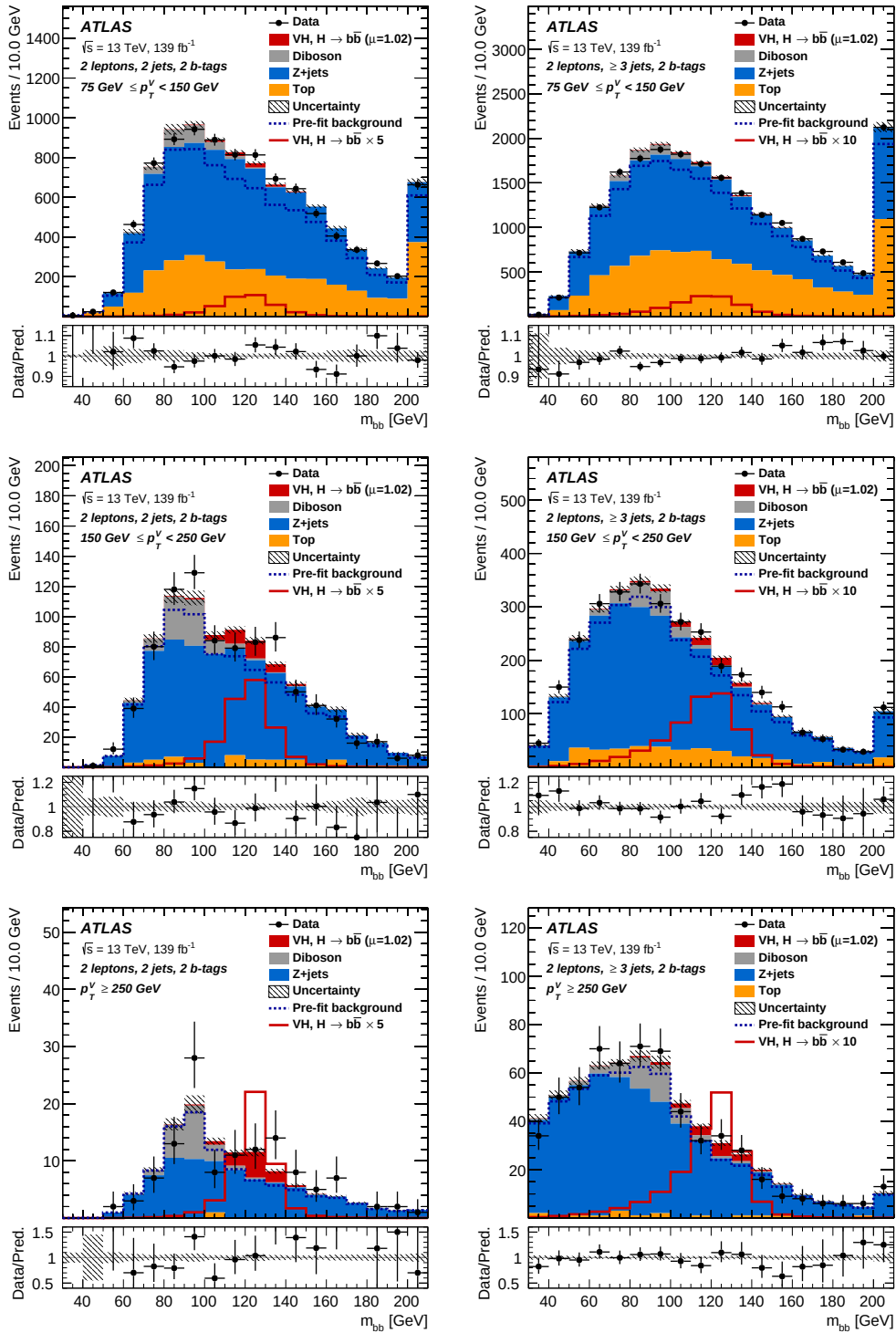


Figure 6.50: Post-fit distributions of the m_{bb} variable in the 2-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $75 < p_T^V < 150$ GeV (top), $150 < p_T^V < 250$ GeV (middle) and $p_T^V > 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The Higgs boson signal ($m_H = 125$ GeV) is shown as a filled red histogram and normalised to the signal strength extracted from data ($\mu = 1.02$), and unstacked scaled by the factor indicated in the legend to illustrate the signal shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

and with two parameters of interests, μ_{WH} and μ_{ZH} are reported in Table 6.29.

Table 6.29: Expected significances from the Asimov dataset and the observed ones from the full Run 2 dataset for the WH and ZH processes in a combined fit with all three channels and with two parameters of interests, μ_{WH} and μ_{ZH} .

Channel	Expected sig.	Observed sig.
MVA VH analysis		
WH	4.1	4.0
ZH	5.1	5.3

Observation of the ZH process can be claimed, together with a strong evidence of the WH production mode.

In Figure 6.51 (a), the signal strength from the unconditional fit to the data in all three channels combined is compared to the case where a signal strength parameter is left floating, independently, for each lepton channel. A very good compatibility of 92% is found between the combined signal strength and those in each individual lepton channel. Figure 6.51 (b) shows the case where a signal strength parameter is left floating independently for each signal process, WH and ZH . A compatibility of 71% is observed between the combined and the individual signal strengths.

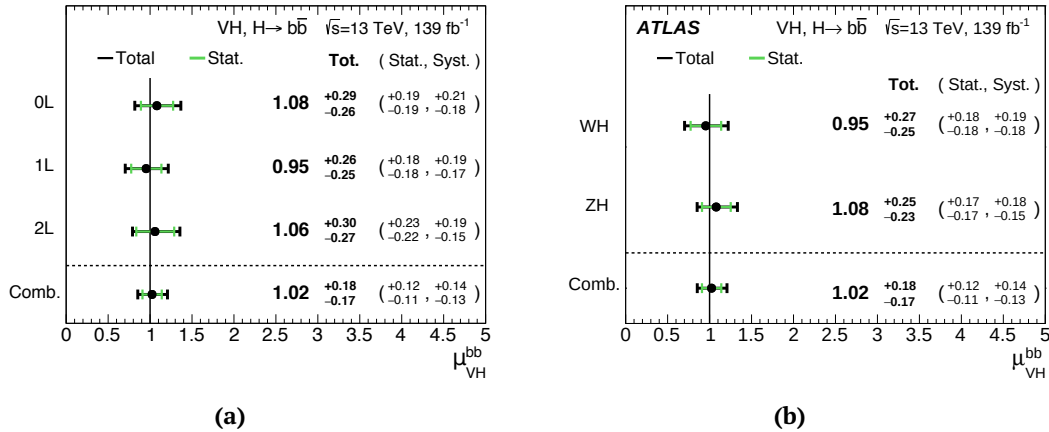


Figure 6.51: Figure (a) shows the comparison between the μ value obtained from a fit to the three lepton channels combined, with the case where a signal strength is left floating independently for each lepton channel. Figure (b) shows the case a signal strength is left floating independently for each signal process WH and ZH [146].

In Figure 6.52 the events for both data and MC prediction are ordered as a function of $\log_{10}(S/B)$, with S (B) the signal (background) yield in each bin, after applying the global likelihood fit results. In the lower panel, the pull corresponds to the difference between data and background-only model, divided by the statistical uncertainty. The full red line indicates the pull expected for a signal + background hypothesis. A nice excess is observed at high values of $\log_{10}(S/B)$.

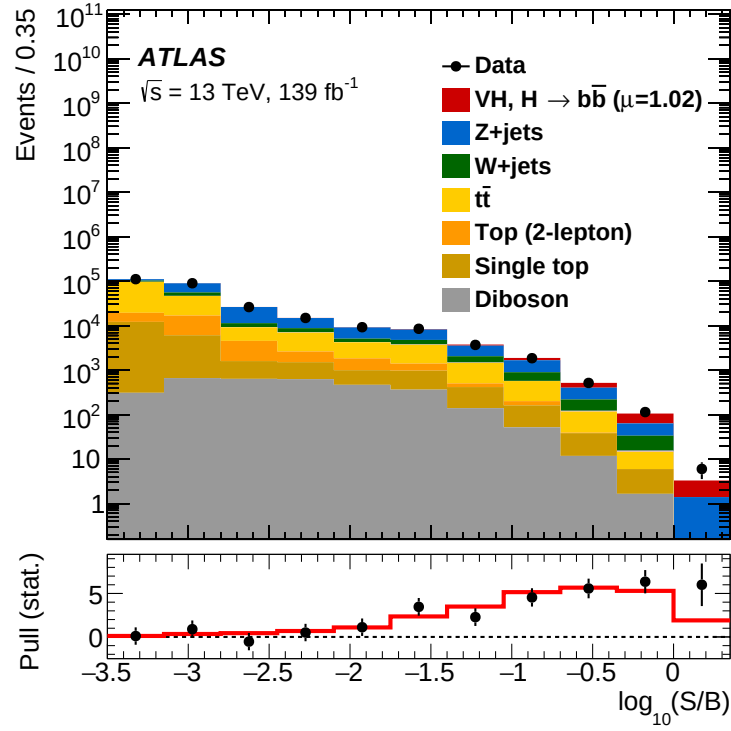


Figure 6.52: The event yields of each bin of the BDT_{VH} discriminants are ordered as a function of $\log_{10}(S/B)$, with S (B) the signal (background) yield in each bin, after applying the global likelihood fit results. The signal contribution is rescaled by the signal strength obtained from an unconditional fit to the Run 2 data. In the lower panel, the pull of the data relative to the background-only hypothesis is shown with statistical error only. The red line indicates the pull expected for a signal+background hypothesis [146].

Nuisance parameters ranking and breakdown To evaluate the impact of the systematic uncertainties on the total uncertainty on the signal strength, $\Delta\mu$, the ranking and breakdown methods are used.

Table 6.30 shows the impact of each group of systematic uncertainties on the fitted combined signal strength, μ_{VH} , calculated using the breakdown method. It can be seen that the analysis is systematically dominated and the largest contributions arise from the systematics on the jets, on the b -tagging, and on the modelling of the signal and W +jets background. This is reflected in Figure 6.53 which shows the nuisance parameters having an impact on the VH signal strength. The largest impact among the experimental systematics is coming from the b -tagging uncertainties and those related to the jets. On the modelling side, the BDTr shape systematic on the W +jets background shows a significant impact on $\Delta\mu$, being the second ranked systematic. Also the modelling of $t\bar{t}$, Z +jets and diboson backgrounds have a significant impact, as well as the theory uncertainties.

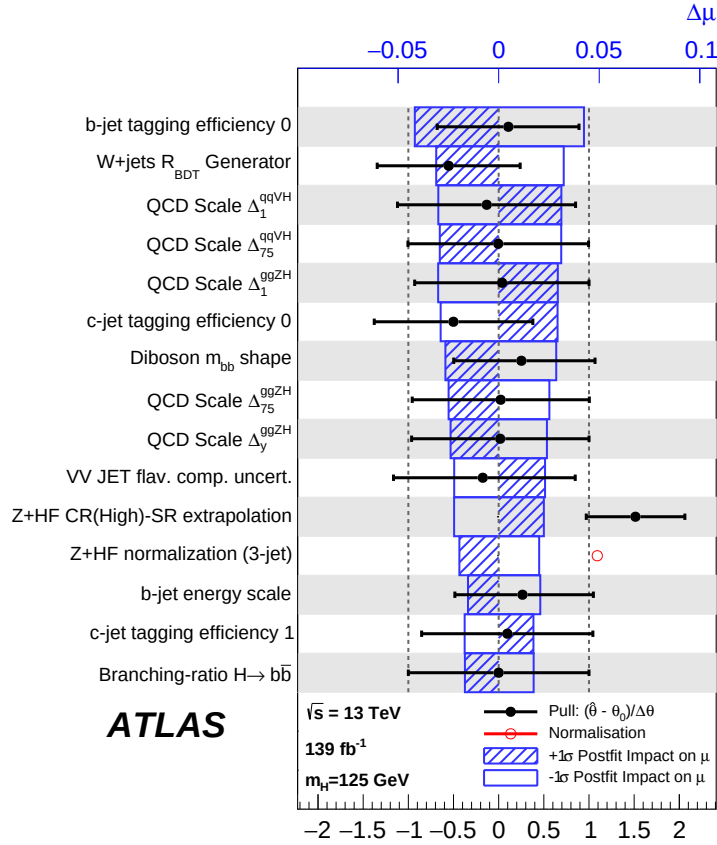


Figure 6.53: Impact of the systematic uncertainties on the VH signal strength. The systematic uncertainties are listed in decreasing order of their impact on μ when evaluated using the ranking method. The blue boxes show the upwards and downwards variations, with the impact referring to the top x -axis. The filled circles, referring to the bottom x -axis, show the deviations of the fitted nuisance parameters $\hat{\theta}$ from their nominal values θ_0 [146].

Table 6.30: Breakdown of the contributions to the total uncertainty on the VH signal strength measurements. The sum in quadrature of the systematic uncertainties relative to the different sources differs from the total systematic uncertainty due to correlations. [146].

Source of uncertainty	VH	σ_μ WH	ZH
Total	0.177	0.260	0.240
Statistical	0.115	0.182	0.171
Systematic	0.134	0.186	0.168
Statistical uncertainties			
Data statistical	0.108	0.171	0.157
$t\bar{t} e\mu$ control region	0.014	0.003	0.026
Floating normalisations	0.034	0.061	0.045
Experimental uncertainties			
Jets	0.043	0.050	0.057
E_T^{miss}	0.015	0.045	0.013
Leptons	0.004	0.015	0.005
b -tagging	b -jets	0.045	0.025
	c -jets	0.035	0.068
	light-flavour jets	0.009	0.004
Pile-up	0.003	0.002	0.007
Luminosity	0.016	0.016	0.016
Theoretical and modelling uncertainties			
Signal	0.072	0.060	0.107
Z + jets	0.032	0.013	0.059
W + jets	0.040	0.079	0.009
$t\bar{t}$	0.021	0.046	0.029
Single top quark	0.019	0.048	0.015
Diboson	0.033	0.033	0.039
Multi-jet	0.005	0.017	0.005
MC statistical	0.031	0.055	0.038

Cross section measurement in the STXS framework Following the description of the STXS scheme in Section 6.8.1, the cross section times branching ratio is measured through a global likelihood fit with five POIs, two for the WH process in the $150 < p_T^V < 250$ GeV and $p_T^V > 250$ GeV regions, and three for the ZH process in the $75 < p_T^V < 150$ GeV, $150 < p_T^V < 250$ GeV and $p_T^V > 250$ GeV regions. The results of these measurements are reported in Table 6.31. Figure 6.54 shows the compatibility between the measured values and the theoretical prediction; very good sensitivity (more than 3σ) can be observed in the WH $p_T^V > 250$ GeV, ZH $150 < p_T^V < 250$ GeV and $p_T^V > 250$ GeV bins.

Table 6.31: Best-fit values and uncertainties for the $VH, H \rightarrow b\bar{b}$ cross-section times the $H \rightarrow b\bar{b}$ BR and the $V \rightarrow \text{leptons}$ BR, in the STXS scheme. Four contributions are accounted for in the total uncertainty and are shown separately: the statistical uncertainty (Stat. unc.), the experimental uncertainties (Exp.) and uncertainties from the signal (Th. sig.) and background (Th. bkg.) modelling. The total systematic uncertainty, equal to the difference in quadrature between the total uncertainty and the statistical one, differs from the sum in quadrature of the signal, backgrounds and experimental uncertainties due to correlations [146].

STXS region		SM prediction [fb]	Result [fb]	Stat. unc.			Syst. unc. [fb]		
Process	$p_T^{V,t}$ interval			[fb]	Th. sig.	Th. bkg.	Exp.		
$W(\ell\nu)H$	150–250 GeV	24.0 $\pm$ 1.1	19.0 $\pm$ 12.1	$\pm$ 7.7	$\pm$ 0.9	$\pm$ 5.5	$\pm$ 6.0		
$W(\ell\nu)H$	> 250 GeV	7.1 $\pm$ 0.3	7.2 $\pm$ 2.2	$\pm$ 1.9	$\pm$ 0.4	$\pm$ 0.8	$\pm$ 0.7		
$Z(\ell\ell/\nu\nu)H$	75–150 GeV	50.6 $\pm$ 4.1	42.5 $\pm$ 35.9	$\pm$ 25.3	$\pm$ 5.6	$\pm$ 17.2	$\pm$ 19.7		
$Z(\ell\ell/\nu\nu)H$	150–250 GeV	18.8 $\pm$ 2.4	20.5 $\pm$ 6.2	$\pm$ 5.0	$\pm$ 2.3	$\pm$ 2.4	$\pm$ 2.3		
$Z(\ell\ell/\nu\nu)H$	> 250 GeV	4.9 $\pm$ 0.5	5.4 $\pm$ 1.7	$\pm$ 1.5	$\pm$ 0.5	$\pm$ 0.5	$\pm$ 0.3		

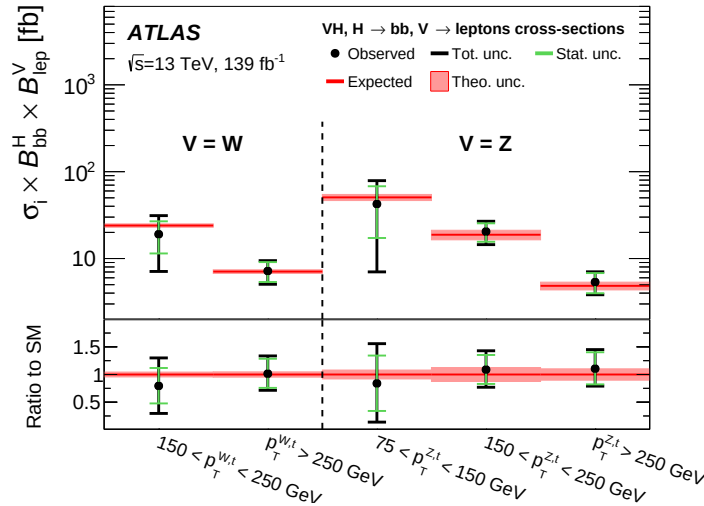


Figure 6.54: Measured $VH, H \rightarrow b\bar{b}$ cross-sections times the $H \rightarrow b\bar{b}$ BR and the $V \rightarrow \text{leptons}$ BR in the stage 1.2 STXS scheme [146].

6.14.2 Diboson $VZ, Z \rightarrow b\bar{b}$ cross-check analysis

In this section, the results from the $VZ, Z \rightarrow b\bar{b}$ analysis, with the BDT_{VZ} output as discriminant of the global likelihood fit, are presented. The event selection, as well as the definition of fit regions and the systematic uncertainties, are identical to those used for the main $VH, H \rightarrow b\bar{b}$ analysis. The $VH, H \rightarrow b\bar{b}$ signal is added as a background process normalised to the predicted SM cross section with a prior uncertainty of 50% to encompass the measurement and uncertainty performed in the $H \rightarrow b\bar{b}$ observation publication [13].

Post-fit distributions Figures 6.55, 6.56 and 6.57 show the BDT_{VZ} output distributions in the SRs in the 2-jet and 3-jet for the 0-lepton, 1-lepton and 2-lepton channels, respectively.

Signal significance and signal strength μ The measured signal strength and the associated uncertainty is:

$$\mu_{VZ}^{b\bar{b}} = 0.93^{+0.07}_{-0.06}(\text{stat.})^{+0.14}_{-0.12}(\text{syst.}) \quad (6.39)$$

In Figure 6.58 (a) the signal strength from the unconditional fit to the data in all three channels combined is compared to the case where a signal strength parameter is left floating, independently, for each lepton channel. A compatibility of 52% is found between the combined signal strength and those obtained in each lepton channel. Figure 6.58 (b) shows the case where a signal strength parameter is left floating independently for each signal process, WZ and ZZ . The compatibility between the two measurements and the combined one is 27%.

6.14.3 $VH, H \rightarrow b\bar{b}$ di-jet mass analysis

In this section, the results of a global likelihood fit with the m_{bb} distribution as discriminant are presented. The event selection, as well as the definition of the fit regions and the systematic uncertainties, are identical to those used for the main $VH, H \rightarrow b\bar{b}$ analysis. The measured signal strength and the associated uncertainty is:

$$\mu_{VH}^{b\bar{b}} = 1.17^{+0.25}_{-0.23} = 1.17 \pm 0.16(\text{stat.})^{+0.19}_{-0.16}(\text{syst.}). \quad (6.40)$$

This is different from the one derived in the main MV analysis, which was used in the m_{bb} post-fit plots shown in Section 6.14.1, but the two are in good agreement.

The background-only hypothesis is rejected with an observed significance of 5.5σ , compared to an expectation of 4.9σ . Therefore, the observation of the $VH, H \rightarrow b\bar{b}$ signal with the di-jet mass analysis can be claimed. It has to be noted, though, that the di-jet mass analysis is 40% less sensitive than the main MV analysis.

Figure 6.59 shows the post-fit m_{bb} distribution summed over all channels and regions, weighted by their respective ratio of the fitted signal yield to the background yield, S/B

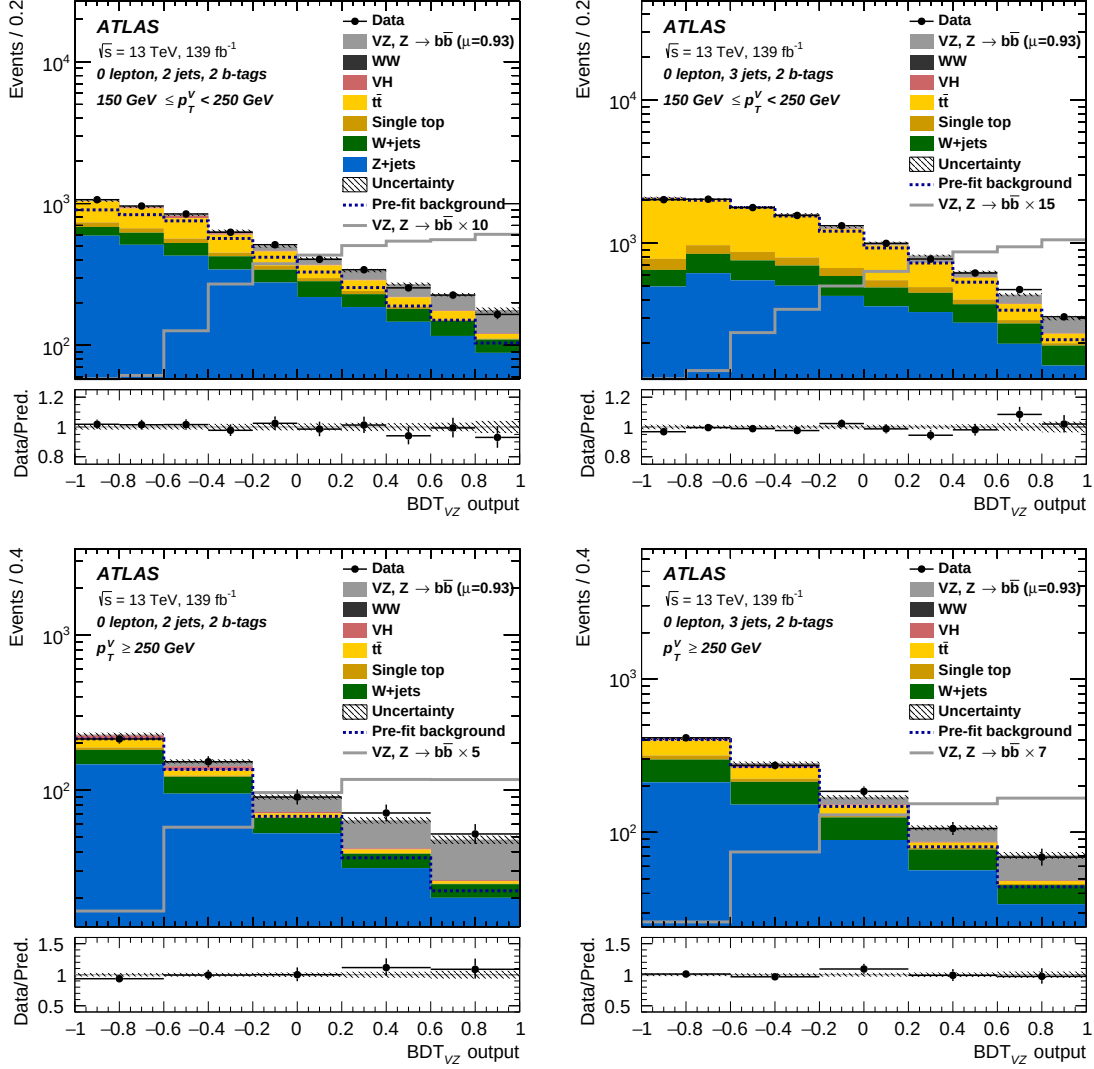


Figure 6.55: Post-fit distributions of the BDT_{VZ} discriminant in the 0-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250$ GeV (top) and $p_T^V \geq 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The diboson signal is shown as a filled grey histogram and normalised to the signal strength extracted from data ($\mu = 0.93$), and unstacked scaled by the factor indicated in the legend to illustrate its shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

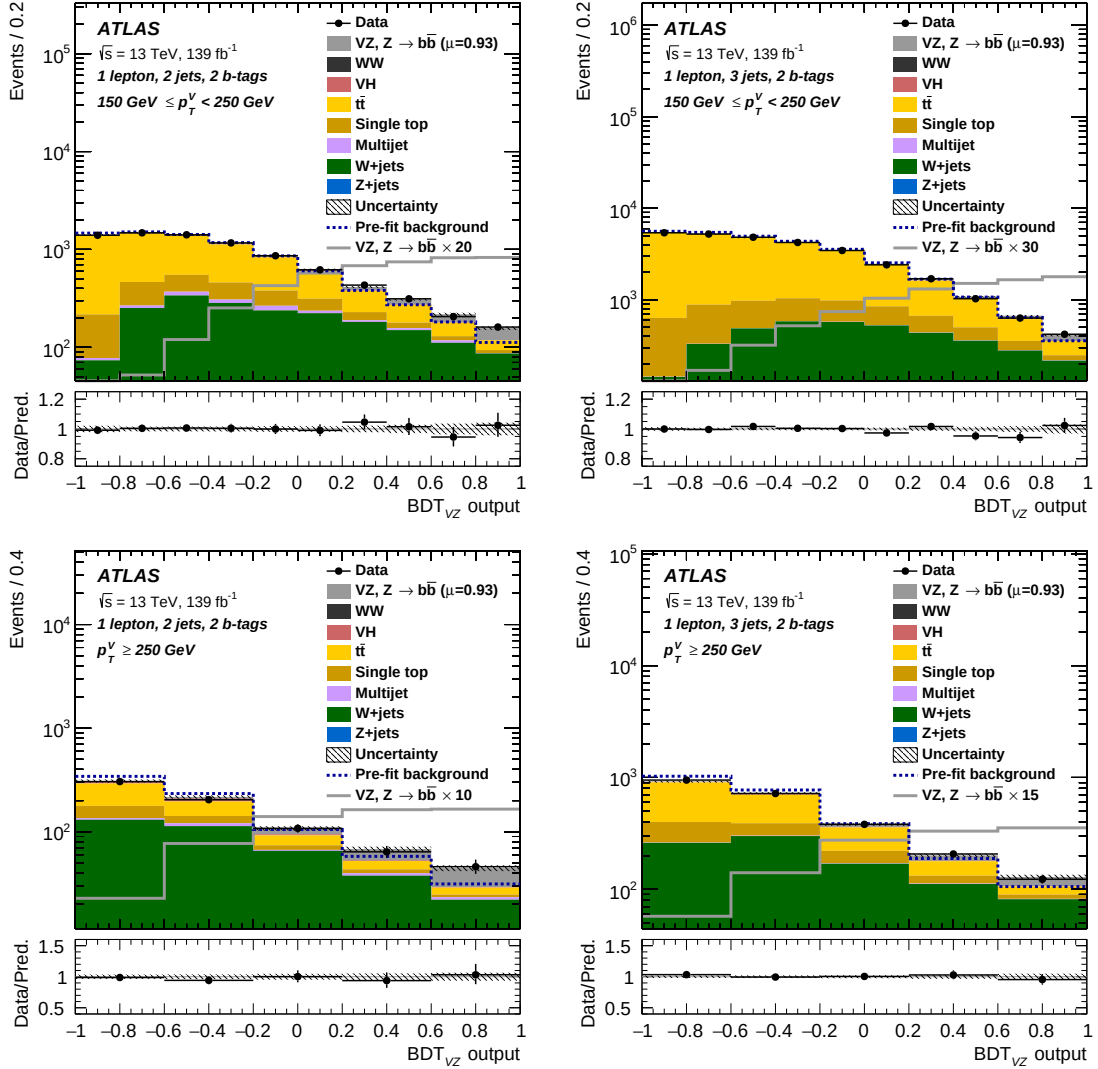


Figure 6.56: Post-fit distributions of the BDT_{VZ} discriminant in the 1-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $150 < p_T^V < 250$ GeV (top) and $p_T^V \geq 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The diboson signal is shown as a filled grey histogram and normalised to the signal strength extracted from data ($\mu = 0.93$), and unstacked scaled by the factor indicated in the legend to illustrate its shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

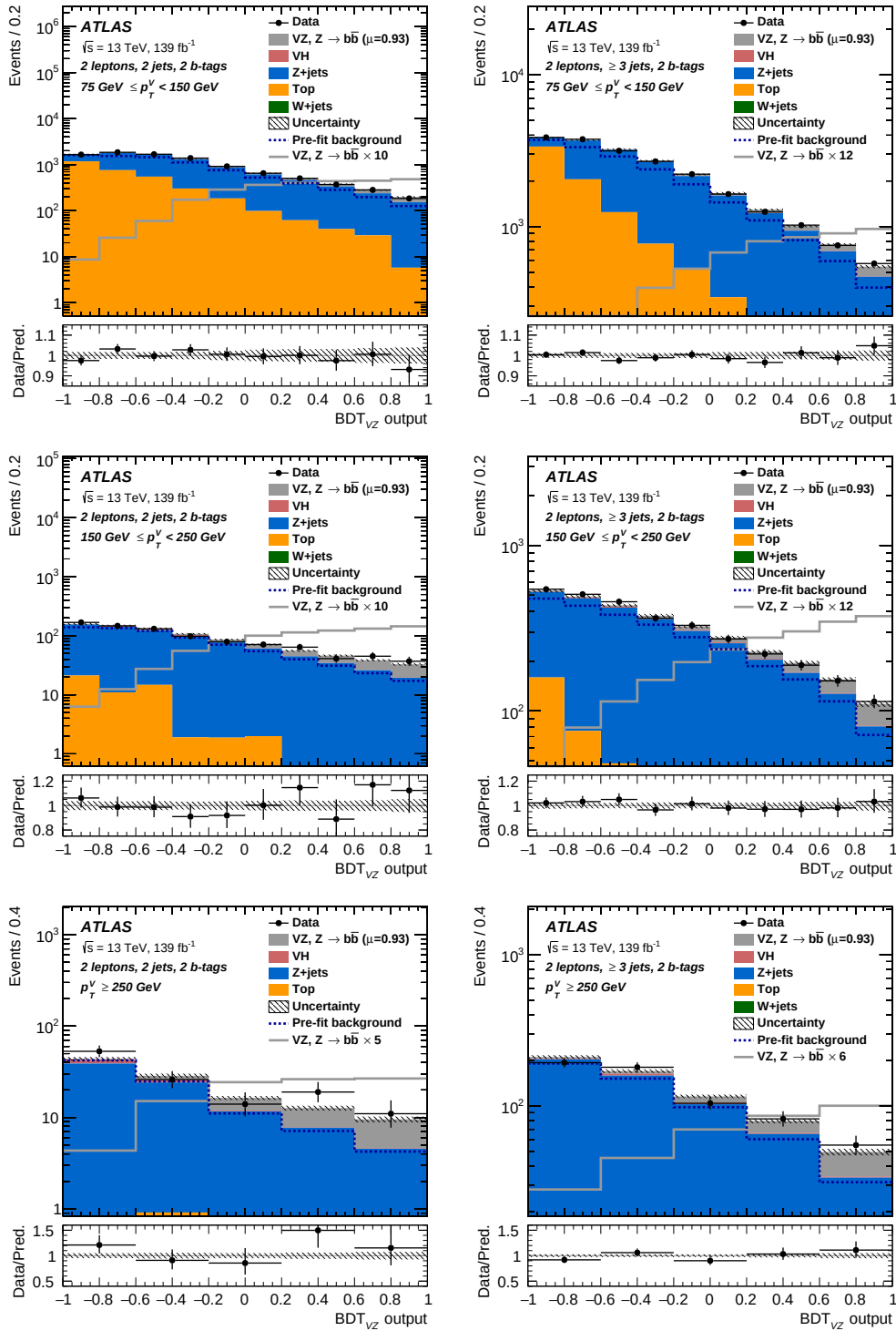


Figure 6.57: Post-fit distributions of the BDT_{VZ} discriminant in the 2-lepton channel for 2- b -tag, 2-jet (on the left) and 3-jet (on the right) events and for the $75 < p_T^V < 150$ GeV (top) $150 < p_T^V < 250$ GeV (middle) and $p_T^V > 250$ GeV (bottom) regions. The black dots correspond to the 139 fb^{-1} of Run 2 data collected at $\sqrt{s} = 13$ TeV while the background prediction is shown by the coloured stacked histograms. The diboson signal is shown as a filled grey histogram and normalised to the signal strength extracted from data ($\mu = 0.93$), and unstacked scaled by the factor indicated in the legend to illustrate its shape. The dashed histogram corresponds to the sum of the MC predictions of all backgrounds before the fit. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The bottom pad shows the ratio of the data to the combined post-fit signal and backgrounds [146].

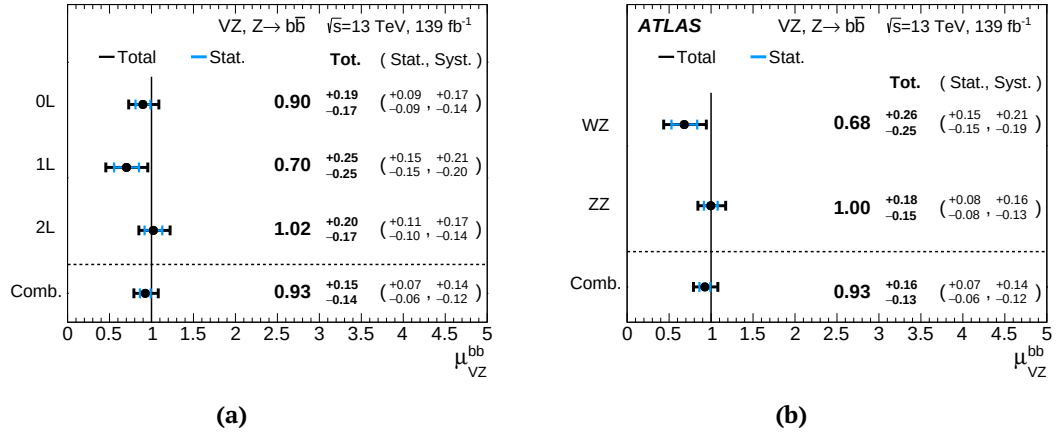


Figure 6.58: Figure (a) shows the comparison between the μ value obtained from a fit to the three lepton channels combined, with the case where a signal strength is left floating independently for each lepton channel. Figure (b) shows the case a signal strength is left floating independently for each signal process WZ and ZZ [146].

and after the subtraction of all the backgrounds, except for the WZ and ZZ processes. Two nice peaks can be observed, the grey one being the diboson peak and the red one being the $VH, H \rightarrow b\bar{b}$ signal peak. Good agreement is found between data and MC prediction.

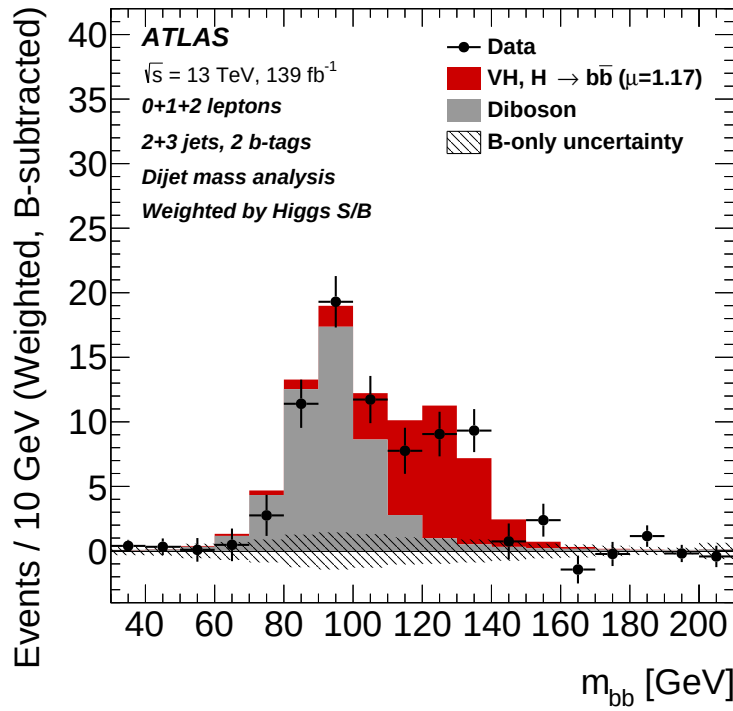


Figure 6.59: The distribution of m_{bb} in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. All lepton channels, p_T^V regions and jet categories are weighted by their respective S/B ratio, with S (B) being the signal (background) fitted yield in that region, and then summed together. The expected contribution of the WH and ZH processes is shown scaled by the signal strength ($\mu = 1.17$) measured in the di-jet mass analysis. The size of the combined statistical and systematic uncertainty for the fitted background is indicated by the hatched band [146].

Conclusions

The observation of the dominant Higgs boson decay mode to a pair of b -quarks in 2018 represented a further validation of the Standard Model (SM) predictions. At the same time, it raised the percentage of observed Higgs boson decays to $\sim 88\%$, thus leaving very little space for new physics beyond the SM in the Higgs sector. However, deviations are still possible in the rate at which the Higgs bosons are produced at high transverse momentum, where the sensitivity to new physics processes is higher.

This thesis described the measurement of the Higgs boson decaying to a pair of b -quarks produced in association with a W or Z boson, whose results were published in Ref. [146]. The analysis targets the leptonic decays of W and Z in order to suppress the large background from multi-jet production, and it is performed in the high p_T regime in order to have a better signal over background ratio. Several SM backgrounds, arising from top quark and vector boson production, make this measurement very challenging.

The main handle to discriminate between the signal from the sum of all backgrounds is the invariant mass, m_{bb} , of the two b -jets in the final state. The analysis performed using this single variable as discriminant is referred to as *di-jet* mass analysis. This is used as a cross-check to a multivariate technique using a Boosted Decision Tree (BDT) to enhance the signal over background ratio by making use of other kinematic variables that show significant differences between the signal and the backgrounds. Two sets of BDT are constructed, the nominal BDT_{VH} is used to separate the $VH, H \rightarrow b\bar{b}$ signal from the sum of the expected backgrounds, while the second one (BDT_{VZ}) is used to validate the analysis strategy and the background modelling of the main analysis by extracting the $VZ, Z \rightarrow b\bar{b}$ process. In both cases, the output distributions of the multivariate analyses are used as discriminants to a binned maximum-likelihood fit which allows the signal yield and the background normalisations to be extracted.

The associated production of a W boson with jets, W +jets, represents one of the main backgrounds of this analysis, contributing mostly in the 1-lepton channel. In the $H \rightarrow b\bar{b}$ observation analysis, the shape uncertainty associated to the vector boson transverse momentum (p_T^V) distribution of this background was found to have the largest impact on the total uncertainty on the signal strength $\Delta\mu$. Studies performed in the 1-lepton channel demonstrated that using p_T^V as fit discriminant in the control regions reduced the impact of the relative shape systematic on $\Delta\mu$ by $\sim 60\%$, improving the analysis sensitivity in the 1-lepton channel by $\sim 9\%$. It was therefore decided to adopt the p_T^V distribution as fit discriminant in all analysis control regions.

Improvements in the technique used to derive the W +jets shape uncertainties were also presented. Instead of deriving them from the ratio of different MC predictions of the m_{bb} and p_T^V distributions, the two main variables entering BDT_{VH}, a BDT-based

reweighting procedure was employed, making use of all the kinematic variables entering the final BDT_{VH} . This approach was demonstrated to parametrise the analysis phase space in a coherent and correlated fashion and to capture the differences between the generators' predictions in almost all corners of the phase space.

The best fit value of the signal strength of $\mu_{VH}^{b\bar{b}} = 1.02_{-0.11}^{+0.12}(\text{stat.})_{-0.13}^{+0.14}(\text{syst.})$ was measured with an observed (expected) significance of 6.7σ (6.7σ). The production of the Higgs boson in association with a Z or W boson, separately, is measured with an observed (expected) significance of 5.3 (5.1) and 4.0 (4.1) standard deviations, respectively, allowing to claim the observation of ZH production and evidence for WH production.

The cross section is measured as a function of the vector boson transverse momentum in the context of the Simplified Template Cross Section framework. This maximises the sensitivity of the measurements while minimising the dependence on the theoretical prediction and, at the same time, provides a common framework for the combination of measurements in different decay channels, as well as among different experiments. The measured values of the differential cross section were found to be all consistent with the Standard Model expectations.

As in the $VH, H \rightarrow b\bar{b}$ analysis, many other analyses utilize b -quarks in the final state. Therefore, the correct identification of the jets originating from b -quarks represents a fundamental task for the ATLAS experiment. In this thesis, two topics were discussed: the studies of the effect of the b - and c -hadron interactions with the ATLAS pixel detector material and the development of a tagging algorithm for very energetic b -jets.

The current version of the GEANT4 software used in the ATLAS simulation does not take into account the direct interaction of b - and c -hadrons with the pixel detector material, but only that of their decay products. A b -hadron with a high transverse momentum is likely to decay after one or more pixel layers, possibly interacting with the material and leaving a hit before decaying. This hit could be mistakenly associated to the track of one of the b -hadron decay products, causing a distortion of the track itself. The size of the impact of this effect was evaluated on b -tagging related variables and on the b -tagging efficiency at high p_T . Some of these studies have been published in Ref. [141]. Furthermore, extrapolation uncertainties on the b -tagging efficiency have been derived as a function of the jet p_T .

The b -tagging efficiency is expected to significantly decrease for very energetic b -hadrons decaying far from the primary vertex. Primary because the track reconstruction efficiency decreases as a function of the b -hadron transverse decay length, but also because the decay products of energetic b -hadrons are very collimated, resulting in an ambiguous hit-to-track association. Profiting from the high spatial resolution of the pixel detector, a small cone around the jet axis can be used to count the hits in between pixel layers. Preliminary studies of the performance of this new algorithm have been presented but more detailed investigation will be needed to assess if this new approach can be used to complement the track-based b -tagging algorithms in order to improve high p_T performances.

Contributions

Most of the material presented in this thesis is based on work performed within the ATLAS Collaboration. Thus, it is typically created from the contributions of many people and cannot be attributed to a single person. The figures and plots containing an ATLAS or ATLAS Preliminary label were created in the context of ATLAS papers or conference notes. All plots without dedicated ATLAS or ATLAS Preliminary labels are private work done by the author of this thesis or, when indicated in the caption, come from private communication with members of the $VH, H \rightarrow b\bar{b}$ analysis. The publications with a significant contribution from the author of this thesis and the details about these contributions are indicated in this section.

- **Studies of the b -hadron interactions with the pixel detector material:**

Debugging of the first version of the quasi-stable simulation ever implemented, in close collaboration with the ATLAS Simulation Working Group [143] and the ATLAS Flavour Tagging Working Group [142].

Studies on the effect of the quasi-stable simulation on b -tagging related variables and on the performance of both low-level and high-level taggers in the estimation of the b -tagging efficiency for all official b -tagging working points. Some of these studies have been published in Ref. [141].

Estimation of extrapolation uncertainties on the b -tagging efficiency evaluated for the DL1r tagger as a function of the jet p_T through a comparison between the ATLAS standard simulation and the quasi-stable one.

- **Trackless b -tagging for very energetic jets:**

Implementation of the trackless b -tagging related variables in the official framework of the Flavour Tagging Working Group.

Studies related to the idea of building an algorithm based not only on the information on the number of hits in the pixel layers, but also on the angular distribution of the hits inside the defined pixel cone.

Estimation of the b -tagging efficiency using the “*max_trackless*” discriminant and first attempt of evaluating its improvement in tagging high p_T jets with respect to the track-based b -tagging algorithms.

Studies of the impact of b -hadron interactions with the pixel detector on the trackless b -tagging algorithm performance.

- **Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector:**

Modelling of the W +jets background with the estimation of normalization and acceptance uncertainties, and the definition of shape uncertainties using a 1-

dimensional approach. The results of this analysis have been published in Ref. [13]. Exploit W -charge asymmetry to test different W +jets models with higher accuracy. These studies are reported in Appendix B.6.

- **Measurement of the $H \rightarrow b\bar{b}$ decay in association with a Vector Boson:**

Editor of the internal note “Data/Monte Carlo Modelling Studies for the Standard Model $VH, H \rightarrow b\bar{b}$ Analysis” for the full Run 2 publication.

Studies on the implementation of the p_T^V distribution as fit discriminant in all control regions of the analysis.

Modelling of the W +jets background with the estimation of normalization and acceptance uncertainties, and the implementation of a novel approach to estimate the shape uncertainties using a BDT-based reweighting procedure.

Studies on the treatment of the nuisance parameters associated to the modelling uncertainties related to the W +jets and $t\bar{t}$ backgrounds in the fit model.

Derivation of all fit results shown in this thesis, with the exception of the STXS cross section measurement, for the main $VH, H \rightarrow b\bar{b}$ analysis, as well as the cross-check $VZ, Z \rightarrow b\bar{b}$ and di-jet mass analyses.

The results of this analysis have been published in Ref. [146].

Bibliography

- [1] S. L. Glashow, *Partial-symmetries of weak interactions*, [Nuclear Physics](#) **22** (1961) 579 – 588.
- [2] S. Weinberg, *A Model of Leptons*, [Phys. Rev. Lett.](#) **19** (1967) 1264–1266, <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [3] A. Salam, *Weak and Electromagnetic Interactions*, [Conf. Proc. C](#) **680519** (1968) 367–377.
- [4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, [Phys. Rev. Lett.](#) **13** (1964) 321–323.
- [5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, [Phys. Rev. Lett.](#) **13** (1964) 508–509.
- [6] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, [Phys. Lett. B](#) **716** (2012) 1, [arXiv:1207.7214 \[hep-ex\]](#).
- [7] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, [Phys. Lett. B](#) **716** (2012) 30, [arXiv:1207.7235 \[hep-ex\]](#).
- [8] ATLAS and CMS Collaboration, *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** (2016) 045, [arXiv:1606.02266 \[hep-ex\]](#).
- [9] 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** (2018) 173–191, [arXiv:1806.00425 \[hep-ex\]](#).
- [10] CMS Collaboration, *Observation of $t\bar{t}H$ Production*, [Phys. Rev. Lett.](#) **120** (2018) 231801, <https://link.aps.org/doi/10.1103/PhysRevLett.120.231801>.
- [11] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension*, [Comput. Phys. Commun.](#) **108** (1998) 56–74, [arXiv:hep-ph/9704448 \[hep-ph\]](#).

- [12] D. Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, R. Tanaka, R. Aggleton, M. Ahmad, B. Allanach, C. Anastasiou, W. Astill, S. Badger, M. Badziak, J. Baglio, E. Bagnaschi, A. Ballestrero, and X. Chen, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, <https://arxiv.org/abs/1610.07922>.
- [13] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector*, *Physics Letters B* **786** (2018) 59 – 86, <http://www.sciencedirect.com/science/article/pii/S0370269318307056>.
- [14] CMS Collaboration, *Observation of Higgs boson decay to bottom quarks*, *Phys. Rev. Lett.* **121** (2018) 121801, [arXiv:1808.08242](https://arxiv.org/abs/1808.08242) [hep-ex].
- [15] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, [arXiv:1005.4568](https://arxiv.org/abs/1005.4568) [physics.ins-det].
- [16] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, *Nucl.Instrum.Meth.* **A506** (2003) 250–303.
- [17] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ -GeV*, *Phys. Lett. B* **122** (1983) 103–116.
- [18] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c² at the CERN SPS Collider*, *Phys. Lett. B* **126** (1983) 398–410.
- [19] CDF Collaboration, *Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab*, *Phys. Rev. Lett.* **74** (1995) 2626–2631, <https://link.aps.org/doi/10.1103/PhysRevLett.74.2626>.
- [20] D0 Collaboration, *Observation of the Top Quark*, *Phys. Rev. Lett.* **74** (1995) 2632–2637, <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632>.
- [21] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. <http://www.slac.stanford.edu/~mpeskin/QFT.html>.
- [22] F. Halzen, A. D. Martin, and N. Mitra, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*, vol. 53. 1985. <https://doi.org/10.1119/1.14146>.
- [23] M. Kaku, *Quantum Field Theory: A Modern Introduction*, vol. 71. 1994. <https://doi.org/10.1021/ed071pA54.2>.
- [24] Y. Nagashima, *Elementary Particle Physics: Quantum Field Theory and Particles*. John Wiley and Sons, Ltd, 2010. <https://onlinelibrary.wiley.com/doi/book/10.1002/9783527630097>.

- [25] A. Buckley et al., *General-purpose event generators for LHC physics*, *Physics Reports* **504** (2011) 145–233, <http://dx.doi.org/10.1016/j.physrep.2011.03.005>.
- [26] J. M. Campbell, J. Huston, and W. Stirling, *Hard Interactions of Quarks and Gluons: A Primer for LHC Physics*, *Rept. Prog. Phys.* **70** (2007) 89, [arXiv:hep-ph/0611148](https://arxiv.org/abs/hep-ph/0611148).
- [27] A. Buckley et al., *Monte Carlo event generators for high energy particle physics event simulation*, 2019. <https://arxiv.org/abs/1902.01674>.
- [28] R. Feynman, *Quantum Electrodynamics*. Advanced Books Classics. Avalon Publishing, 1998. <https://books.google.fr/books?id=xt-Vvhloo8YC>.
- [29] J. C. Maxwell, *A dynamical theory of the electromagnetic field*, Philosophical Transactions of the Royal Society of London **155** (1865), <https://royalsocietypublishing.org/doi/pdf/10.1098/rstl.1865.0008>.
- [30] S. Mandelstam, *Feynman Rules for Electromagnetic and Yang-Mills Fields from the Gauge-Independent Field-Theoretic Formalism*, *Phys. Rev.* **175** (1968) 1580–1603, <https://link.aps.org/doi/10.1103/PhysRev.175.1580>.
- [31] G. Sardanashvily, *Noether's Theorems: Applications in Mechanics and Field Theory*, 2016.
- [32] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, *Phys. Rev. Lett.* **10** (1963) 531–533.
- [33] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, *Prog. Theor. Phys.* **49** (1973) 652–657.
- [34] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567, <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [35] V. Lobashev, *The search for the neutrino mass by direct method in the tritium beta-decay and perspectives of study it in the project KATRIN*, *Nucl. Phys. A* **719** (2003) 153–160.
- [36] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, *Sov. Phys. JETP* **7** (1958) 172–173.
- [37] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, *Prog. Theor. Phys.* **28** (1962) 870–880.
- [38] CERN Webfest 2012, *Standard Model infographic developed at the webfest*, 2014. <https://texample.net/tikz/examples/model-physics/>.

- [39] Y. Kim and M. Noz, *Theory and applications of the Poincaré group*, vol. 17. Reidel, Netherlands, 1986.
- [40] E. Fermi, *Tentativo di una Teoria Dei Raggi β* , *Il Nuovo Cimento* (1924-1942) **11** (1934) 1–19.
- [41] F. Hasert et al., *Search for Elastic ν_μ Electron Scattering*, *Phys. Lett. B* **46** (1973) 121–124.
- [42] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental Test of Parity Conservation in Beta Decay*, *Phys. Rev.* **105** (1957) 1413–1415, <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.
- [43] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, *Nucl. Phys. B* **44**: No. 1, 189–213(1972). (1972).
- [44] D. Azevedo, *Probing the CP nature of the Higgs' couplings in ttH events at the LHC*, 2017. <https://www.semanticscholar.org/paper/Probing-the-CP-nature-of-the-Higgs'-couplings-in-at-Azevedo/2c0382b9c5f5a337917134caaf4568892ce3092b>.
- [45] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev. D* **98** (2018) 030001, <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [46] LHC Higgs Cross Section Working Group. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [47] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, *Physical Review Letters* **114** (2015), <http://dx.doi.org/10.1103/PhysRevLett.114.191803>.
- [48] P. A. M. Dirac, *The Quantum Theory of the Emission and Absorption of Radiation*, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **114** (1927) 243–265, <http://www.jstor.org/stable/94746>.
- [49] E. Fermi, *Nuclear Physics*. The University of Chicago Press, 1950. <https://press.uchicago.edu/ucp/books/book/chicago/N/bo3631242.html>.
- [50] A. De Roeck and R. Thorne, *Structure functions*, *Progress in Particle and Nuclear Physics* **66** (2011) 727–781, <http://dx.doi.org/10.1016/j.ppnp.2011.06.001>.

- [51] H1 and ZEUS Collaborations, *Combination of Measurements of Inclusive Deep Inelastic $e^\pm p$ Scattering Cross Sections and QCD Analysis of HERA Data*, 2015.
<https://link.springer.com/article/10.1140/epjc/s10052-015-3710-4>.
- [52] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nuclear Physics B* **126** (1977) 298 – 318, <http://www.sciencedirect.com/science/article/pii/0550321377903844>.
- [53] R. D. Ball, V. Bertone, S. Carrazza, L. D. Debbio, S. Forte, P. Groth-Merrild, A. Guffanti, N. P. Hartland, Z. Kassabov, J. Latorre, E. R. Nocera, J. Rojo, L. Rottoli, E. Slade, and M. Ubiali, *Parton distributions from high-precision collider data*, *The European Physical Journal C* **77** (2017) 663,
<https://doi.org/10.1140/epjc/s10052-017-5199-5>.
- [54] *Proton-proton collision Scheme*,
<https://imperialhep.blogspot.com/2011/08/>.
- [55] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, *New parton distribution functions from a global analysis of quantum chromodynamics*, *Physical Review D* **93** (2016),
<http://dx.doi.org/10.1103/PhysRevD.93.033006>.
- [56] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849](https://arxiv.org/abs/1410.8849) [hep-ph].
- [57] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *The European Physical Journal C* **75** (2015), <http://dx.doi.org/10.1140/epjc/s10052-015-3397-6>.
- [58] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43** (2016) 023001, [arXiv:1510.03865](https://arxiv.org/abs/1510.03865) [hep-ph].
- [59] J. C. Collins, *Sudakov Form Factors*, 2003.
<https://arxiv.org/abs/hep-ph/0312336>.
- [60] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196, [arXiv:1512.01178](https://arxiv.org/abs/1512.01178) [hep-ph].
- [61] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Computer Physics Communications* **191** (2015) 159–177,
<http://dx.doi.org/10.1016/j.cpc.2015.01.024>.
- [62] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31–145.

- [63] D. Amati and G. Veneziano, *Preconfinement as a Property of Perturbative QCD*, *Phys. Lett. B* **83** (1979) 87–92.
- [64] Sherpa Collaboration, E. Bothmann et al., *Event Generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [arXiv:1905.09127](https://arxiv.org/abs/1905.09127) [hep-ph].
- [65] D. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152–155.
- [66] S. Jadach, J. H. Kuhn, and Z. Was, *TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons*, *Comput. Phys. Commun.* **64** (1990) 275–299.
- [67] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) S08001.
- [68] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. <https://cds.cern.ch/record/782076>.
- [69] D. H. Perkins, *Introduction to high energy physics*. Cambridge Univ. Press, 2000.
- [70] E. Mobs, *The CERN accelerator complex - August 2018*. *Complexe des accélérateurs du CERN - Août 2018*, <https://cds.cern.ch/record/2636343>, General Photo.
- [71] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, <https://cds.cern.ch/record/2668326>.
- [72] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [73] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [74] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.
- [75] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.
- [76] Luminosity Public Results Run 2. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_201.
- [77] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. <https://cds.cern.ch/record/338080>.
- [78] ATLAS Magnet System Webpage. <http://atlas-magnet-dfs.web.cern.ch>.

- [79] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997.
<https://cds.cern.ch/record/331063>.
- [80] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run-2*, 2020.
- [81] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Atlas-tdr-19, 2010, <https://cds.cern.ch/record/1291633>.
- [82] C. Da Via, M. Boscardin, G.-F. Dalla Betta, G. Darbo, C. Fleta, C. Gemme, P. Grenier, S. Grinstein, T.-E. Hansen, J. Hasi, C. Kenney, A. Kok, S. Parker, G. Pellegrini, E. Vianello, and N. Zorzi, *3D silicon sensors: Design, large area production and quality assurance for the ATLAS IBL pixel detector upgrade*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **694** (2012) 321 – 330, <http://www.sciencedirect.com/science/article/pii/S0168900212008509>.
- [83] ATLAS Collaboration, *IBL Efficiency and Single Point Resolution in Collision Events*, Tech. Rep. ATL-INDET-PUB-2016-001, CERN, Geneva, Aug, 2016.
<https://cds.cern.ch/record/2203893>.
- [84] J. R. Pater, *The ATLAS SemiConductor Tracker operation and performance*, *Journal of Instrumentation* **7** (2012) C04001–C04001,
<https://doi.org/10.1088%2F1748-0221%2F7%2F04%2Fc04001>.
- [85] A. Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, Tech. Rep. ATL-INDET-PROC-2013-005, CERN, Geneva, Apr, 2013.
<https://cds.cern.ch/record/1537991>.
- [86] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, *Journal of Instrumentation* **12** (2017) P05002–P05002,
<http://dx.doi.org/10.1088/1748-0221/12/05/P05002>.
- [87] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996.
<http://cds.cern.ch/record/331061>.
- [88] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, *Eur. Phys. J. C* **70** (2010) 1193, [arXiv:1007.5423](https://arxiv.org/abs/1007.5423) [hep-ex].
- [89] ATLAS Collaboration, *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*, *Eur. Phys. J. C* **70** (2010) 723–753, [arXiv:0912.2642](https://arxiv.org/abs/0912.2642) [physics.ins-det].
- [90] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*.

- [91] T. Barillari, *The ATLAS liquid argon hadronic end-cap calorimeter: construction and selected beam test results*, *Nuclear Physics B - Proceedings Supplements* **150** (2006) 102 – 105, <http://www.sciencedirect.com/science/article/pii/S0920563205008194>, Proceedings of the 9th Topical Seminar on Innovative Particle and Radiation Detectors.
- [92] D. Gillberg and the ATLAS Liquid Argon Calorimeter Group, *Performance of the ATLAS Forward Calorimeters in First LHC Data*, *Journal of Physics: Conference Series* **293** (2011) 012041, <https://doi.org/10.1088%2F1742-6596%2F293%2F1%2F012041>.
- [93] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. <https://cds.cern.ch/record/331068>.
- [94] M. Mikuz, V. Cindro, I. Dolenc, H. Kagan, G. Kramberger, H. Fraiss-Kolbl, A. Gorisek, E. Griesmayer, B. Macek, I. Mandic, M. Niegl, H. Pernegger, D. Smith, D. Tardif, W. Trischuk, P. Weilhammer, and M. Zavrtanik, *The ATLAS beam conditions monitor*, *Journal of Instrumentation* **3** (2008) 1914 – 1917.
- [95] ATLAS Collaboration, *Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System*, Tech. Rep. CERN-LHCC-2013-018. ATLAS-TDR-023, Sep, 2013. <https://cds.cern.ch/record/1602235>.
- [96] ATLAS Collaboration, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003. <https://cds.cern.ch/record/616089>.
- [97] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *The European Physical Journal C* **77** (2017), <http://dx.doi.org/10.1140/epjc/s10052-017-5225-7>.
- [98] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444–450.
- [99] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-018, 2015, <https://cds.cern.ch/record/2037683>.
- [100] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, <https://cds.cern.ch/record/1449796>.
- [101] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332, [arXiv:1611.10235](https://arxiv.org/abs/1611.10235) [physics.ins-det].

- [102] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, Tech. Rep. ATL-PHYS-PUB-2015-026, CERN, Geneva, Jul, 2015. <https://cds.cern.ch/record/2037717>.
- [103] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys. G* **34** (2007) N343.
- [104] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, ATL-PHYS-PUB-2017-022, 2017, <https://cds.cern.ch/record/2298955>.
- [105] D. Scott, *Multivariate density estimation: Theory, practice, and visualization: Second edition*, *Multivariate Density Estimation: Theory, Practice, and Visualization: Second Edition* (2015) 1–360.
- [106] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, *Journal of Instrumentation* **14** (2019) P12006–P12006, <https://doi.org/10.1088%2F1748-0221%2F14%2F12%2Fp12006>.
- [107] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Physics Letters B* **659** (2008) 119–126, <http://dx.doi.org/10.1016/j.physletb.2007.09.077>.
- [108] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) 292, [arXiv:1603.05598](https://arxiv.org/abs/1603.05598) [hep-ex].
- [109] J. Illingworth and J. Kittler, *A survey of the hough transform*, *Computer Vision, Graphics, and Image Processing* **44** (1988) 87 – 116, <http://www.sciencedirect.com/science/article/pii/S0734189X88800331>.
- [110] J. J. Junggeburth, *Muon identification and reconstruction efficiencies in full Run-2 dataset*, Tech. Rep. ATL-COM-PHYS-2019-1001, CERN, Geneva, Aug, 2019. <https://cds.cern.ch/record/2685295>.
- [111] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, <https://cds.cern.ch/record/2064383>.
- [112] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al., *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *The European Physical Journal C* **77** (2017), <http://dx.doi.org/10.1140/epjc/s10052-017-5031-2>.

- [113] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063–063, <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [114] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. Seymour, and B. Webber, *HERWIG 6.5 Release Note*, 2002.
- [115] ATLAS Collaboration, *Jet energy scale and uncertainties in 2015-2017 data and simulation*, 2018. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2018-006/>.
- [116] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, 2019.
- [117] ATLAS Collaboration, *Jet energy resolution in 2017 data and simulation*, 2018. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2018-005/>.
- [118] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, <https://cds.cern.ch/record/1700870>.
- [119] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **78** (2018) 903, [arXiv:1802.08168](https://arxiv.org/abs/1802.08168) [hep-ex].
- [120] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, 2018, <https://cds.cern.ch/record/2625233>.
- [121] D. Adams et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, Tech. Rep. ATL-PHYS-INT-2014-018, CERN, Geneva, Jul, 2014. <https://cds.cern.ch/record/1743654>.
- [122] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, <https://cds.cern.ch/record/2270366>.
- [123] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics.*, CERN-OPEN-2008-020, (December, 2018), <http://arxiv.org/abs/0901.0512>.
- [124] ATLAS Collaboration, *Performance of b -jet identification in the ATLAS experiment*, *Journal of Instrumentation* **11** (2016) P04008–P04008, <https://doi.org/10.1088%2F1748-0221%2F11%2F04%2Fp04008>.

- [125] ATLAS Collaboration, *Expected performance of the ATLAS b -tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015,
<https://cds.cern.ch/record/2037697>.
- [126] ATLAS Collaboration, *Commissioning of the ATLAS high performance b -tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102, 2011,
<https://cds.cern.ch/record/1369219>.
- [127] ATLAS Collaboration, *Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016,
<https://cds.cern.ch/record/2160731>.
- [128] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017,
<https://cds.cern.ch/record/2255226>.
- [129] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, *Studies in Computational Intelligence* 385 (2012).
- [130] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b -jet tagging in ATLAS*, *Journal of Physics: Conference Series* **119** (2008) 032032,
<https://doi.org/10.1088%2F1742-6596%2F119%2F3%2F032032>.
- [131] P. Billoir and S. Qian, *Simultaneous pattern recognition and track fitting by the Kalman filtering method*, *Nucl. Instrum. Meth. A* **294** (1990) 219–228.
- [132] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017,
<https://cds.cern.ch/record/2273281>.
- [133] A. Hoecker et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007.
<https://arxiv.org/abs/physics/0703039>.
- [134] F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- [135] T. Team, *Theano: A Python framework for fast computation of mathematical expressions*, [arXiv:1605](https://arxiv.org/abs/1605).
- [136] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014.
<https://arxiv.org/abs/1412.6980>.
- [137] ATLAS Collaboration, *ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C* **79** (2019),
<http://dx.doi.org/10.1140/epjc/s10052-019-7450-8>.

- [138] ATLAS Collaboration, *Measurement of b -tagging efficiency of c -jets in $t\bar{t}$ events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001, 2018, <https://cds.cern.ch/record/2306649>.
- [139] J. Erdmann, S. Guindon, K. Kröninger, B. Lemmer, O. Nackenhorst, A. Quadt, and P. Stolte, *A likelihood-based reconstruction algorithm for top-quark pairs and the KLfitter framework*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **748** (2014) 18 – 25, <http://www.sciencedirect.com/science/article/pii/S0168900214001855>.
- [140] 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, <https://cds.cern.ch/record/2314418>.
- [141] ATLAS Collaboration, *Impact of ATLAS quasi-stable particle simulation on quantities related to flavour tagging and tau reconstruction*, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2020-002/>.
- [142] ATLAS Flavour Tagging Working Group. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/FlavourTagging>.
- [143] ATLAS Simulation Working Group. <https://twiki.cern.ch/twiki/bin/view/AtlasComputing/AtlasSimulation>.
- [144] B. T. Huffman, C. Jackson, and J. Tseng, *Tagging b -quarks at extreme energies without tracks*, *Journal of Physics G: Nuclear and Particle Physics* **43** (2016) 085001, <http://dx.doi.org/10.1088/0954-3899/43/8/085001>.
- [145] E. Perez Codina and P. G. Roloff, *Hit multiplicity approach to b -tagging in FCC-hh*, Tech. Rep. CERN-ACC-2018-0023, CERN, Geneva, Jul, 2018. <https://cds.cern.ch/record/2631478>.
- [146] 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*, [arXiv:2007.02873](https://arxiv.org/abs/2007.02873) [hep-ex], <https://arxiv.org/pdf/2007.02873>.
- [147] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [arXiv:1002.2581](https://arxiv.org/abs/1002.2581) [hep-ph].
- [148] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola, T. Reiter, and F. Tramontano, *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) 1889, [arXiv:1111.2034](https://arxiv.org/abs/1111.2034) [hep-ph].
- [149] K. Hamilton, P. Nason, and G. Zanderighi, *MINLO: multi-scale improved NLO*, *JHEP* **10** (2012) 155, [arXiv:1206.3572](https://arxiv.org/abs/1206.3572) [hep-ph].

- [150] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, *HW[±]/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) 083, [arXiv:1306.2542 \[hep-ph\]](#).
- [151] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820 \[hep-ph\]](#).
- [152] 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** (2014) 145, [arXiv:1406.3660 \[hep-ex\]](#).
- [153] M. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003) 073003, [arXiv:hep-ph/0306234 \[hep-ph\]](#).
- [154] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) 149–156, [arXiv:hep-ph/0307206 \[hep-ph\]](#).
- [155] G. Ferrera, M. Grazzini, and F. Tramontano, *Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* **107** (2011) 152003, [arXiv:1107.1164 \[hep-ph\]](#).
- [156] O. Brein, R. Harlander, M. Wiesemann, and T. Zirke, *Top-quark mediated effects in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) 1868, [arXiv:1111.0761 \[hep-ph\]](#).
- [157] G. Ferrera, M. Grazzini, and F. Tramontano, *Higher-order QCD effects for associated WH production and decay at the LHC*, *JHEP* **04** (2014) 039, [arXiv:1312.1669 \[hep-ph\]](#).
- [158] G. Ferrera, M. Grazzini, and F. Tramontano, *Associated ZH production at hadron colliders: the fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015) 51–55, [arXiv:1407.4747 \[hep-ph\]](#).
- [159] J. M. Campbell, R. K. Ellis, and C. Williams, *Associated production of a Higgs boson at NNLO*, *JHEP* **06** (2016) 179, [arXiv:1601.00658 \[hep-ph\]](#).
- [160] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak, and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) 078, [arXiv:1211.5015 \[hep-ph\]](#).
- [161] B. Hespel, F. Maltoni, and E. Vryonidou, *Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) 065, [arXiv:1503.01656 \[hep-ph\]](#).

- [162] R. V. Harlander, A. Kulesza, V. Theeuwes, and T. Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) 082, [arXiv:1410.0217 \[hep-ph\]](#).
- [163] R. V. Harlander, S. Liebler, and T. Zirke, *Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) 023, [arXiv:1307.8122 \[hep-ph\]](#).
- [164] O. Brein, R. V. Harlander, and T. J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, *Comput. Phys. Commun.* **184** (2013) 998–1003, [arXiv:1210.5347 \[hep-ph\]](#).
- [165] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadron production*, *JHEP* **09** (2007) 126, [arXiv:0707.3088 \[hep-ph\]](#).
- [166] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, <https://cds.cern.ch/record/1966419>.
- [167] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) 2930, [arXiv:1112.5675 \[hep-ph\]](#).
- [168] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111, [arXiv:0907.4076 \[hep-ph\]](#), [Erratum: *JHEP* **02** (2010) 011].
- [169] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, *Phys. Rev. D* **81** (2010) 054028, [arXiv:1001.5034 \[hep-ph\]](#).
- [170] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547, [arXiv:1009.2450 \[hep-ph\]](#).
- [171] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, *Phys. Rev. D* **82** (2010) 054018, [arXiv:1005.4451 \[hep-ph\]](#).
- [172] F. Cascioli, P. Maierhöfer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, [arXiv:1111.5206 \[hep-ph\]](#).
- [173] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, [arXiv:0808.3674 \[hep-ph\]](#).
- [174] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038, [arXiv:0709.1027 \[hep-ph\]](#).
- [175] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) 027, [arXiv:1207.5030 \[hep-ph\]](#).

- [176] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, *Vector Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at Next-to-Next-to-Leading Order*, *Phys. Rev. Lett.* **103** (2009) 082001, [arXiv:0903.2120 \[hep-ph\]](#).
- [177] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B867** (2013) 244–289, [arXiv:1207.1303 \[hep-ph\]](#).
- [178] A. Denner, S. Dittmaier, S. Kallweit, and A. Muck, *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and LHC with HAWK*, *JHEP* **03** (2012) 075, [arXiv:1211.5142 \[hep-ph\]](#).
- [179] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [arXiv:0409146 \[hep-ph\]](#).
- [180] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, P. Richardson, M. Seymour, A. Sherstnev, and B. Webber, *Herwig++ physics and manual*, *The European Physical Journal C* **58** (2008) 639–707, <http://dx.doi.org/10.1140/epjc/s10052-008-0798-9>.
- [181] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301 \[hep-ph\]](#).
- [182] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *Journal of High Energy Physics* **2009** (2009) 007, <http://stacks.iop.org/1126-6708/2009/i=02/a=007>.
- [183] L. Lonnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046, [arXiv:hep-ph/0112284 \[hep-ph\]](#).
- [184] N. Lavesson and L. Lonnblad, *W+jets matrix elements and the dipole cascade*, *JHEP* **07** (2005) 054, [arXiv:hep-ph/0503293 \[hep-ph\]](#).
- [185] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008. <https://cds.cern.ch/record/1099735>.
- [186] A. D. Bukin, *Fitting function for asymmetric peaks*, 2007.
- [187] ATLAS Collaboration, *Auxiliary material to "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"*, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2018-51/>.

- [188] Y. Coadou, *Boosted Decision Trees and Applications*, In: EPJ Web Conf. 55 (2013), p. 02004 .
- [189] J. H. Friedman, *Stochastic gradient boosting*, *Computational Statistics and Data Analysis* **38** (2002) 367 – 378, <http://www.sciencedirect.com/science/article/pii/S0167947301000652>, Nonlinear Methods and Data Mining.
- [190] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, *Data analysis in high energy physics: a practical guide to statistical methods*, <https://cds.cern.ch/record/1517556>.
- [191] ATLAS Collaboration, *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** (2015) 069, [arXiv:1409.6212](https://arxiv.org/abs/1409.6212) [hep-ex].
- [192] D. Gonçalves and J. Nakamura, *Role of the Z polarization in the $H \rightarrow b\bar{b}$ measurement*, *Phys. Rev. D* **98** (2018) 093005, <https://link.aps.org/doi/10.1103/PhysRevD.98.093005>.
- [193] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [arXiv:1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an], [Erratum: *Eur. Phys. J. C* **73** (2013) 2501].
- [194] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys. Commun.* **77** (1993) 219–228.
- [195] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Ann. Math. Statist.* **9** (1938) 60–62, <https://doi.org/10.1214/aoms/1177732360>.
- [196] A. L. Read, *Presentation of search results: the CL_S technique*, *J. Phys. G* **28** (2002) 2693.
- [197] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, *Transactions of the American Mathematical Society* **54** (1943) 426–482, <http://www.jstor.org/stable/1990256>.
- [198] M. Jansen and G. Claeskens, *Cramér–Rao Inequality*, *International Encyclopedia of Statistical Science* (2011) 322–323, https://doi.org/10.1007/978-3-642-04898-2_197.
- [199] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, [arXiv:1506.02169](https://arxiv.org/abs/1506.02169) [stat.AP].
- [200] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN-2011-002 (2011), [arXiv:1101.0593](https://arxiv.org/abs/1101.0593) [hep-ph].

- [201] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN-2012-002 (2012), [arXiv:1201.3084 \[hep-ph\]](#).
- [202] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN-2013-004 (2013), [arXiv:1307.1347 \[hep-ph\]](#).
- [203] ATLAS Collaboration, *Measurement of VH , $H \rightarrow b\bar{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*, JHEP **05** (2019) 141, [arXiv:1903.04618 \[hep-ex\]](#).
- [204] S. Badger et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, 2016. [arXiv:1605.04692 \[hep-ph\]](#). <https://lss.fnal.gov/archive/2016/conf/fermilab-conf-16-175-ppd-t.pdf>.
- [205] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, [arXiv:1906.02754 \[hep-ph\]](#).
- [206] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, [arXiv:1608.03953 \[hep-ex\]](#).
- [207] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.
- [208] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC*, Phys. Rev. Lett. **117** (2016) 182002, [arXiv:1606.02625 \[hep-ex\]](#).
- [209] 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*, submitted to Eur. Phys. J. (2019), [arXiv:1902.04655 \[hep-ex\]](#).
- [210] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002, [arXiv:1703.09665 \[hep-ex\]](#).
- [211] ATLAS Collaboration, *Jet energy resolution in proton–proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C **73** (2013) 2306, [arXiv:1210.6210 \[hep-ex\]](#).
- [212] M. Muškinja, *Simulation for quasistable particles*, https://indico.cern.ch/event/844658/contributions/3546573/attachments/1905249/3146415/miham_2019_09_10.QS-Sim.pdf.

Appendix

A Additional material for the b -hadron interactions with the pixel detector

A.1 The issue of the higher lifetimes of neutral b -hadrons

In the quasi-stable simulation, the neutral b -hadrons have larger lifetimes with respect to the standard simulation. This is due to the fact that during the event generation step, the oscillation vertex of the neutral b -hadron¹⁵ is added immediately before its decay vertex. However, during the quasi-stable simulation, the oscillation vertex is artificially positioned at the middle of the flight path and moved back on top of the decay vertex after the simulation is complete. In this way, GEANT4 does not immediately decay the b -hadron as it should but rather propagates it, artificially increasing its lifetime [212]. A schematic view of this issue is shown in Figure A.60.

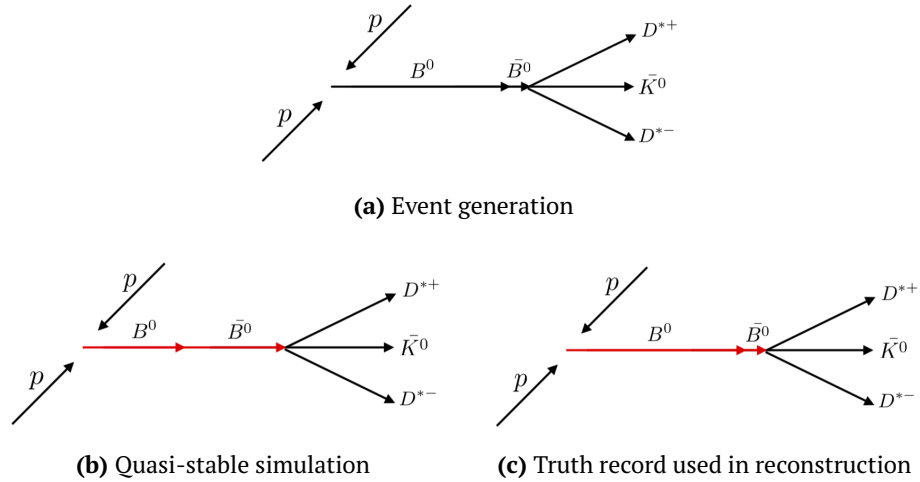


Figure A.60: Figure (a) shows how neutral b -hadrons undergoing oscillation are treated in the event generation in which the oscillation vertex is placed immediately before the decay vertex. During the quasi-stable simulation the oscillation vertex is placed at the middle of the b -hadron flight path, as shown in (b). However, once the simulation is finished, the oscillation vertex is moved back on top of the decay vertex, as in (c), and this truth value of L_{xy} is passed to the digitization and reconstruction step.

To overcome this issue a fix in the GEANT4 software has been implemented to place the oscillation vertex on top of the decay one during and at the end of the quasi-stable simulation step. This way, the truth record of L_{xy} saved in the hits file is expected to be

¹⁵Other long-lived particles (such as Σ^0) are affected by this issue and subjected to the same correction as the b -hadrons.

identical to the one stored in the case of the standard ATLAS simulation. However, a residual $\sim 8\%$ effect was found and it is still under investigation.

A.2 Performance of high-level b -tagging discriminants

Efficiency at the 77% working point of the tagger

Figure A.61 shows the b -tagging efficiency at the 77% WP of the MV2c10 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

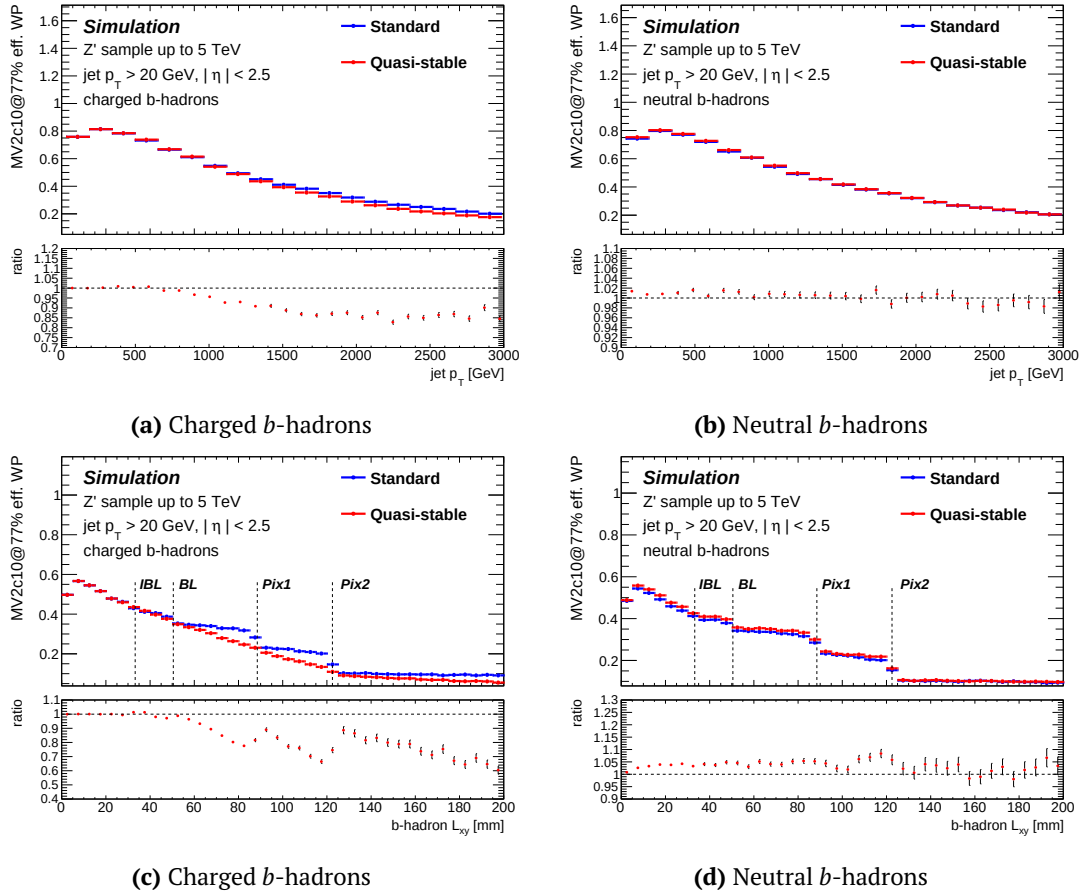


Figure A.61: b -tagging efficiency at the 77% WP of the MV2c10 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

Figure A.62 shows the b -tagging efficiency at the 77% WP of the DL1 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

Figure A.63 shows the b -tagging efficiency at the 77% WP of the DL1r discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

Efficiency at the 85% working point of the tagger

Figure A.64 shows the b -tagging efficiency at the 85% WP of the MV2c10 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral

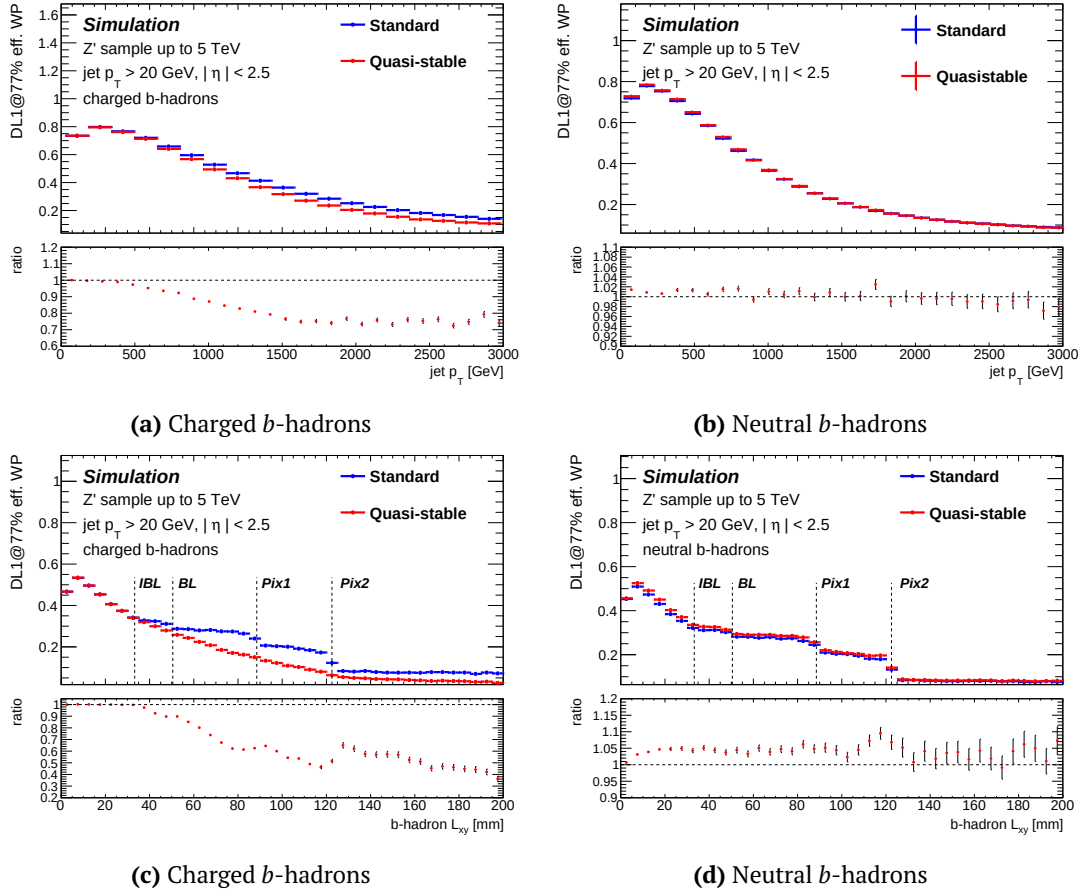


Figure A.62: b -tagging efficiency at the 77% WP of the DL1 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

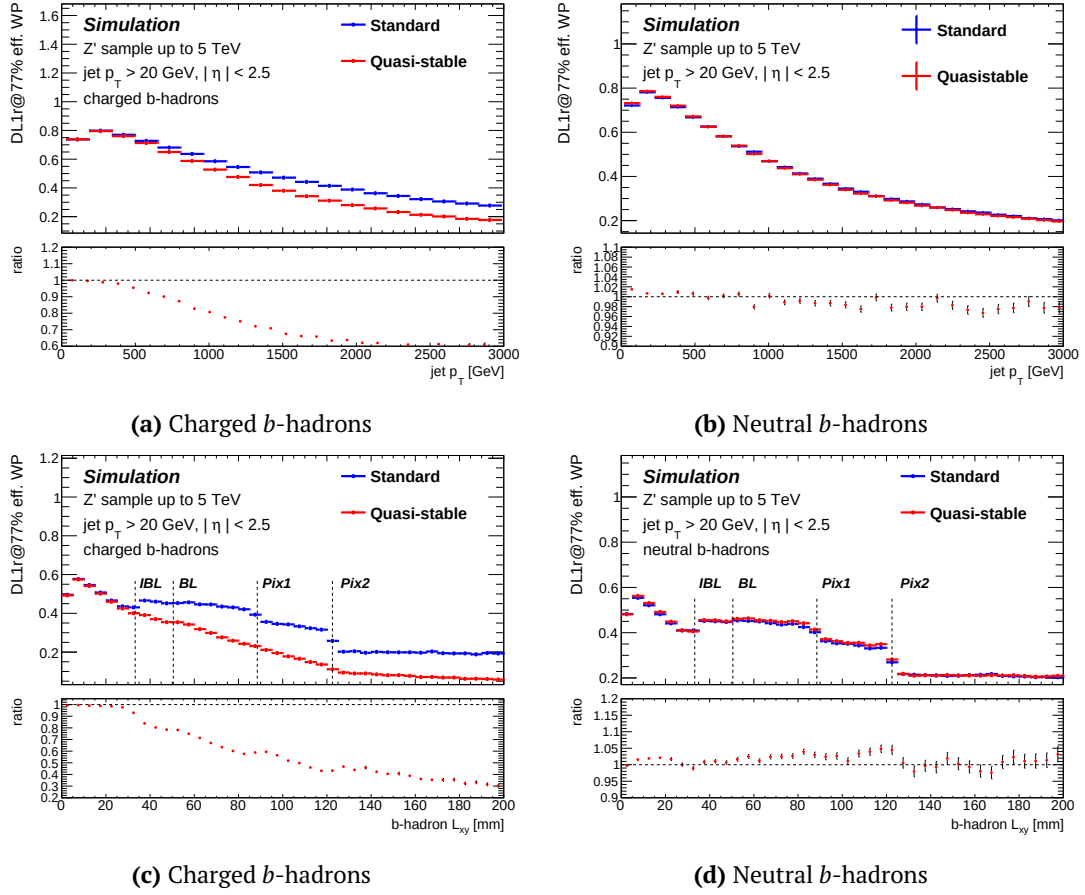


Figure A.63: b -tagging efficiency at the 77% WP of the DL1r tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

b -hadrons.

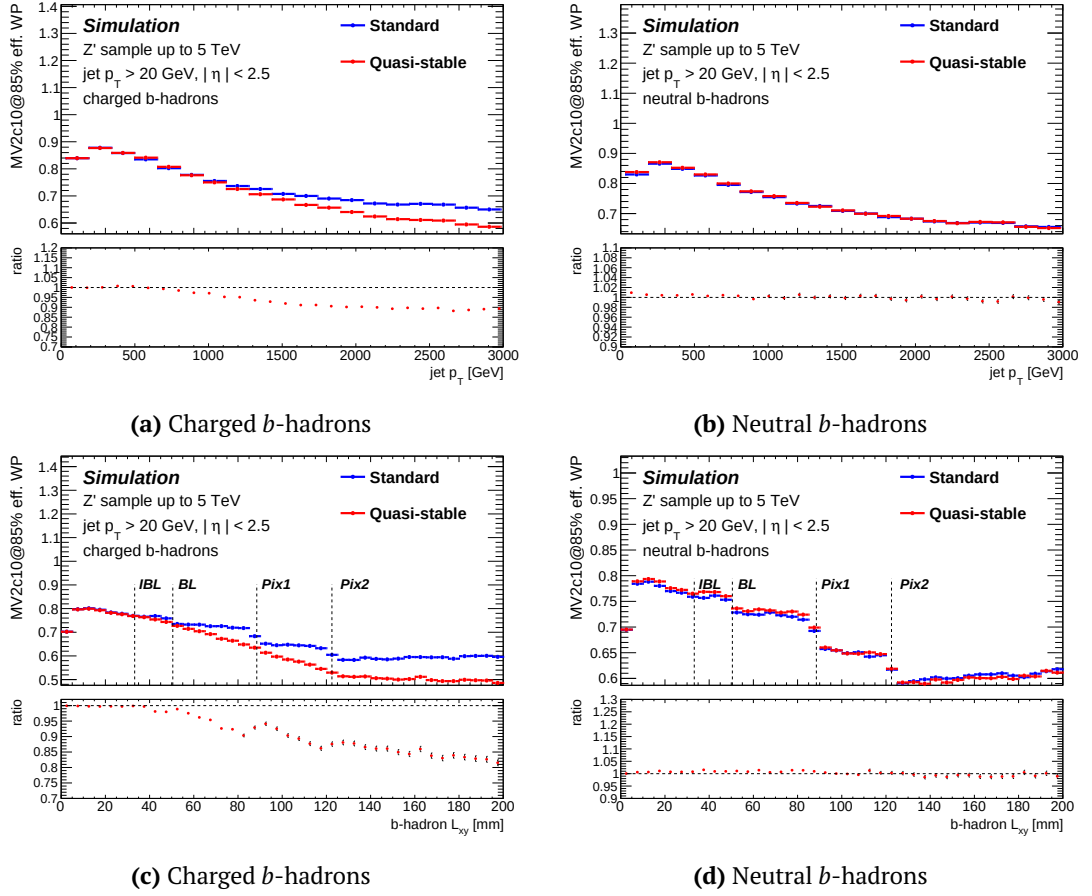


Figure A.64: b -tagging efficiency at the 85% WP of the MV2c10 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

Figure A.65 shows the b -tagging efficiency at the 85% WP of the DL1 discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

Figure A.66 shows the b -tagging efficiency at the 85% WP of the DL1r discriminant as a function of the jet p_T and the b -hadron L_{xy} , separately for charged and neutral b -hadrons.

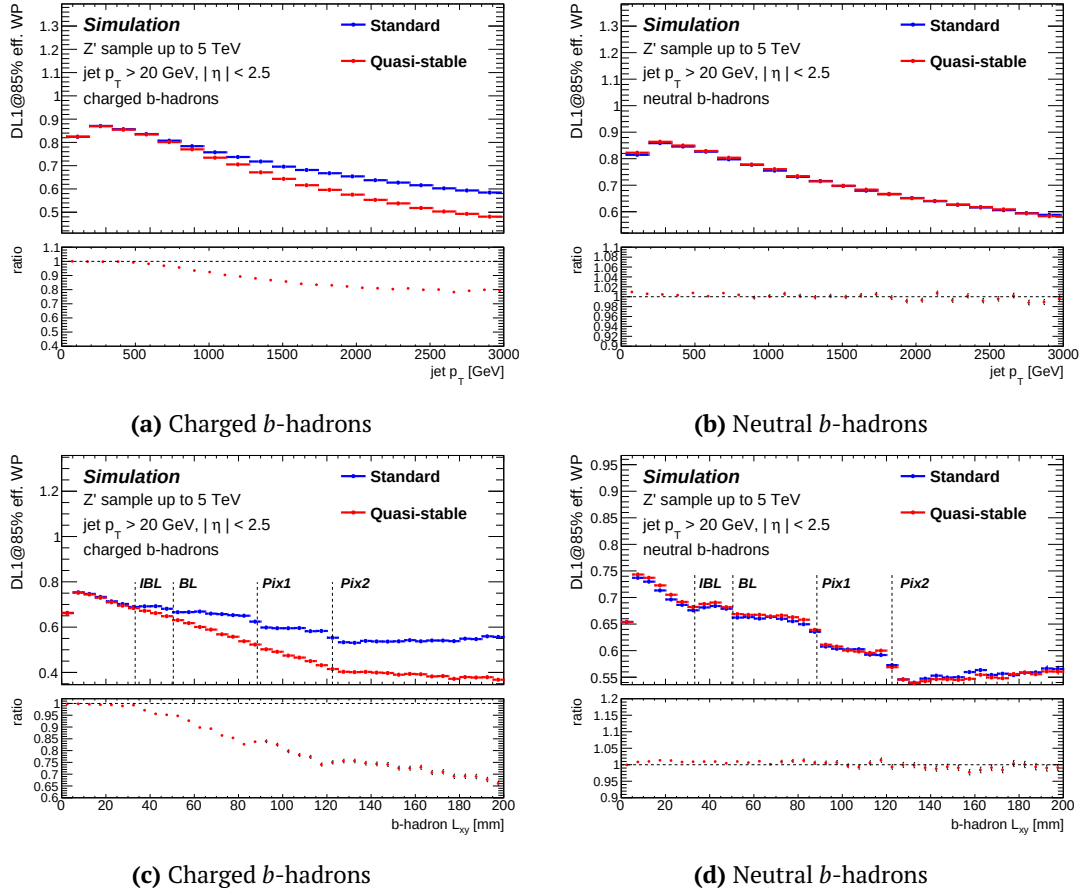


Figure A.65: b -tagging efficiency at the 85% WP of the DL1 tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

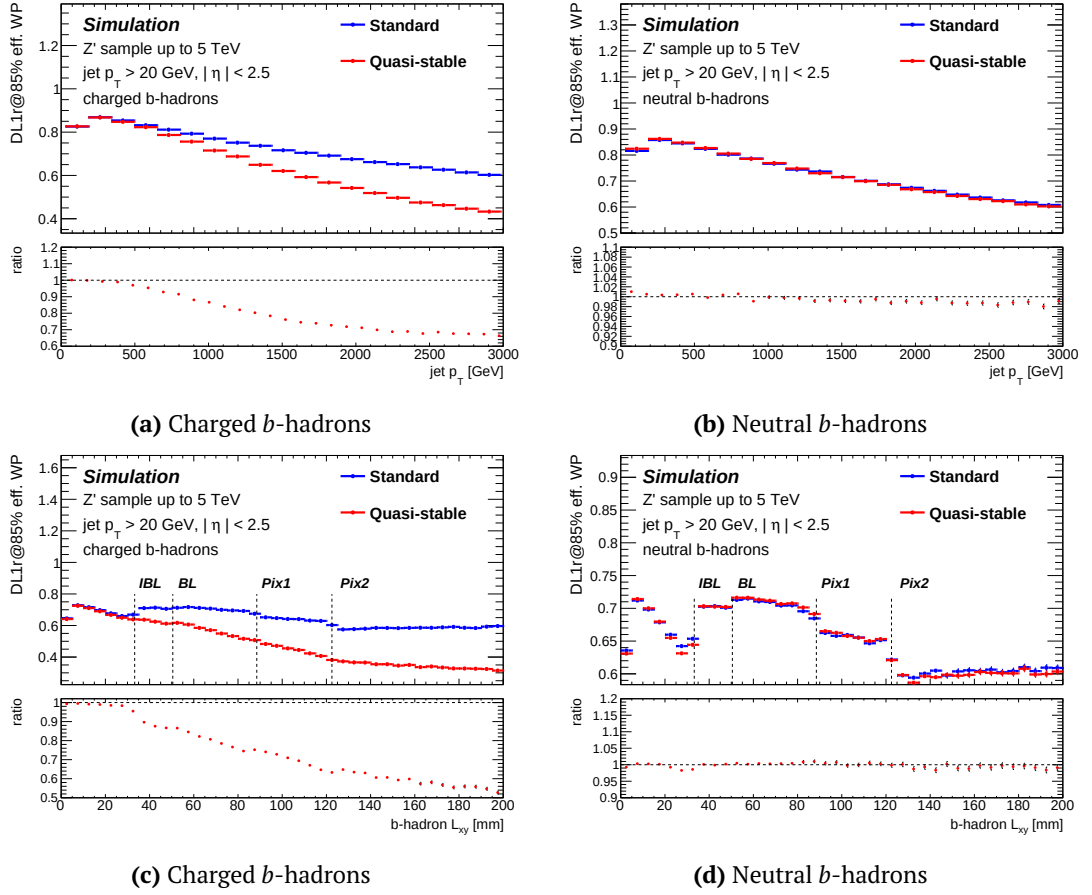


Figure A.66: b -tagging efficiency at the 85% WP of the DL1r tagger, as a function of the jet p_T for charged (a) and neutral (b) b -hadrons, and the b -hadron L_{xy} for charged (c) and neutral (d) b -hadrons.

B Additional material for the W +jets modelling systematics

B.1 Breakdown of extrapolation uncertainties induced by the p_T^V shape uncertainty

Table B.32: Summary of the extrapolation uncertainties induced by the p_T^V shape systematic on the W +jets samples across all the analysis regions of the analysis and in the 0 and 1-lepton channels. The column "weighted-sum" shows the normalization effect induced by the shape systematic weighted by the yields in each of the analysis regions. The raw $\Delta(\text{high-low}) p_T^V$ shows the normalization effect induced by the p_T^V shape systematic across the different p_T^V bins of the analysis. The last two columns show the migration effect, inclusive in the two p_T^V bins induced by the p_T^V shape systematic from Low- ΔR CR to SR and from High- ΔR CR to SR.

0-lepton						
2-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum $\Delta(\text{Low-}\Delta R \text{ CR} - \text{SR})$	weighted-sum $\Delta(\text{High-}\Delta R \text{ CR} - \text{SR})$
150-250 p_T^V	5.08	5.49	5.83	5.44	-1.64	-1.64
250 p_T^V	14.08	17.52	18.37	17.48		
$\Delta(\text{high-low}) p_T^V$	9.00	12.03	12.54	12.04	-	-
3-jet						
	CR _{low}	SR	High- ΔR CR	weighted-sum	weighted-sum $\Delta(\text{Low-}\Delta R \text{ CR} - \text{SR})$	weighted-sum $\Delta(\text{High-}\Delta R \text{ CR} - \text{SR})$
150-250 p_T^V	4.12	4.41	4.77	4.42	-1.26	-1.41
250 p_T^V	14.95	16.97	18.44	17.33		
$\Delta(\text{high-low}) p_T^V$	10.83	12.56	13.67	12.91	-	-
1-lepton						
2-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum $\Delta(\text{Low-}\Delta R \text{ CR} - \text{SR})$	weighted-sum $\Delta(\text{High-}\Delta R \text{ CR} - \text{SR})$
150-250 p_T^V	5.14	5.41	5.29	5.30	-1.95	+0.60
250 p_T^V	14.08	17.34	18.30	17.47		
$\Delta(\text{high-low}) p_T^V$	8.94	11.93	13.01	12.17	-	-
3-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum $\Delta(\text{Low-}\Delta R \text{ CR} - \text{SR})$	weighted-sum $\Delta(\text{High-}\Delta R \text{ CR} - \text{SR})$
150-250 p_T^V	4.08	4.38	4.61	4.39	-1.89	+1.34
250 p_T^V	15.24	18.02	19.01	18.24		
$\Delta(\text{high-low}) p_T^V$	11.16	13.64	14.40	13.85	-	-

B.2 Breakdown of flavour composition uncertainties

In this section, the full breakdown of the flavour composition uncertainties for the W +jets background modelling is reported. Table B.33 and Table B.34 summarise the full breakdown of the uncertainty on the ratio between the $W + bc$ and the main $W + bb$ components in the 1-lepton and the 0-lepton channel, respectively. Table B.33 and Table B.34 show the full breakdown of the uncertainty on the ratio between the $W + bl$ and the main $W + bb$ components in the 1-lepton and the 0-lepton channel, respectively. And Table B.33 and Table B.34 are the breakdowns for the uncertainty on the ratio between the $W + cc$ and the main $W + bb$ components in the 1-lepton and the 0-lepton channel, respectively.

Table B.33: Systematic uncertainties on bc/bb extrapolation in $W+hf$ events in the 1-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**1-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+1.8 / -1.3 %	+3.0 / -1.8 %	+2.8 / -1.1 %	+3.2 / -1.9 %
μ_F	-0.4 / +0.4 %	-0.2 / +0.2 %	+0.0 / +0.3 %	-0.1 / +0.2 %
$\mu_R + \mu_F$	-1.0 / +1.3 %	-1.5 / +2.9 %	-0.8 / +2.7 %	-1.8 / +3.0 %
PDF13000	± 3.2 %	± 4.2 %	± 4.2 %	± 3.9 %
PDF25300	± 1.7 %	± 0.4 %	± 2.0 %	± 1.2 %
PDF MAX	± 3.2 %	± 4.2 %	± 4.2 %	± 3.9 %
SH vs MG	-23.8 / -23.8 %	-19.4 / -19.4 %	-27.8 / -27.8 %	-18.2 / -18.2 %
Final uncertainty	± 24.1 %	± 20.0 %	± 28.2 %	± 18.8 %

1-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+1.0 / +0.2 %	+1.0 / +0.6 %	+2.3 / -1.0 %	+0.0 / +0.2 %
μ_F	-0.2 / +0.3 %	+0.3 / -0.1 %	-0.0 / +0.4 %	+0.0 / +0.2 %
$\mu_R + \mu_F$	+0.1 / +0.9 %	-0.0 / +1.3 %	+0.8 / +0.7 %	-0.9 / +2.3 %
PDF13000	± 2.3 %	± 2.7 %	± 2.8 %	± 2.6 %
PDF25300	± 1.1 %	± 0.2 %	± 1.2 %	± 0.6 %
PDF MAX	± 2.3 %	± 2.7 %	± 2.8 %	± 2.6 %
SH vs MG	-21.4 / -21.4 %	-20.5 / -20.5 %	-24.5 / -24.5 %	-20.3 / -20.3 %
Final uncertainty	± 21.5 %	± 20.7 %	± 24.7 %	± 20.6 %

Table B.34: Systematic uncertainties on bc/bb extrapolation in $W+hf$ events in the 0-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**0-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+3.0 / -1.7 %	+3.0 / -1.7 %	+1.9 / -1.0 %	+2.7 / -1.7 %
μ_F	-0.1 / +0.1 %	-0.1 / +0.1 %	-0.1 / +0.1 %	-0.1 / +0.1 %
$\mu_R + \mu_F$	-1.6 / +3.1 %	-1.6 / +3.1 %	-0.9 / +2.0 %	-1.7 / +2.7 %
PDF13000	± 1.7 %	± 1.7 %	± 1.7 %	± 1.6 %
PDF25300	± 1.0 %	± 1.0 %	± 0.0 %	± 0.2 %
PDF MAX	± 1.7 %	± 1.7 %	± 1.7 %	± 1.6 %
SH vs MG	-22.1 / -22.1 %	-22.1 / -22.1 %	-35.1 / -35.1 %	-24.1 / -24.1 %
Final uncertainty	± 22.3 %	± 22.3 %	± 35.2 %	± 24.2 %

0-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+0.6 / +0.6 %	+0.6 / +0.6 %	+1.1 / +0.4 %	+1.9 / -0.5 %
μ_F	+0.4 / -0.5 %	+0.4 / -0.5 %	-0.3 / +0.2 %	∓ 0.0 %
$\mu_R + \mu_F$	+0.1 / +1.1 %	+0.1 / +1.1 %	+0.3 / +0.9 %	-0.7 / +1.9 %
PDF13000	± 0.9 %	± 0.9 %	± 1.0 %	± 0.9 %
PDF25300	∓ 0.6 %	∓ 0.6 %	∓ 0.2 %	∓ 0.3 %
PDF MAX	± 0.9 %	± 0.9 %	± 1.0 %	± 0.9 %
SH vs MG	-23.2 / -23.2 %	-23.2 / -23.2 %	-24.4 / -24.4 %	-22.2 / -22.2 %
Final uncertainty	± 23.2 %	± 23.2 %	± 24.4 %	± 22.2 %

Table B.35: Systematic uncertainties on bl/bb extrapolation in W +hf events in the 1-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**1-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+2.5 / -1.8 %	+3.0 / -1.9 %	+1.7 / -0.6 %	+3.1 / -2.0 %
μ_F	-0.0 / -0.0 %	± 0.5 %	+0.2 / -0.2 %	+0.5 / -0.4 %
$\mu_R + \mu_F$	-2.0 / +2.3 %	-2.3 / +3.4 %	-0.9 / +1.9 %	-2.6 / +3.3 %
PDF13000	∓ 0.4 %	∓ 0.3 %	∓ 0.5 %	∓ 0.6 %
PDF25300	∓ 0.6 %	∓ 1.3 %	∓ 0.9 %	∓ 1.1 %
PDF MAX	∓ 0.4 %	∓ 0.3 %	∓ 0.5 %	∓ 0.6 %
SH vs MG	-29.8 / -29.8 %	-12.1 / -12.1 %	-28.6 / -28.6 %	-11.3 / -11.3 %
Final uncertainty	± 29.9 %	± 12.3 %	± 28.6 %	± 11.6 %

1-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+1.8 / -1.0 %	+1.1 / -0.2 %	+0.9 / +0.3 %	+2.7 / -1.6 %
μ_F	-0.2 / +0.2 %	+0.7 / -0.5 %	+0.3 / +0.1 %	+0.4 / -0.2 %
$\mu_R + \mu_F$	-0.9 / +1.7 %	-0.8 / +1.5 %	+0.2 / +0.9 %	-1.9 / +2.8 %
PDF13000	∓ 0.5 %	∓ 0.3 %	∓ 0.5 %	∓ 0.5 %
PDF25300	∓ 0.4 %	∓ 0.6 %	∓ 0.5 %	∓ 0.5 %
PDF MAX	∓ 0.4 %	∓ 0.3 %	∓ 0.5 %	∓ 0.5 %
SH vs MG	-15.8 / -15.8 %	-8.5 / -8.5 %	-19.5 / -19.5 %	-9.5 / -9.5 %
Final uncertainty	± 15.8 %	± 8.6 %	± 19.5 %	± 9.7 %

Table B.36: Systematic uncertainties on bl/bb extrapolation in W +hf events in the 0-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**0-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+2.2 / -1.7 %	+2.2 / -1.7 %	+1.1 / -0.7 %	+2.1 / -1.7 %
μ_F	+0.5 / -0.4 %	+0.5 / -0.4 %	+0.2 / -0.1 %	+0.3 / -0.3 %
$\mu_R + \mu_F$	-2.0 / +2.7 %	-2.0 / +2.7 %	-0.9 / +1.4 %	-2.0 / +2.3 %
PDF13000	∓ 0.2 %	∓ 0.2 %	∓ 0.3 %	∓ 0.4 %
PDF25300	∓ 1.3 %	∓ 1.3 %	∓ 1.0 %	∓ 1.1 %
PDF MAX	∓ 0.2 %	∓ 0.2 %	∓ 0.3 %	∓ 0.4 %
SH vs MG	-12.3 / -12.3 %	-12.3 / -12.3 %	-29.3 / -29.3 %	-14.1 / -14.1 %
Final uncertainty	± 12.5 %	± 12.5 %	± 29.3 %	± 14.3 %

0-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+0.1 / +0.9 %	+0.1 / +0.9 %	+1.1 / +0.6 %	+2.0 / -0.6 %
μ_F	+0.7 / -0.6 %	+0.7 / -0.6 %	+0.1 / -0.1 %	+0.3 / -0.2 %
$\mu_R + \mu_F$	+0.2 / +1.1 %	+0.2 / +1.1 %	+0.2 / +1.3 %	-1.0 / +2.4 %
PDF13000	∓ 0.4 %	∓ 0.4 %	∓ 0.4 %	∓ 0.5 %
PDF25300	∓ 0.7 %	∓ 0.7 %	∓ 0.6 %	∓ 0.7 %
PDF MAX	∓ 0.4 %	∓ 0.4 %	∓ 0.4 %	∓ 0.5 %
SH vs MG	-14.7 / -14.7 %	-14.7 / -14.7 %	-17.7 / -17.7 %	-13.7 / -13.7 %
Final uncertainty	± 14.7 %	± 14.7 %	± 17.7 %	± 13.8 %

Table B.37: Systematic uncertainties on cc/bb extrapolation in $W+hf$ events in the 1-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**1-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+0.4 / -0.5 %	+1.9 / -1.1 %	+0.8 / -0.3 %	+1.3 / -0.9 %
μ_F	-0.0 / +0.1 %	+0.1 / -0.1 %	$\mp$ 0.0 %	+0.0 / -0.0 %
$\mu_R + \mu_F$	-0.4 / +0.4 %	-1.1 / +2.1 %	-0.3 / +0.7 %	-0.9 / +1.3 %
PDF13000	$\pm$ 1.3 %	$\pm$ 4.2 %	$\pm$ 2.5 %	$\pm$ 2.9 %
PDF25300	$\pm$ 0.7 %	$\pm$ 0.7 %	$\pm$ 1.0 %	$\pm$ 0.8 %
PDF MAX	$\pm$ 1.3 %	$\pm$ 4.2 %	$\pm$ 2.5 %	$\pm$ 2.9 %
SH vs MG	-2.3 / -2.3 %	-2.9 / -2.9 %	-4.6 / -4.6 %	-0.8 / -0.8 %
Final uncertainty	$\pm$ 2.6 %	$\pm$ 5.3 %	$\pm$ 5.3 %	$\pm$ 3.2 %

1-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	-0.2 / +0.4 %	+0.5 / +0.0 %	+0.2 / +0.4 %	+0.9 / -0.3 %
μ_F	+0.1 / +0.0 %	+0.6 / -0.3 %	+0.2 / +0.1 %	+0.3 / -0.1 %
$\mu_R + \mu_F$	+0.4 / -0.2 %	-0.4 / +1.0 %	+0.4 / +0.3 %	-0.4 / +1.1 %
PDF13000	$\pm$ 0.9 %	$\pm$ 3.1 %	$\pm$ 1.7 %	$\pm$ 2.2 %
PDF25300	$\pm$ 0.4 %	$\pm$ 0.6 %	$\pm$ 0.7 %	$\pm$ 0.6 %
PDF MAX	$\pm$ 0.9 %	$\pm$ 3.1 %	$\pm$ 1.7 %	$\pm$ 2.2 %
SH vs MG	-4.4 / -4.4 %	-3.1 / -3.1 %	-6.2 / -6.2 %	-3.6 / -3.6 %
Final uncertainty	$\pm$ 4.5 %	$\pm$ 4.4 %	$\pm$ 6.5 %	$\pm$ 4.2 %

Table B.38: Systematic uncertainties on cc/bb extrapolation in $W+hf$ events in the 0-lepton channel for the 2-jet (top) and 3-jet (bottom) categories.**0-lep 2-jet category**

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+1.9 / -1.2 %	+1.9 / -1.2 %	+0.5 / -0.4 %	+0.7 / -0.5 %
μ_F	+0.3 / -0.0 %	+0.3 / -0.0 %	+0.1 / -0.2 %	+0.2 / -0.1 %
$\mu_R + \mu_F$	-1.3 / +2.2 %	-1.3 / +2.2 %	-0.4 / +0.6 %	-0.6 / +0.9 %
PDF13000	$\pm$ 2.4 %	$\pm$ 2.4 %	$\pm$ 1.3 %	$\pm$ 1.4 %
PDF25300	$\mp$ 0.1 %	$\mp$ 0.1 %	$\pm$ 0.0 %	$\pm$ 0.1 %
PDF MAX	$\pm$ 2.4 %	$\pm$ 2.4 %	$\pm$ 1.3 %	$\pm$ 1.4 %
SH vs MG	-1.6 / -1.6 %	-1.6 / -1.6 %	-3.2 / -3.2 %	-0.7 / -0.7 %
Final uncertainty	$\pm$ 3.3 %	$\pm$ 3.3 %	$\pm$ 3.5 %	$\pm$ 1.7 %

0-lep 3-jet category

Systematic	Low- ΔR CR	High- ΔR CR	SR	SR+CRs
μ_R	+0.9 / +0.1 %	+0.9 / +0.1 %	+0.0 / +0.6 %	+0.6 / +0.2 %
μ_F	+0.5 / -0.6 %	+0.5 / -0.6 %	-0.1 / +0.1 %	+0.2 / -0.1 %
$\mu_R + \mu_F$	-0.5 / +1.8 %	-0.5 / +1.8 %	+0.4 / +0.0 %	-0.1 / +0.9 %
PDF13000	$\pm$ 0.9 %	$\pm$ 0.9 %	$\pm$ 0.4 %	$\pm$ 0.6 %
PDF25300	$\mp$ 0.6 %	$\mp$ 0.6 %	$\mp$ 0.2 %	$\mp$ 0.2 %
PDF MAX	$\pm$ 0.9 %	$\pm$ 0.9 %	$\pm$ 0.4 %	$\pm$ 0.6 %
SH vs MG	-2.8 / -2.8 %	-2.8 / -2.8 %	-2.3 / -2.3 %	-1.0 / -1.0 %
Final uncertainty	$\pm$ 3.2 %	$\pm$ 3.2 %	$\pm$ 2.4 %	$\pm$ 1.3 %

B.3 Derivation of W +jets shape uncertainties using a 1-dimensional parametrization

Many studies have been performed for the estimation of the W +jets shape systematics for the previous round of the $VH, H \rightarrow b\bar{b}$ analysis. In Ref [13], the derivation of the W +jets shape uncertainties was based on the 1-dimensional parametrization explained in Section 6.7; the nominal MC generator is compared to its internal weights, such as the renormalization and the factorization scales and the PDF variations, and then to the alternative MC prediction by MADGRAPH 5.

Comparison of SHERPA 2.2.1 and MADGRAPH 5 predictions The shape uncertainties are derived using truth events only and on the two variables of the analysis-BDT having the highest discrimination power between the nominal SHERPA 2.2.1 sample and the alternative MADGRAPH 5 one, m_{bb} and p_T^V . The shapes are estimated separately in the 0-lepton and 1-lepton channels and in the 2-jet and 3-jet categories. To define an uncertainty, the distributions of the nominal and alternative generators in each channel and jet category are normalized to the same area and compared. A functional form is fitted through the ratio between the two MC predictions, converted into a binned histogram to match the format of the fit framework, and then symmetrized. If some shape uncertainties are found to have a similar impact in certain analysis categories, a common uncertainty is taken and implemented accross all regions.

The comparison between the MC predictions in the p_T^V distribution can be parametrized by a straight line in the range of $p_T^V > 150$ GeV. In fact, no attempt of adding the medium p_T^V ($75 < p_T^V < 150$ GeV) region was made for the previous round of this analysis. Figure B.67 (Figure B.68) shows the comparison of the p_T^V distributions between the SHERPA 2.2.1 and MADGRAPH 5 samples in the 1-lepton (0-lepton) channel for the 2-jet and 3-jet categories.

As for the case of the p_T^V shape, the comparison between the MC predictions in the m_{bb} distribution can be parametrized by a straight line to the unit-normalized ratio between the nominal and the alternative samples. Figure B.69 (Figure B.70) shows the comparison of the m_{bb} distributions between the SHERPA 2.2.1 and MADGRAPH 5 samples in the 1-lepton (0-lepton) channel for the 2-jet and 3-jet categories.

Once derived, the p_T^V and m_{bb} shape systematics in each lepton channel and jet category are compared to see if they have a similar impact on the p_T^V and m_{bb} distributions, respectively.

Figure B.71 shows the four p_T^V shape uncertainties derived in the 1-lepton and 0-lepton channels and for the 2-jet and the 3-jet categories. As a similar trend is observed, the maximal variation is taken and applied accross all analysis regions in a correlated fashion. In this case, the maximum variation would be the one in the 0-lepton channel and in the 3-jet region, however, the 1-lepton channel as much more statistics available so the shape in the 1-lepton channel and in the 3-jet category is chosen as a common uncertainty applied accross all analysis regions in a correlated fashion.

Figure B.72 shows the four m_{bb} shape uncertainties derived in the 1-lepton and 0-lepton

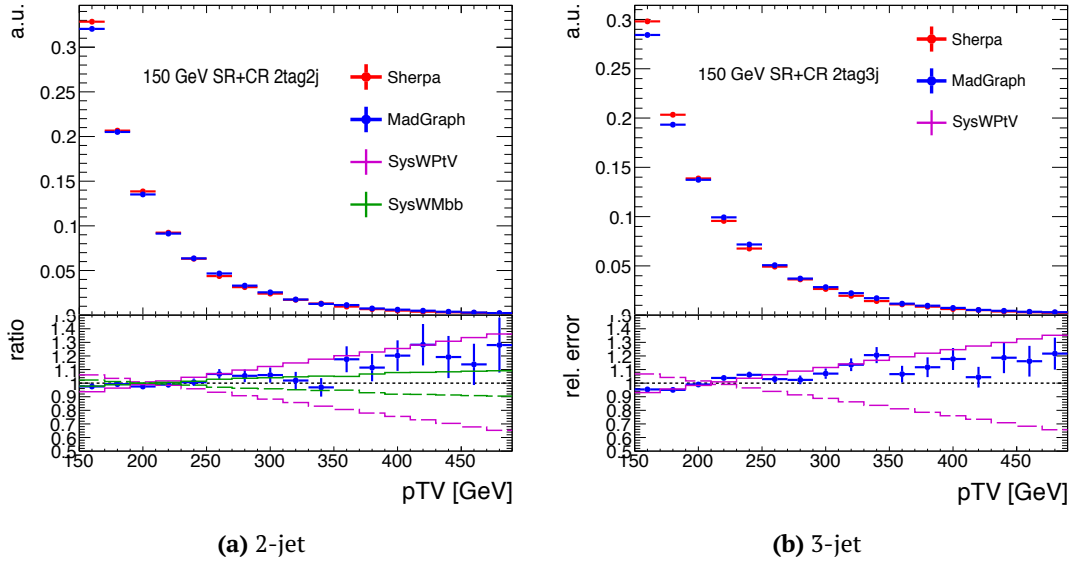


Figure B.67: p_T^V shape uncertainties derived in 2-jet (a) and 3-jet (b) of the 1-lepton channel. The red and blue histograms represent the nominal SHERPA 2.2.1 and the alternative MADGRAPH 5 predictions, while the violet histogram is the p_T^V shape uncertainty already converted into a binned distribution.

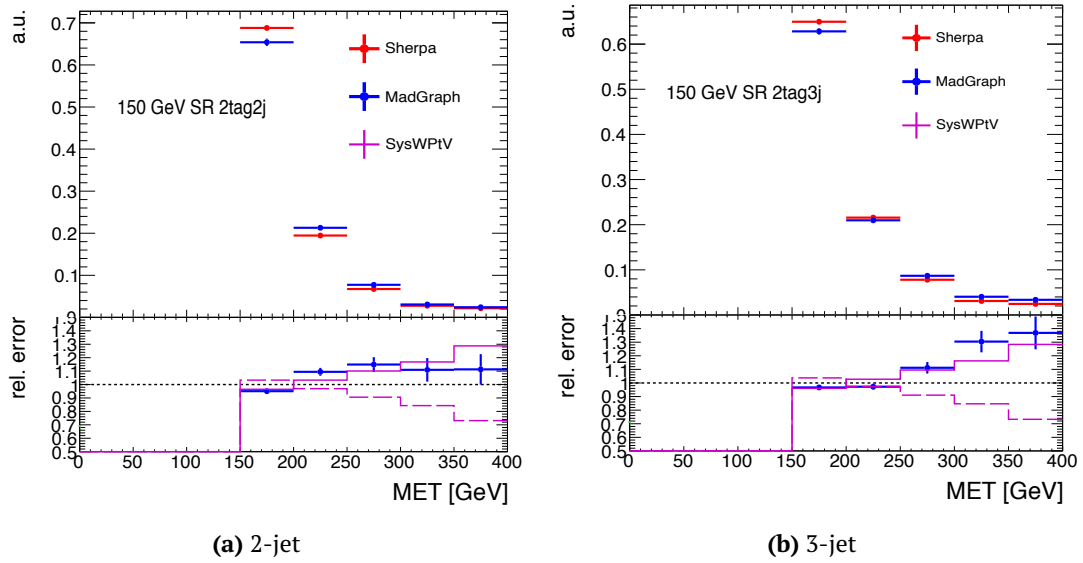


Figure B.68: $p_T^V (E_T^{\text{miss}})$ shape uncertainties derived in 2-jet (a) and 3-jet (b) of the 0-lepton channel. The red and blue histograms represent the nominal SHERPA 2.2.1 and the alternative MADGRAPH 5 predictions, while the violet histogram is the p_T^V shape uncertainty already converted into a binned distribution.

channels and for the 2-jet and the 3-jet categories. As a similar trend is observed, the maximal variation is taken and applied across all analysis regions in a correlated fashion. The maximal variation is represented by the shape in the 1-lepton channel and in the 2-jet category.

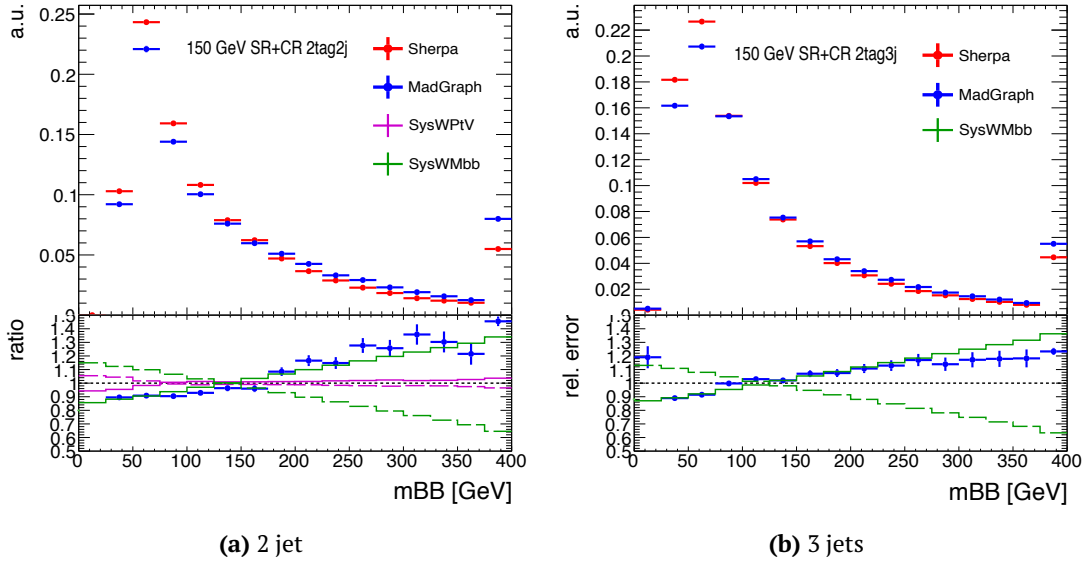


Figure B.69: m_{bb} shape uncertainties derived in 2-jet (a) and 3-jet (b) of the 1-lepton channel. The red and blue histograms represent the nominal SHERPA 2.2.1 and the alternative MADGRAPH 5 predictions, while the green histogram is the m_{bb} shape uncertainty already converted into a binned distribution.

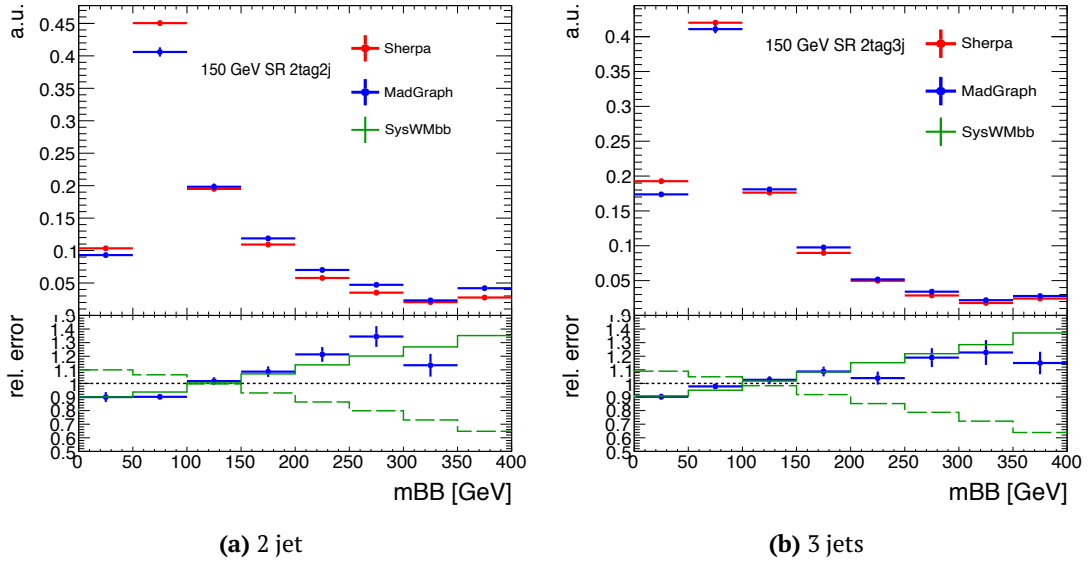


Figure B.70: m_{bb} shape uncertainties derived in 2-jet (a) and 3-jet (b) of the 0-lepton channel. The red and blue histograms represent the nominal SHERPA 2.2.1 and the alternative MADGRAPH 5 predictions, while the green histogram is the m_{bb} shape uncertainty already converted into a binned distribution.

Comparison of the nominal SHERPA 2.2.1 with its internal weights Figure B.73 and Figure B.74 show the comparison between the nominal SHERPA 2.2.1 and two alternative versions with renormalization and factorization scales doubled and halved both at the same time, for all the kinematic variables entering the analysis BDT¹⁶ in the 1-lepton channel and for the 2-jet and 3-jet categories, respectively. Similar comparisons are

¹⁶The variable $\Delta Y(W, H)$ is missing but similar conclusions are expected for it.

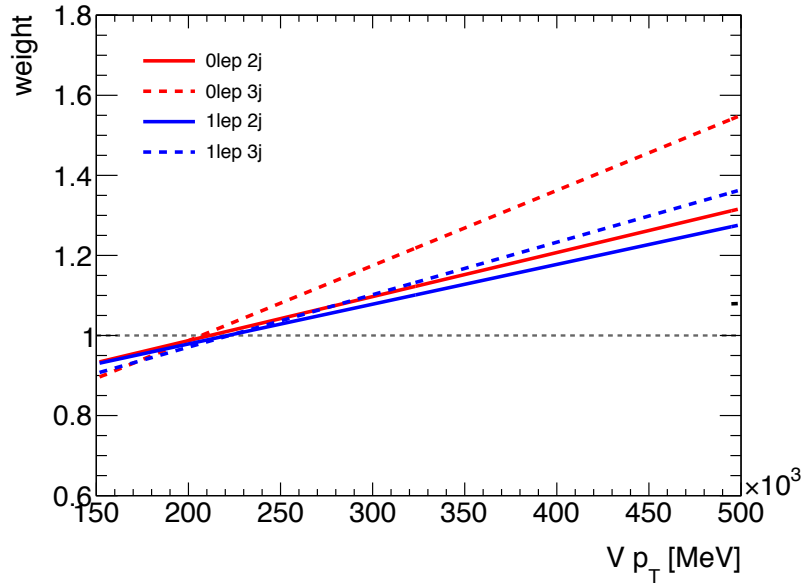


Figure B.71: Comparison of the p_T^V shape uncertainties derived in the 0-lepton and 1-lepton channels and in the 2-jet and 3-jet categories. The largest shape is the one derived in the 1-lepton, 3-jet region and it is applied in a correlated fashion across all the analysis regions.

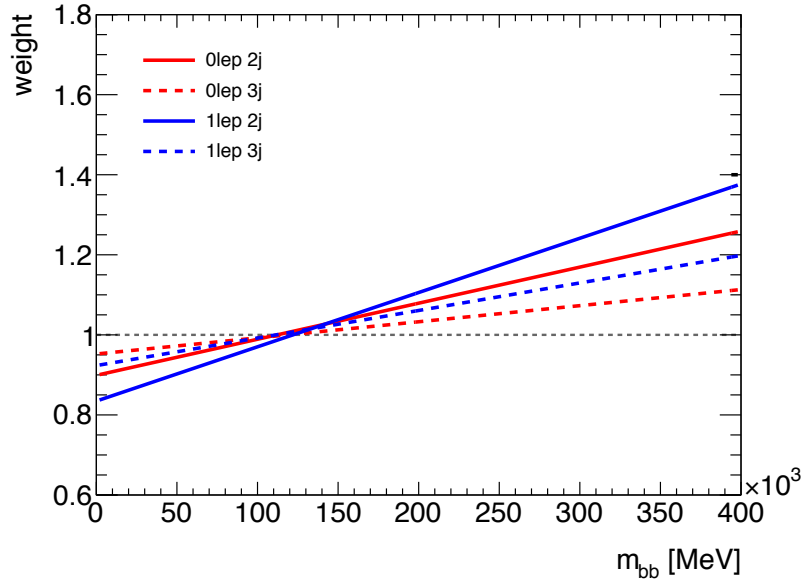


Figure B.72: Comparison of the m_{bb} shape uncertainties derived in the 0-lepton and 1-lepton channels and in the 2-jet and 3-jet categories. The largest shape is the one derived in the 1-lepton, 2-jet region and it is applied in a correlated fashion across all the analysis regions.

done also in the case where μ_R and μ_F are doubled or halved one at the time but are not shown in the following as the differences with the nominal SHERPA 2.2.1 generator are smaller than the ones obtained when doubling/halving the scales together. The same set of studies has been performed also in the 0-lepton channel and similar conclusions can be drawn.

The red line corresponds to the nominal SHERPA 2.2.1 prediction, the black (orange) line shows the SHERPA 2.2.1 prediction with halved (doubled) renormalization and factorization scales. The p_T^V and m_{bb} shape systematics, derived from the comparison between the nominal SHERPA 2.2.1 and the alternative MADGRAPH 5 samples, are shown in magenta and green, respectively. It can be observed that the contribution from the difference between the nominal SHERPA 2.2.1 and the one with varied scale settings is much smaller with respect to the MC generators comparison. For this reason, the variation coming from the μ_R and μ_F scales has not been considered as a systematic uncertainty on the shape of the p_T^V and m_{bb} W +jets distributions.

B.4 Derivation of W +jets shape uncertainties through a BDT-based approach with particle-level events

This section summarises the first studies performed to derive the W +jets shape uncertainties using the BDT-reweighting approach with truth events. In this studies, the p_T^V shape uncertainties is not factorized out from the BDTr procedure as in the final hybrid BDTr approach used in the analysis. Distributions of $W + bb$ events are shown in the 1-lepton channel and for both the 2-jet and 3-jet categories. Very good closure can be seen and similar performances were observed also for $W + bc$, $W + bl$ and $W + cc$ events.

Figure B.75 shows the closure of the BDTr shape systematic on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 2-jet category and for $W + bb$ events. Figure B.76 shows the closure for the 1-lepton, 2-tag, 3-jet category and for $W + bb$ events. The BDTr shape systematic was derived in the range of $p_T^V > 150$ GeV as when these studies were performed there was no plan of adding the medium p_T^V ($75 < p_T^V < 150$ GeV) as a fitted region in the 1-lepton channel.

The effect of the BDTr shape systematic on the analysis-BDT is shown in Figure B.77 for $W + bb$ events in the 2-tag, 2-jet category on the left and 2-tag, 3-jet category on the right for the 1-lepton channel.

B.5 Derivation of W +jets shape uncertainties through a hybrid BDT-based approach using reconstructed event

This section contains the closure performances of the p_T^V -factorized BDTr approach on all the input variables of the BDT (excluding the m_{bb} and the p_T^V or E_T^{miss} distributions that are included in Section 6.9.3) for $W + bb$ events in both the 0-lepton and 1-lepton channels and for the 3-jet category. Very good closure can be seen, especially on those variables for which the 1-dimensional parametrization exhibited some issues. Similar performances are observed for $W + bc$, $W + bl$ and $W + cc$ events as shown in the following.

Figure B.78 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 3-jet category and for $W + bb$ events.

Figure B.79 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 2-jet category and for $W + bc$

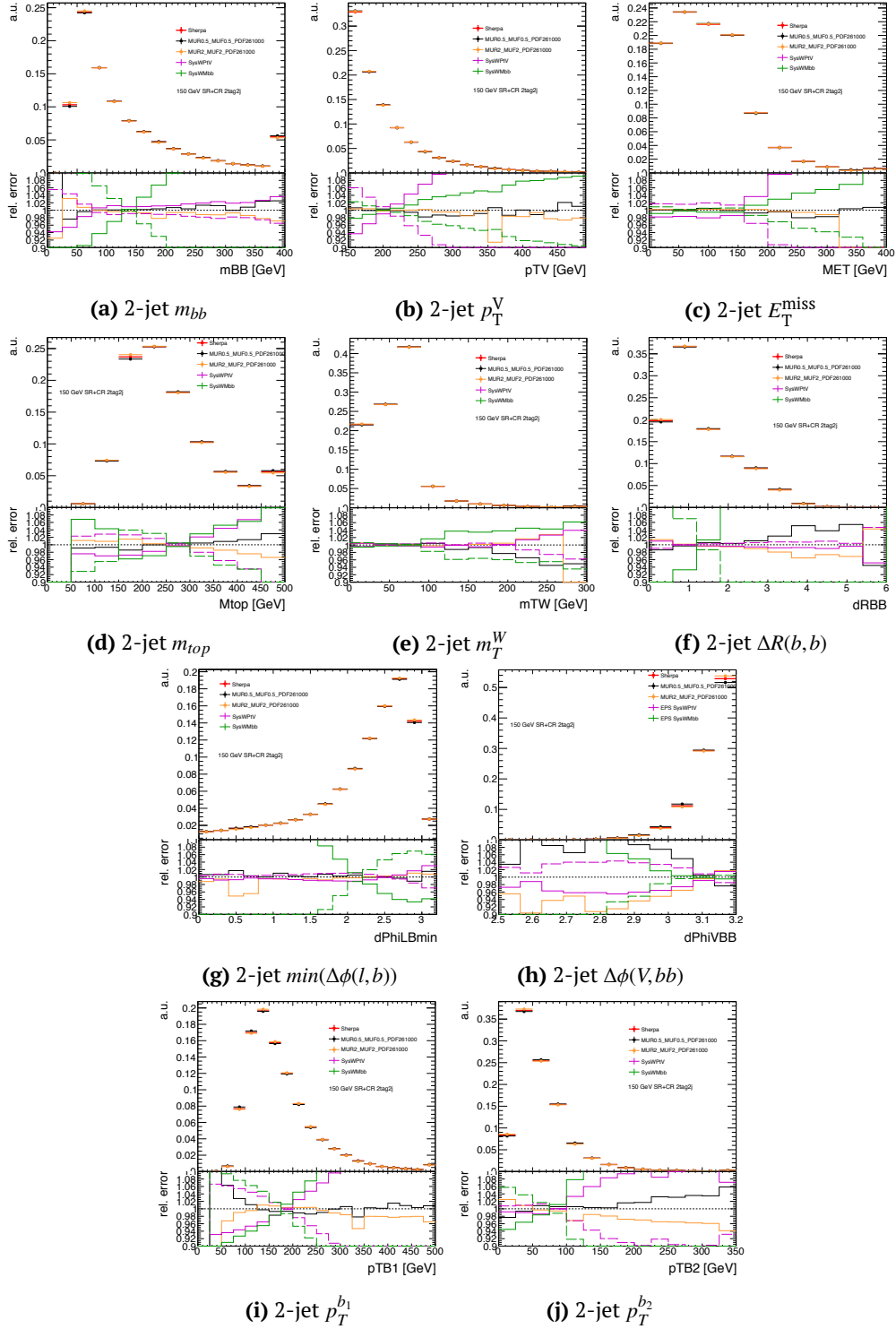


Figure B.73: Comparison of the nominal SHERPA 2.2.1 prediction (red line) with two alternative versions in which both the μ_R and μ_F are doubled (black line) or halved (orange line) in the 1-lepton, 2-tag, 2-jet category, on all the kinematic variables entering the analysis-BDT (but $\Delta Y(W, H)$).

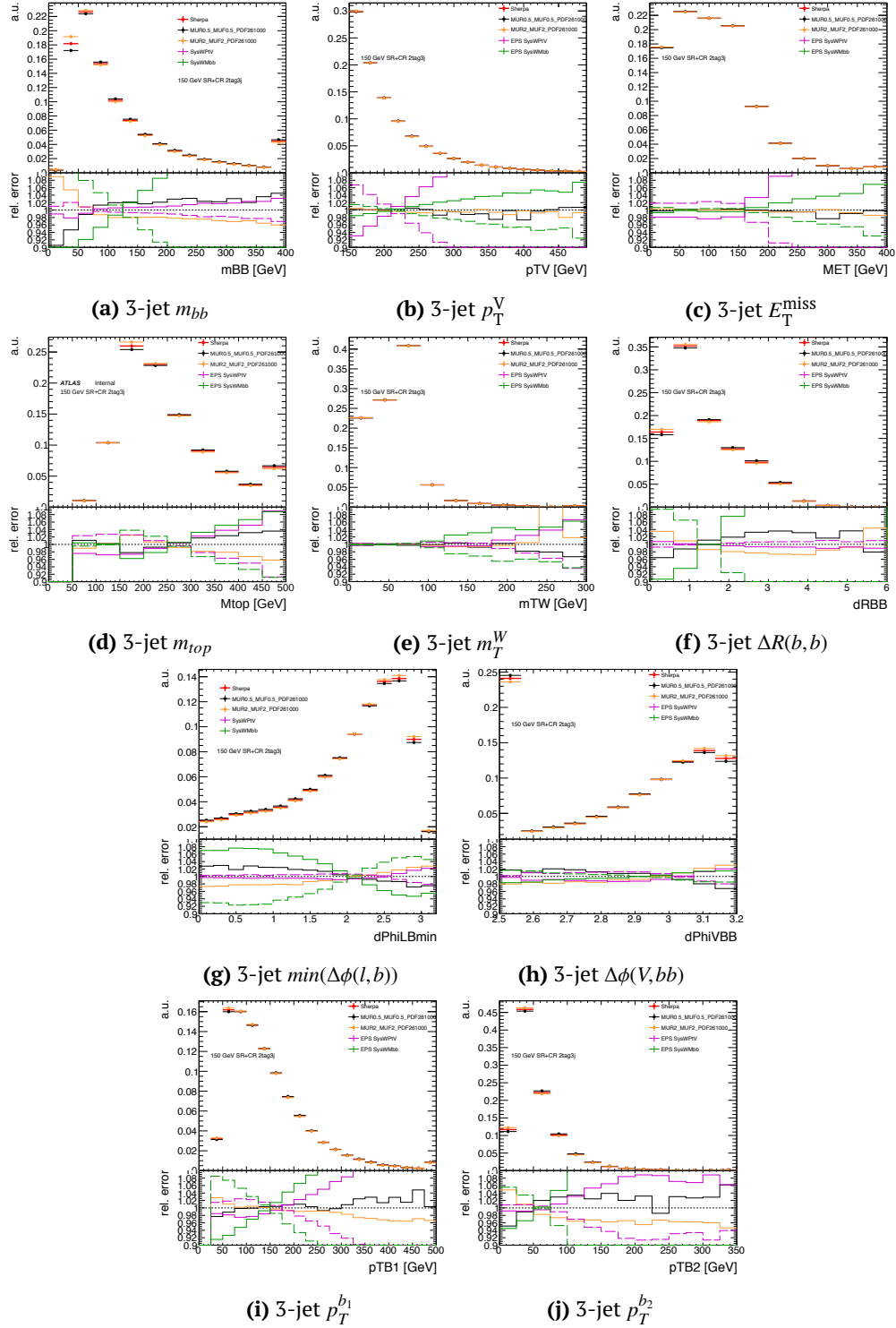


Figure B.74: Comparison of the nominal SHERPA 2.2.1 prediction (red line) with two alternative versions in which both the μ_R and μ_F are doubled (black line) or halved (orange line) in the 1-lepton, 2-tag, 3-jet category, on all the kinematic variables entering the analysis-BDT (but $\Delta Y(W, H)$).

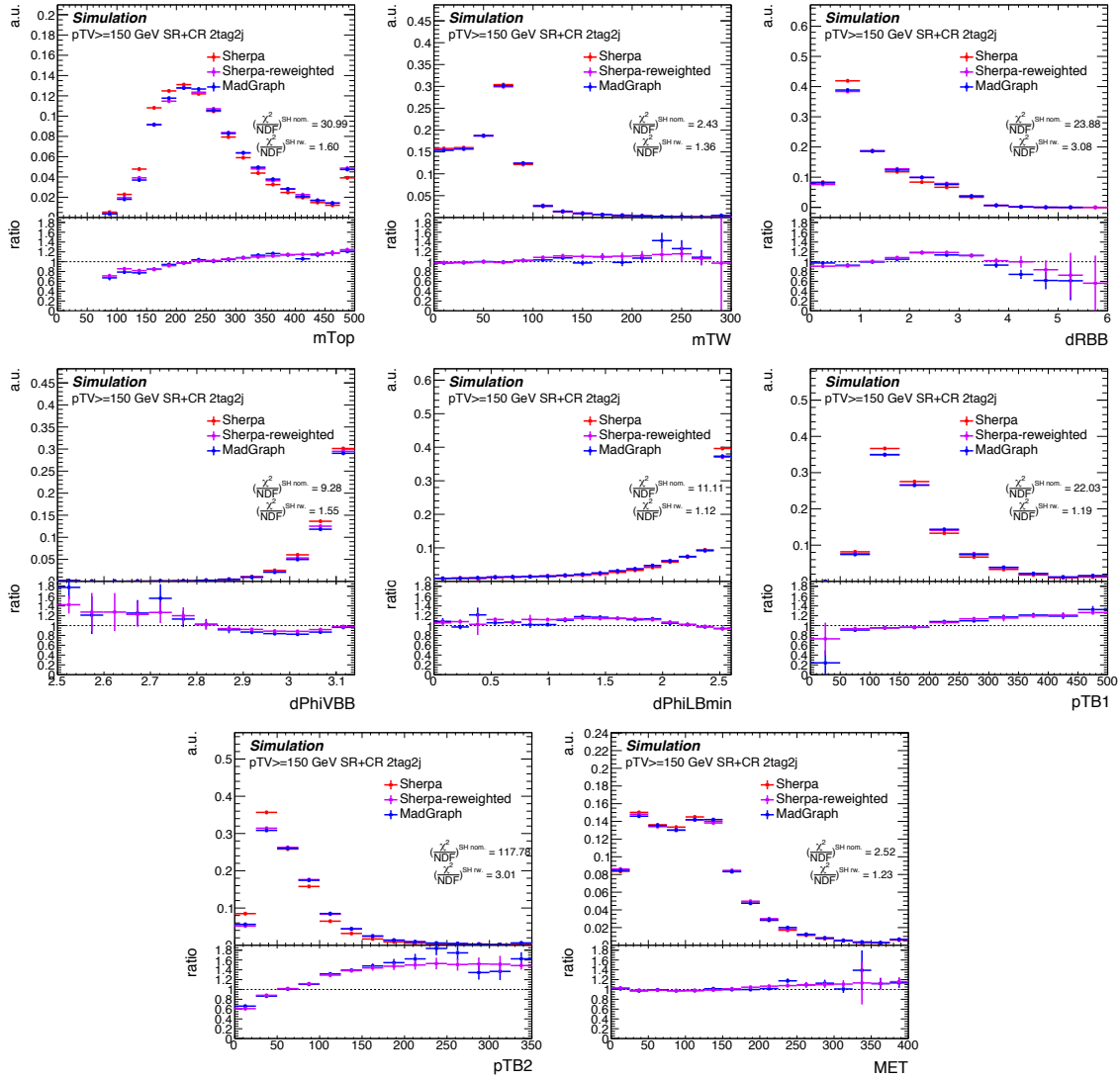


Figure B.75: Closure of the BDT approach on the variables entering the analysis-BDT in the 2-tag, 2-jet region of the 1-lepton channel and for $W + bb$ events. The nominal SHERPA 2.2.1 prediction is plotted in red, the alternative one from MADGRAPH 5 is plotted in blue and the BDT systematic is plotted in magenta.

events.

Figure B.80 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 2-jet category and for $W + bl$ events.

Figure B.81 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 2-jet category and for $W + cc$ events.

Figure B.82 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 3-jet category and for $W + bc$ events.

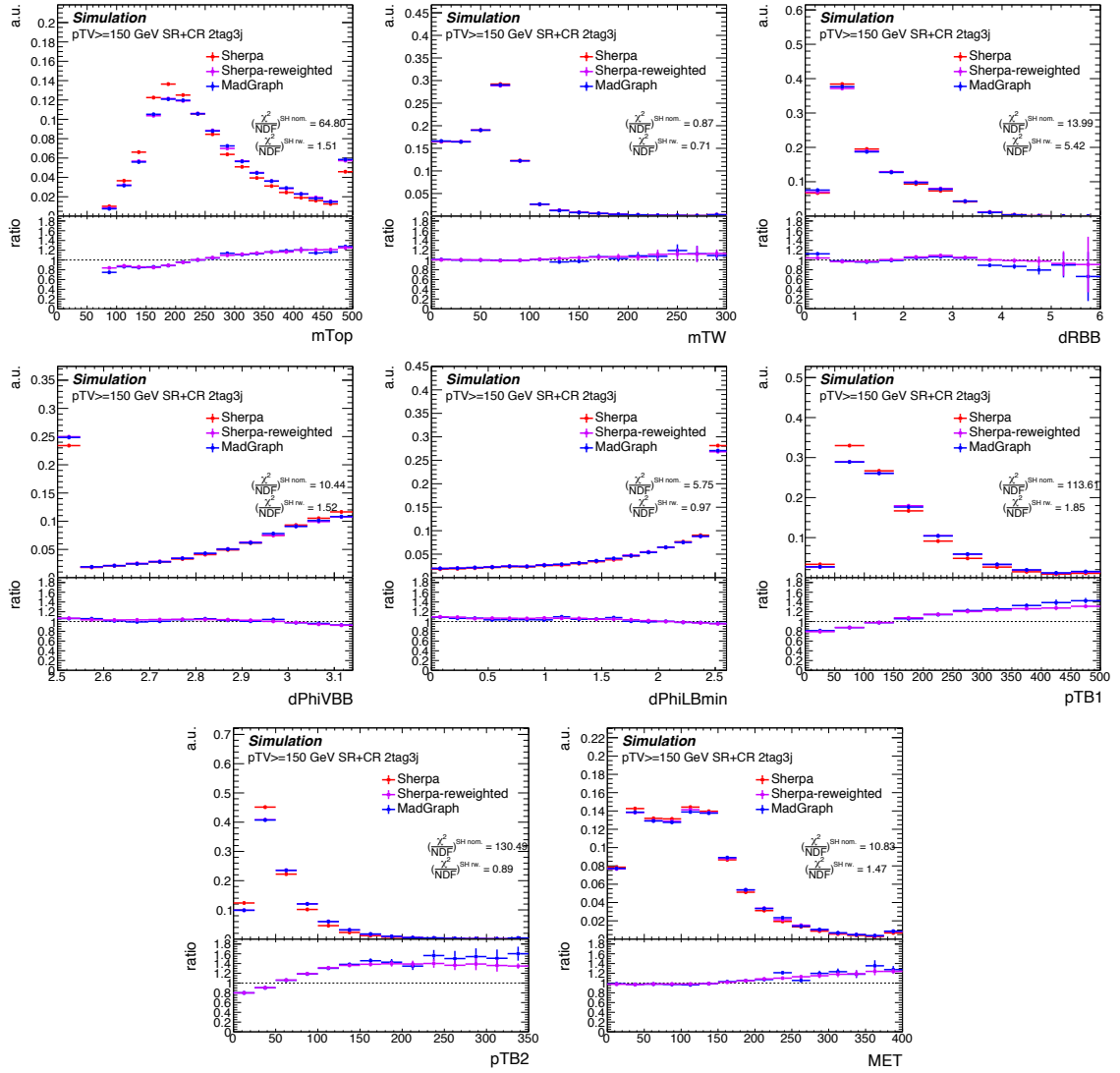


Figure B.76: Closure of the BDT approach on the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 1-lepton channel and for $W + bb$ events. The nominal SHERPA 2.2.1 prediction is plotted in red, the alternative one from MADGRAPH 5 is plotted in blue and the BDT systematic is plotted in magenta.

Figure B.83 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 3-jet category and for $W + bl$ events.

Figure B.84 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 1-lepton, 2-tag, 3-jet category and for $W + cc$ events.

Figure B.85 shows the validation of the BDT-reweighting procedure on some of the variables entering the analysis-BDT for the 0-lepton, 2-tag, 3-jet category and for $W + bb$ events.

Figure B.86 shows the validation of the BDT-reweighting procedure on some of the

are shown in Table B.39 across all regions and p_T^V bins of the analysis for both the 0-lepton and the 1-lepton channels.

Table B.39: Summary of the extrapolation uncertainties of the BDTr shape systematic on the W +jets samples across all the analysis regions of the analysis and in the 0- and 1-lepton channels. The third column for both 2-jet and 3-jet categories shows the normalization effect induced by the shape systematic weighted by the yields in each of the analysis regions. The last row for both 0-lepton and 1-lepton channels shows the normalization effect induced by the BDTr shape systematic across the different p_T^V bins of the analysis. The last two columns show the migration effect, inclusive in the two p_T^V bins induced by the p_T^V shape systematic from Low- ΔR CR to SR and from High- ΔR CR to SR.

0-lepton						
2-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum Δ (Low- ΔR CR - SR)	weighted-sum Δ (High- ΔR CR - SR)
150-250 p_T^V	-11.87	-4.31	12.87	-2.85	-7.32	+15.70
250 p_T^V	-9.96	-5.59	3.21	-2.84		
Δ (high-low) p_T^V	1.91	-1.28	-9.66	0.01	-	-
3-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum Δ (Low- ΔR CR - SR)	weighted-sum Δ (High- ΔR CR - SR)
150-250 p_T^V	-4.76	-1.73	5.84	-0.62	-2.70	+6.70
250 p_T^V	<0.5	-1.93	1.32	-0.34		
Δ (high-low) p_T^V	4.76	-0.20	-4.52	0.29	-	-
1-lepton						
2-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum Δ (Low- ΔR CR - SR)	weighted-sum Δ (High- ΔR CR - SR)
150-250 p_T^V	-11.64	-1.05	21.17	2.65	-10.13	+21.47
250 p_T^V	-7.64	-2.30	14.95	4.67		
Δ (high-low) p_T^V	4.00	-1.25	-6.76	2.03	-	-
3-jet						
	Low- ΔR CR	SR	High- ΔR CR	weighted-sum	weighted-sum Δ (Low- ΔR CR - SR)	weighted-sum Δ (High- ΔR CR - SR)
150-250 p_T^V	-3.70	-2.15	9.56	1.55	-1.40	+9.54
250 p_T^V	2.47	<0.5	2.73	1.60		
Δ (high-low) p_T^V	6.17	2.15	-6.83	0.05	-	-

B.6 W boson charge asymmetry for the modelling of the W +jets background

In a pp collider, the W^+ is produced through a valence up-quark and a down anti-quark from the sea, while the W^- is produced through an up anti-quark from the sea and a valence down-quark. Given that each proton is formed by two up-quark and a single down-quark, it is more likely to produce a W^+ than a W^- boson.

The uncertainty on the modelling of the W +jets background is dominated by the difference between the nominal SHERPA 2.2.1 generator and the alternative MADGRAPH 5 one. This section summarises the first attempt of using data/MC comparisons to assess whether one of the two generators (or both) is not supported by data.

For each kinematic variable entering the analysis-BDT, the distribution of the signal (background) processes having a negative charged lepton in the final state, coming from the decay of a W^- , is subtracted from the one having a positive charged lepton, from the decay of a W^+ . This way, the contribution of the largest background of the analysis, $t\bar{t}$, as

well as the WW , ZZ processes and the QCD multijet background largely cancels.

Figure B.92 shows the distribution of $\frac{W^+ - W^-}{W^+ + W^-}$ as a function of the transverse momentum of the lepton in the 1-lepton, 2-tag region and inclusively in SR and CRs and jet category, for the signal and background processes surviving the $W^+ - W^-$ subtraction. Among these processes, the W +jets and the single-top t -channel backgrounds have the largest $W^\pm$ charge asymmetry.

Figure B.93 shows the distributions of the m_{bb} , p_T^V and ΔR_{bb} variables in the 1-lepton, 1-tag region, and inclusively in SR and CRs, p_T^V range, and jet categories for the $qq \rightarrow WH$ signal, the diboson WZ background, all flavour components of the W +jets background, the single-top t -channel, and the single-top Wt channel, whose contribution is negligible. The 1-tag region has the advantage of having a higher number of events with respect to the 2-tag region. In the first column, both signal and backgrounds are simulated using the nominal SHERPA 2.2.1 prediction, while on the second column they are simulated using the alternative MADGRAPH 5 prediction. A large mismodelling in the case all processes are simulated using the MADGRAPH 5 prediction can be observed at low values of m_{bb} , p_T^V , and ΔR_{bb} . This is not present in the case the processes are simulated using the nominal SHERPA 2.2.1 generator, suggesting that the data do not support the prediction of the alternative generator.

Figure B.94 shows the distributions of the m_{bb} , p_T^V and ΔR_{bb} variables in the 1-lepton, 2-tag region, and inclusively in SR and CRs, p_T^V range, and jet categories for the $qq \rightarrow WH$ signal, the diboson WZ background, all flavour components of the W +jets background, the single-top t -channel, and the single-top Wt channel, whose contribution is negligible. The 2-tag region has much less events with respect to the 1-tag region but it is the one relevant for the analysis. In the first column, both signal and backgrounds are simulated using the nominal SHERPA 2.2.1 prediction, while on the second column they are simulated using the alternative MADGRAPH 5 prediction. The two generators show a very similar response in the 2-tag region and they are both in good agreement with the data. However, the large error bars due to the small number of events in this region do not allow any further conclusion.

Due to lack of time, these studies have not been further investigated but could provide a powerful tool to assess the goodness of a generator with respect to the other in modelling the W +jets background, as well as provide useful feedbacks to the theorists developing the MC simulation.

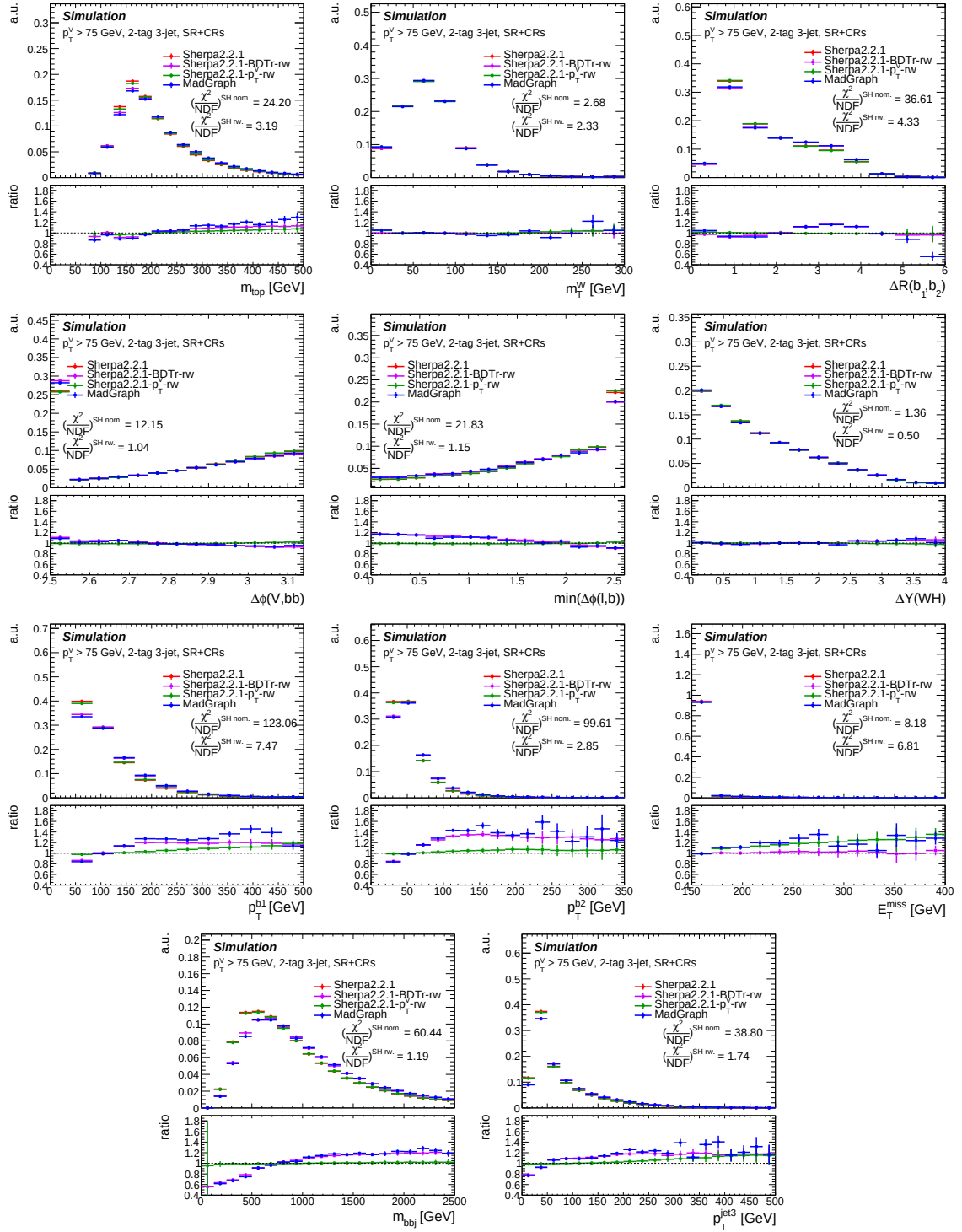


Figure B.78: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 1-lepton channel and for $W + bb$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

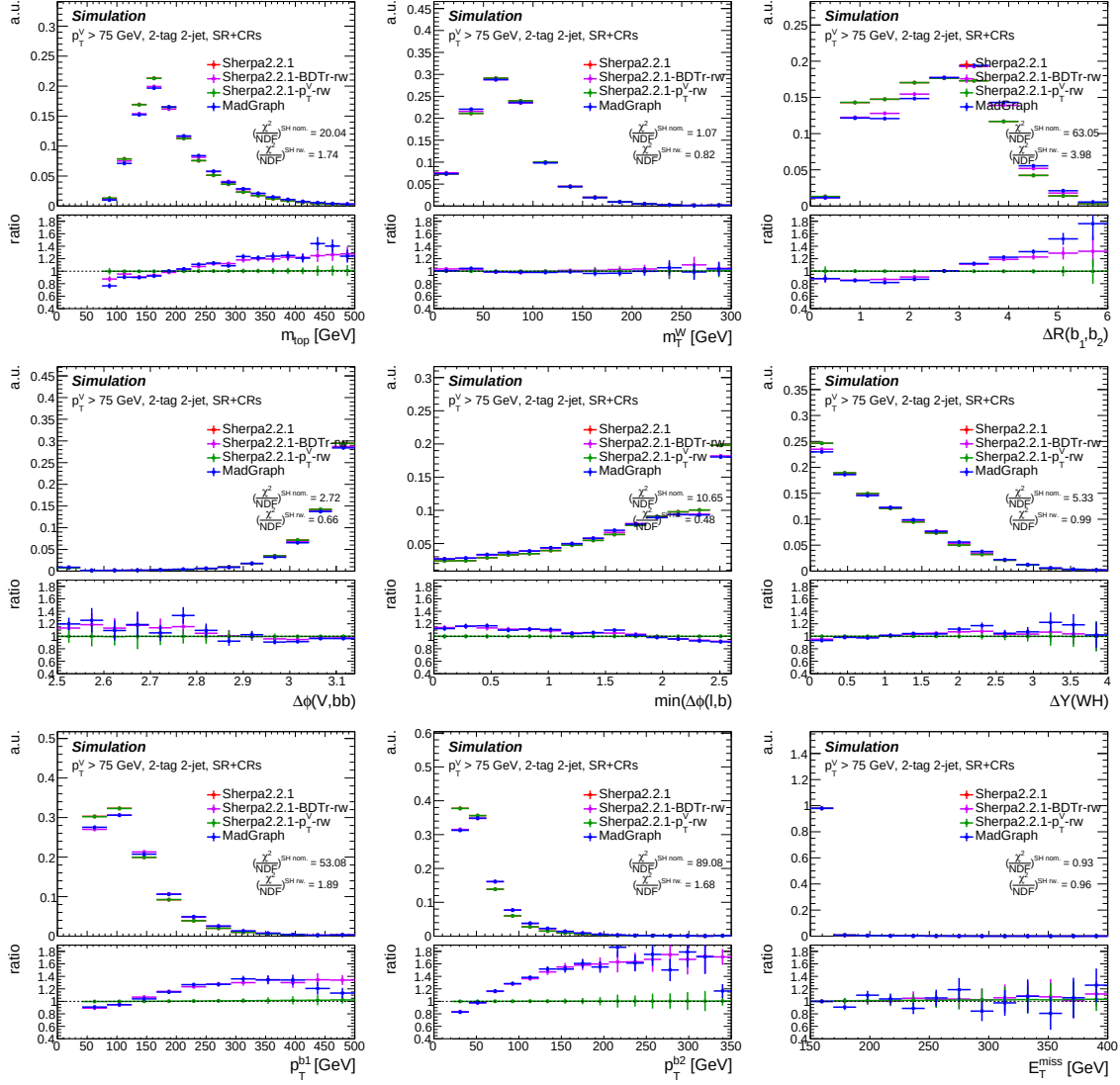


Figure B.79: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet reweighting of the 1-lepton channel and for $W + bc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f.}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

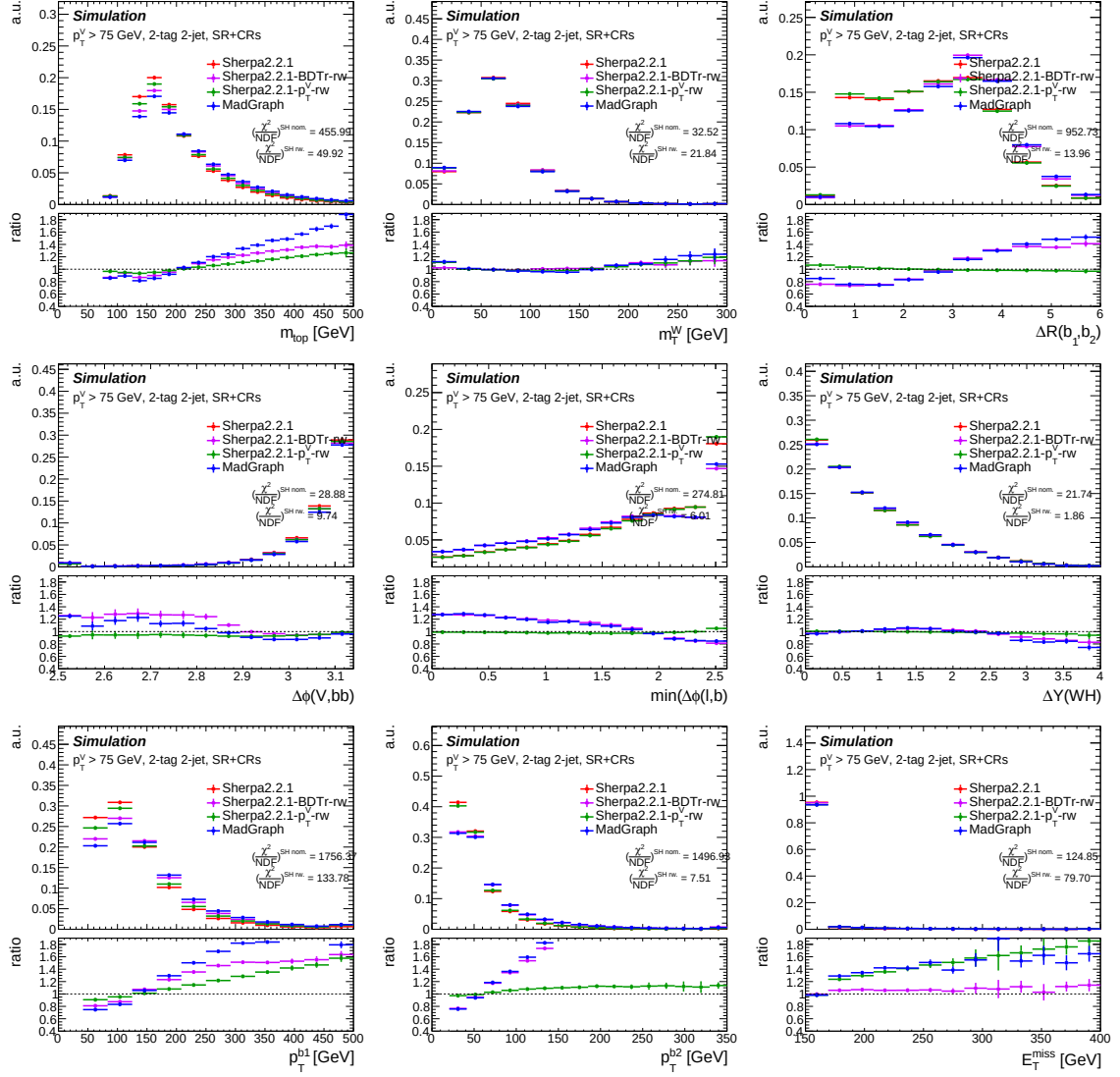


Figure B.80: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet region of the 1-lepton channel and for $W + bl$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

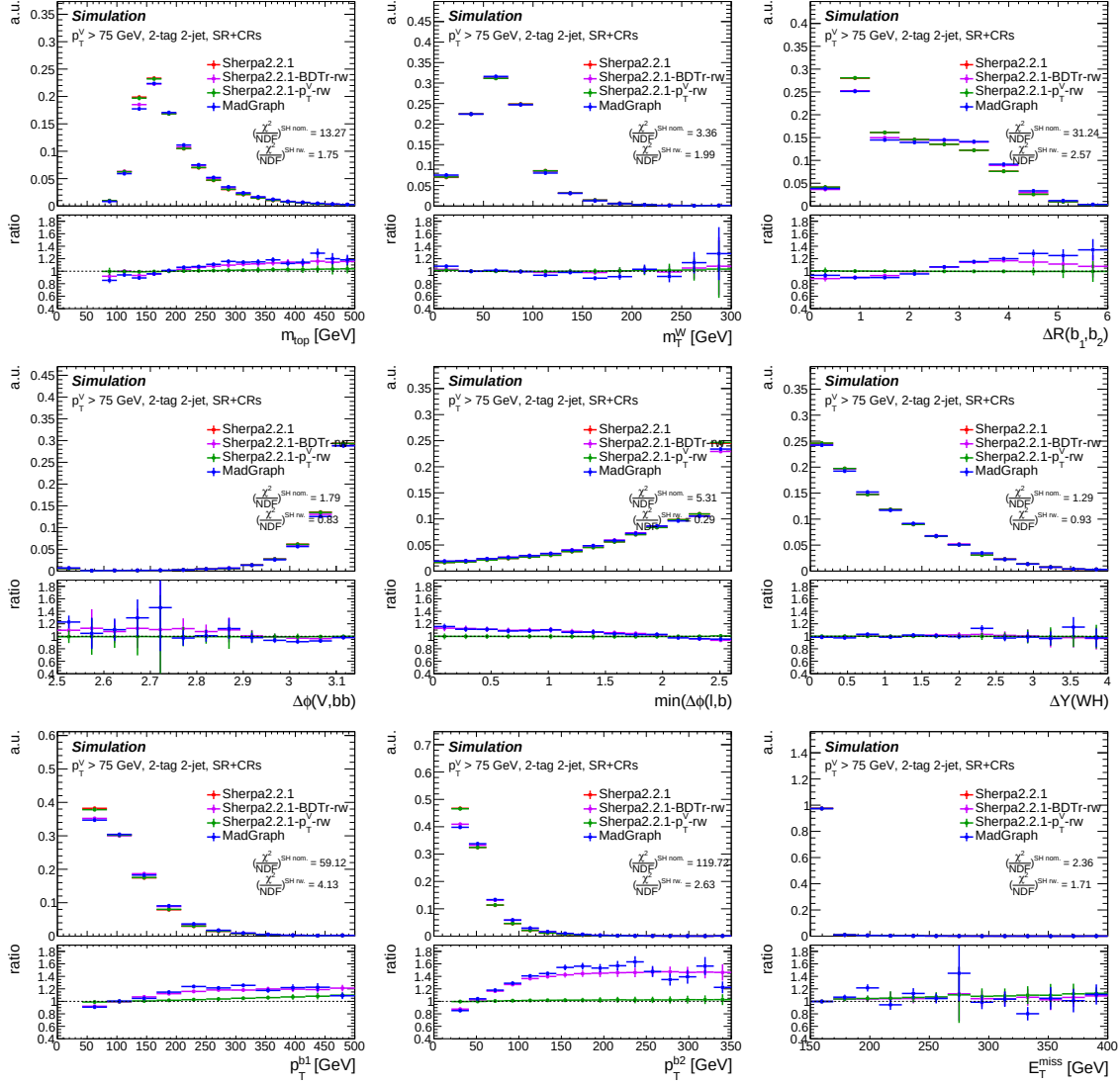


Figure B.81: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet reweighting of the 1-lepton channel and for $W + cc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

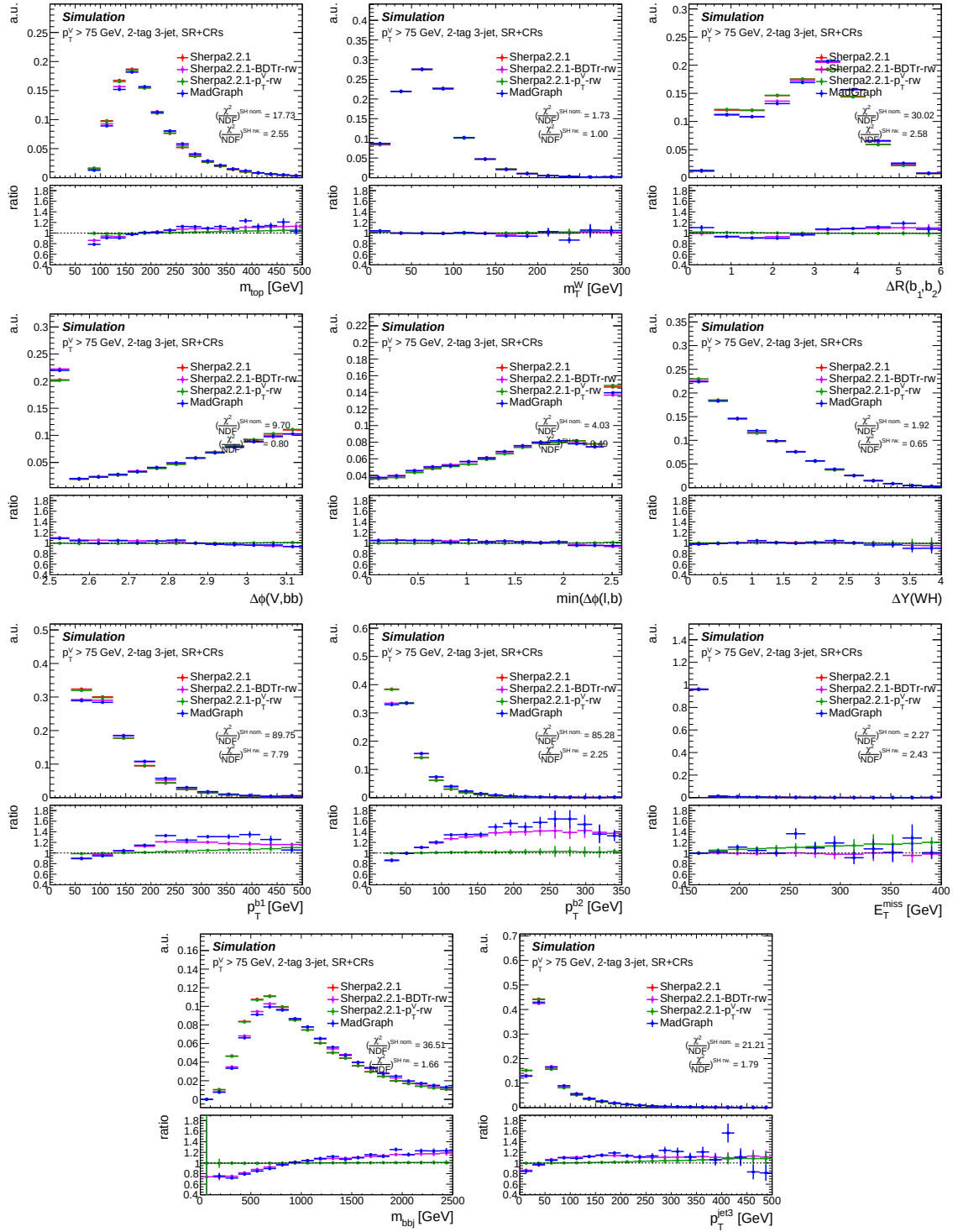


Figure B.82: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 1-lepton channel and for $W + bc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

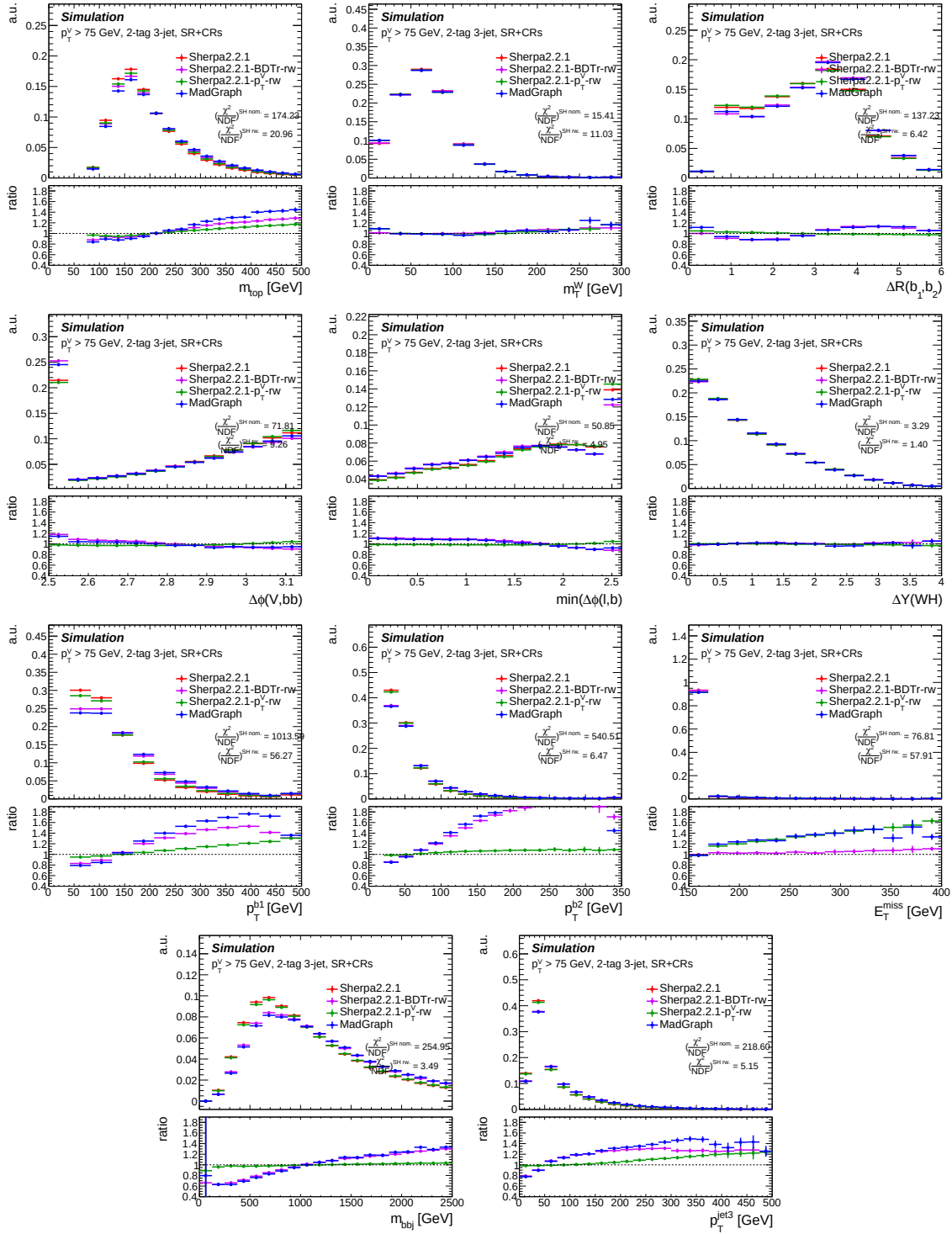


Figure B.83: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 1-lepton channel and for $W + bl$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

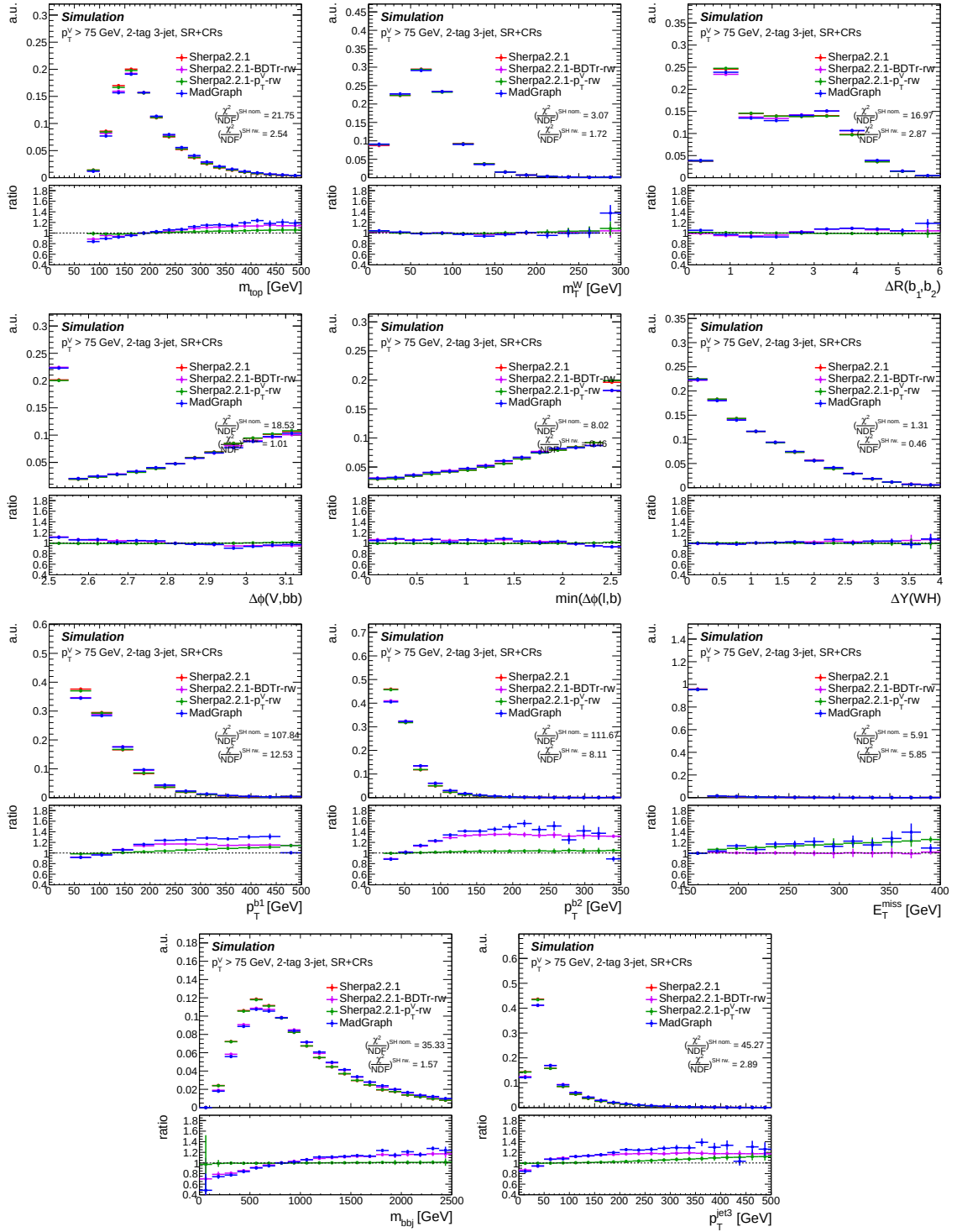


Figure B.84: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 1-lepton channel and for $W + cc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

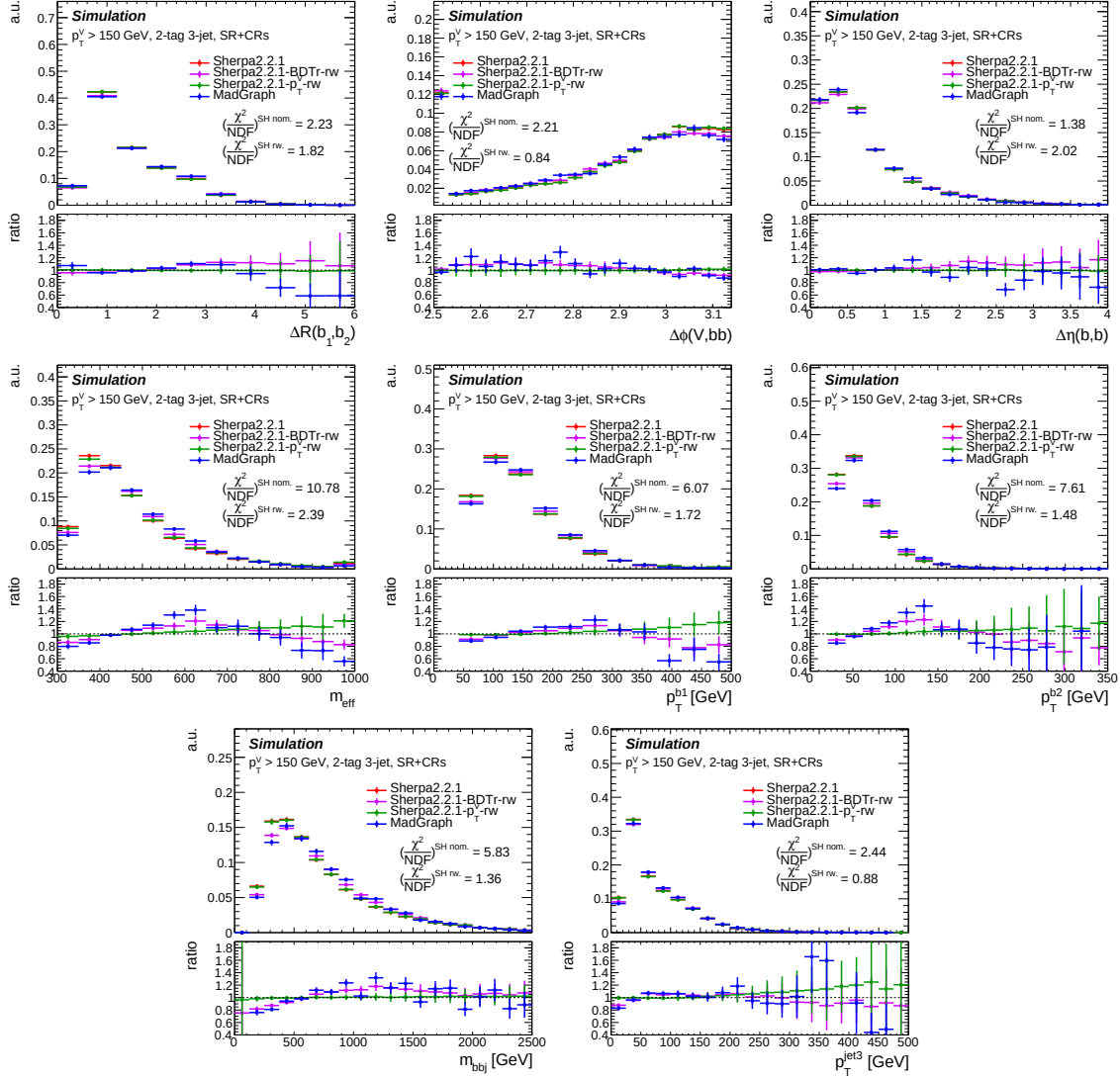


Figure B.85: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 0-lepton channel and for $W + bb$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is plotted in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

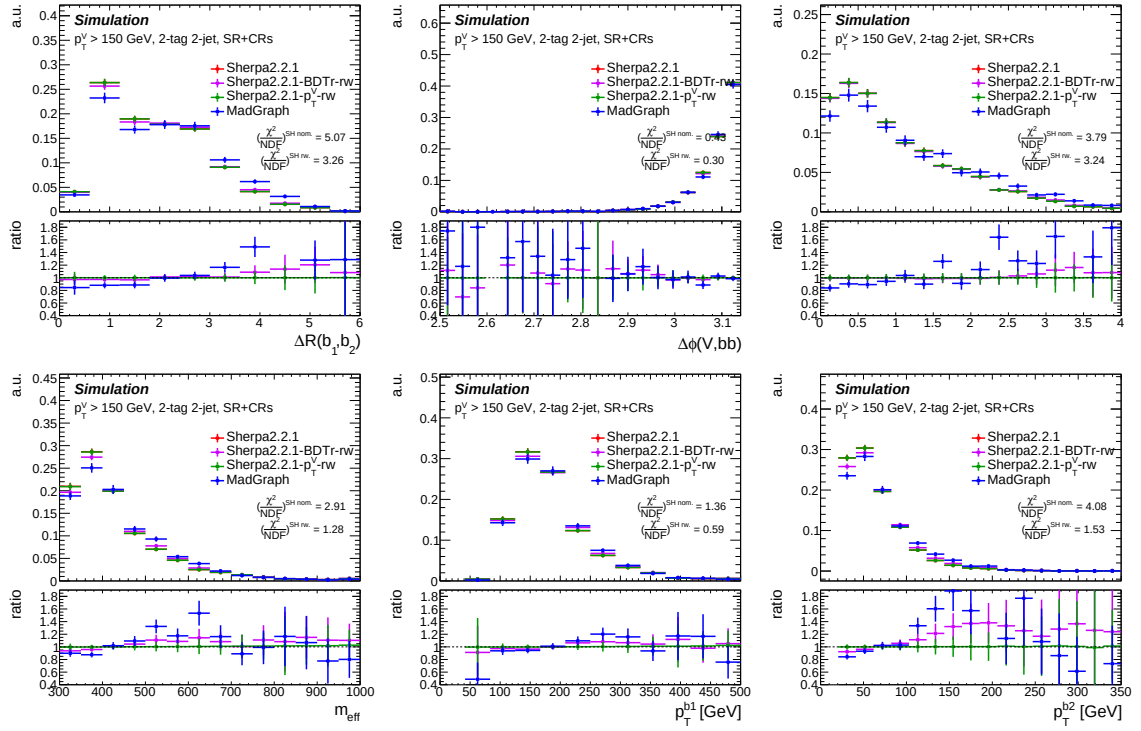


Figure B.86: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet region of the 0-lepton channel and for $W + bc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

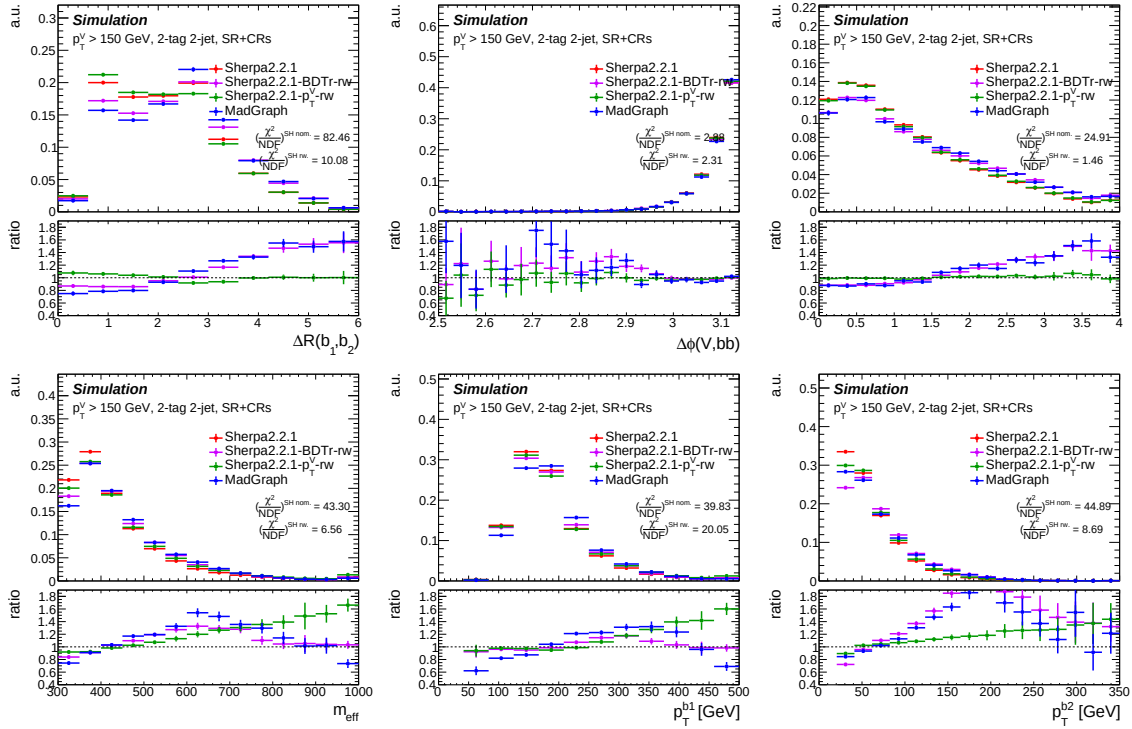


Figure B.87: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet region of the 0-lepton channel and for $W + b\ell$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f.}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

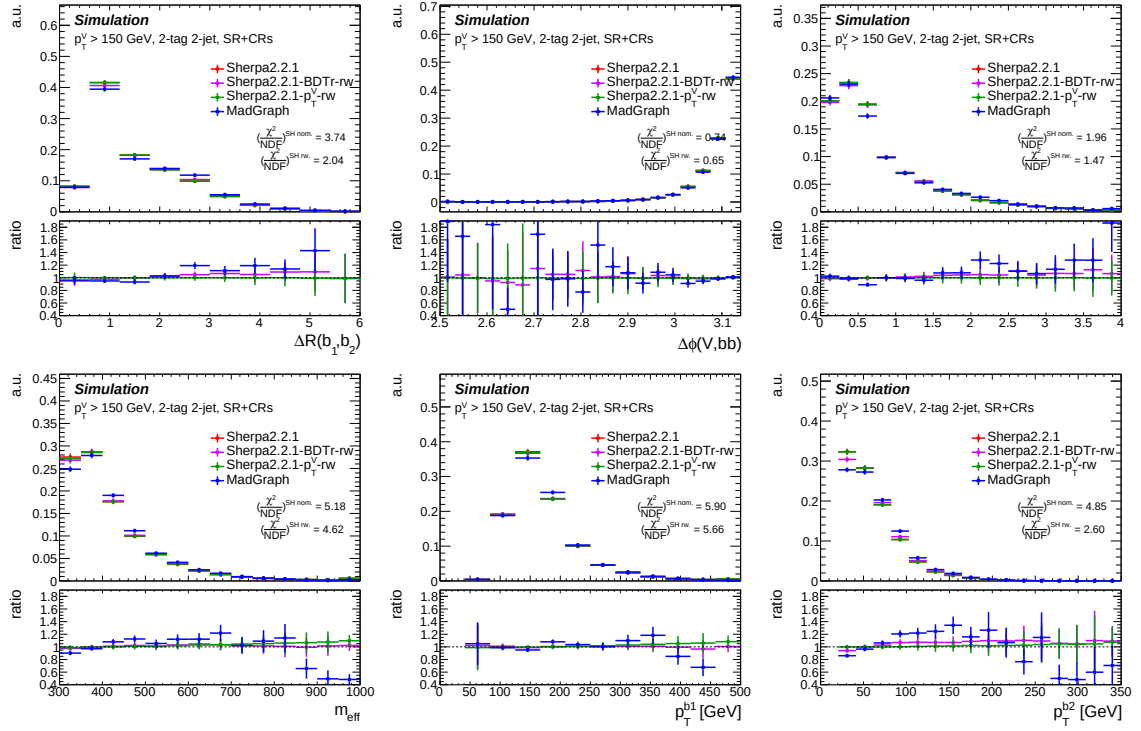


Figure B.88: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 2-jet region of the 0-lepton channel and for $W + cc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

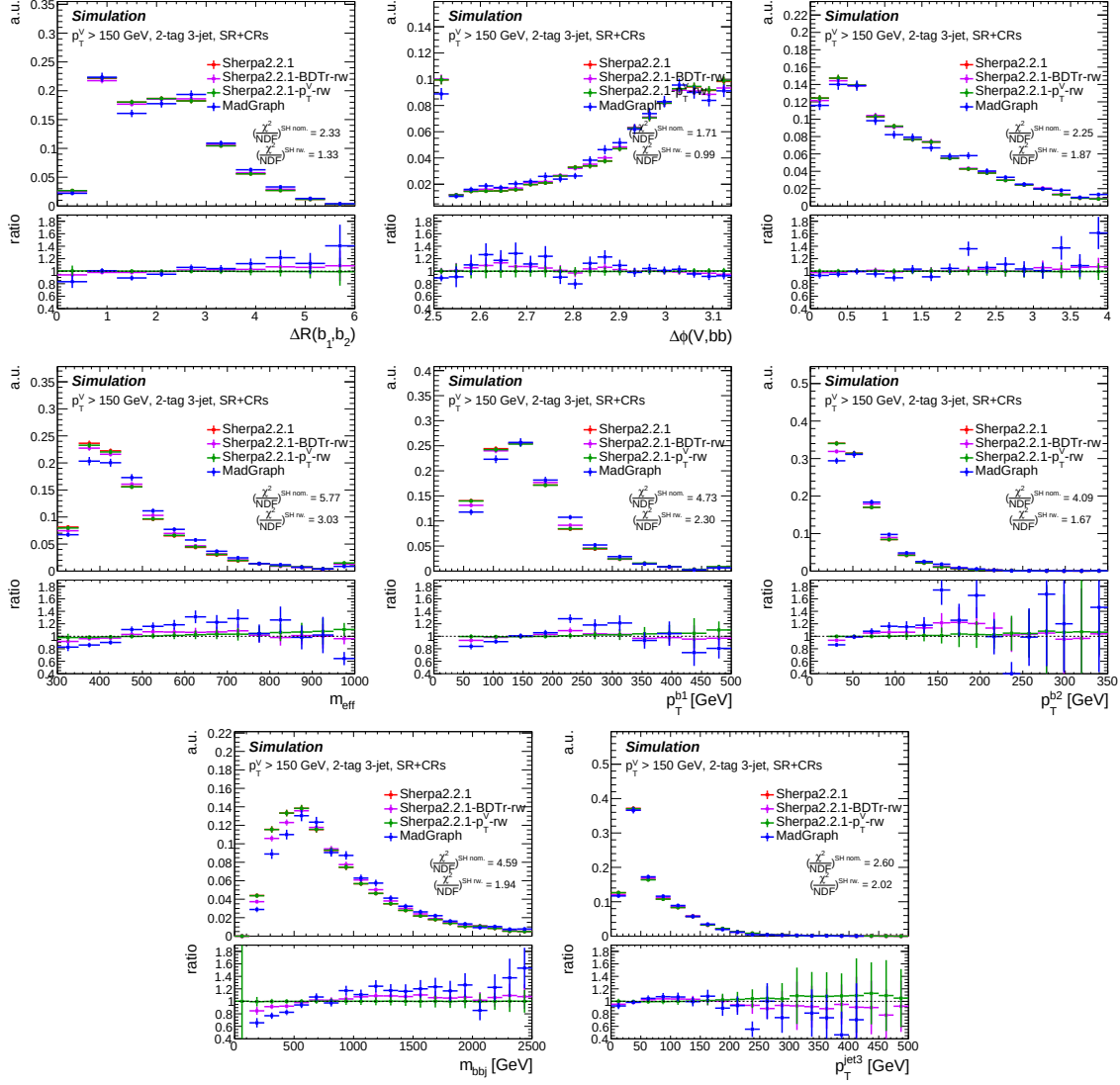


Figure B.89: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 0-lepton channel and for $W + bc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{\text{d.o.f.}}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

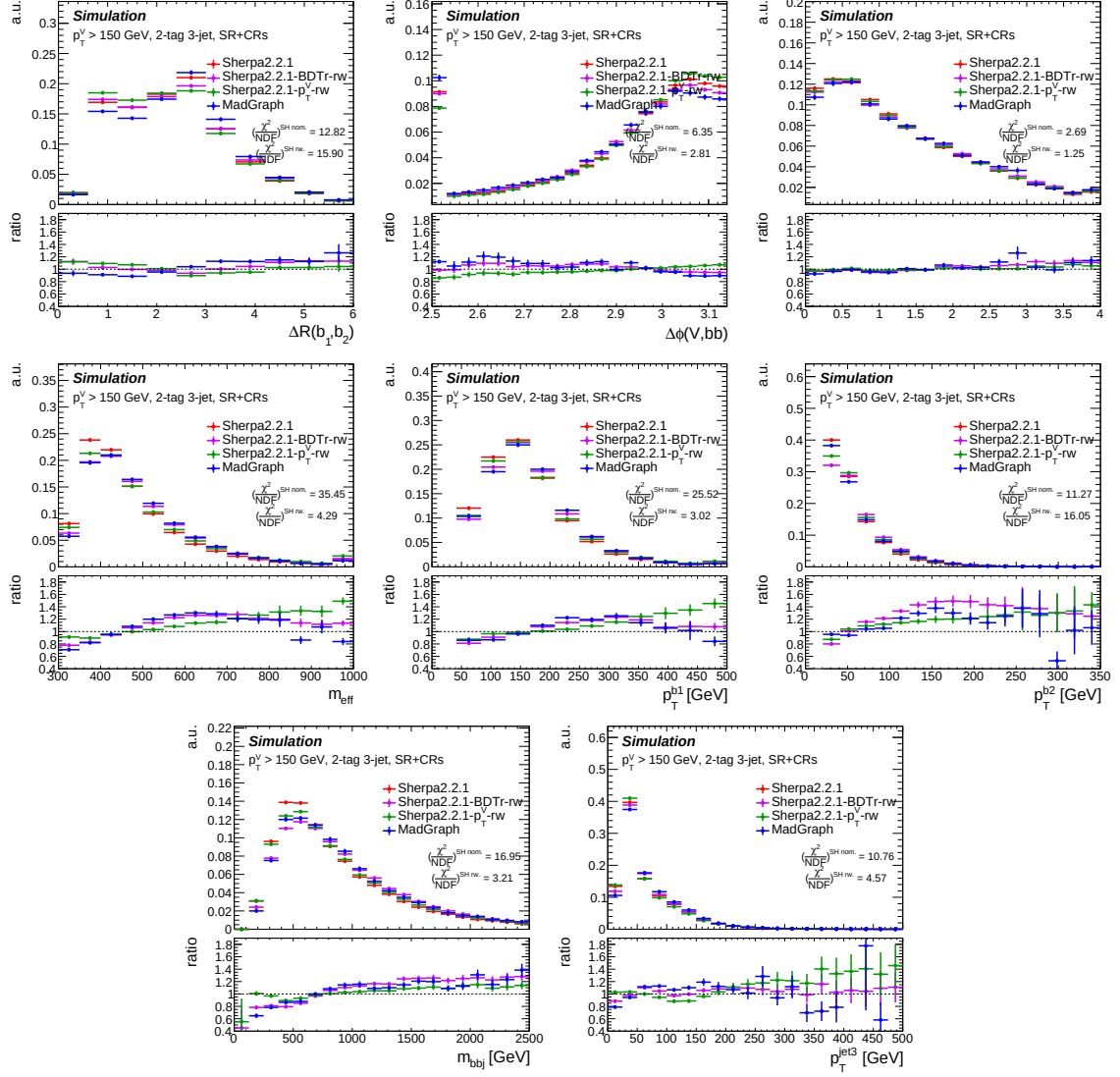


Figure B.90: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet region of the 0-lepton channel and for $W + bl$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

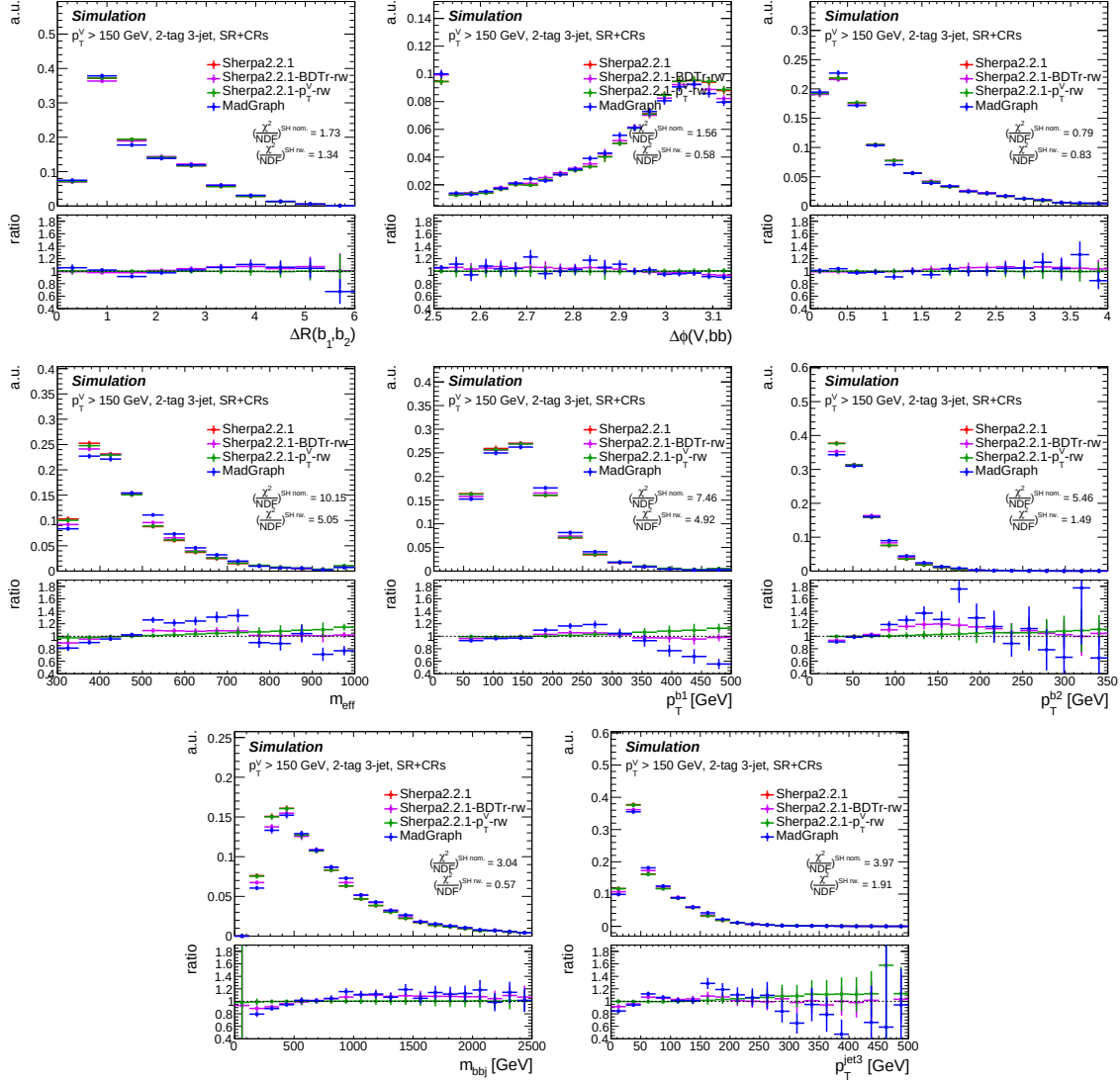


Figure B.91: Validation of the $p-2$ reweighting procedure on some of the variables entering the analysis-BDT in the 2-tag, 3-jet reweighting of the 0-lepton channel and for $W + cc$ events. The nominal SHERPA 2.2.1 prediction is shown in red, while the alternative MADGRAPH 5 one is shown in blue. The nominal distribution weighted by the p_T^V density ratio is plotted in green, while the one weighted by the BDTr density ratio is shown in magenta. The $\chi^2/n_{d.o.f}$ metric is evaluated to quantify the difference between the original nominal SHERPA 2.2.1 distribution and the MADGRAPH 5 one, as well as the difference between the nominal SHERPA 2.2.1 distribution weighted by the BDTr density ratio and the alternative one.

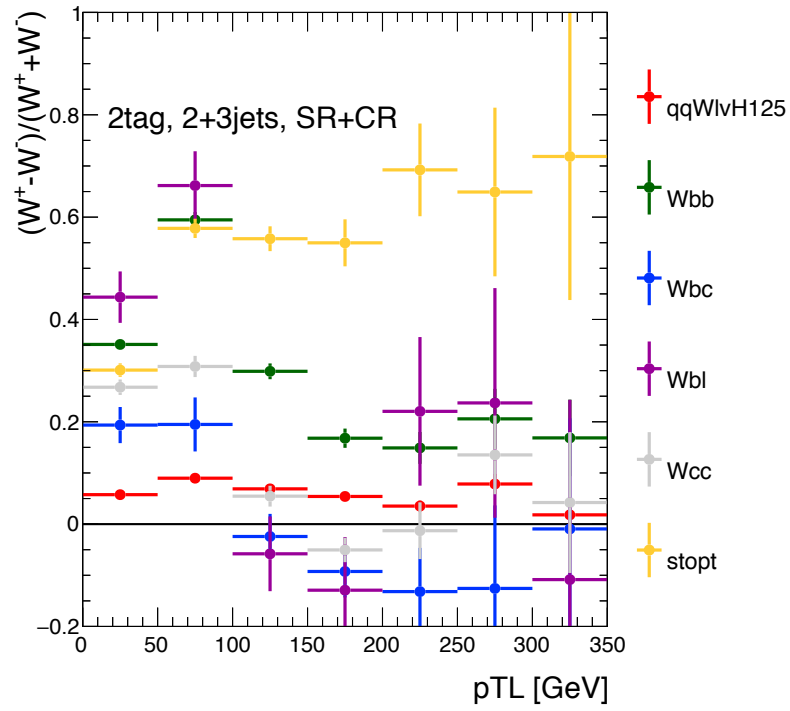


Figure B.92: The distribution of the quantity $\frac{W^+ - W^-}{W^+ + W^-}$ as a function of the transverse momentum of the lepton, p_{T}^L is shown for the signal and background processes surviving the $W^+ - W^-$ subtraction. The W +jets and single-top t -channel have the largest asymmetry.

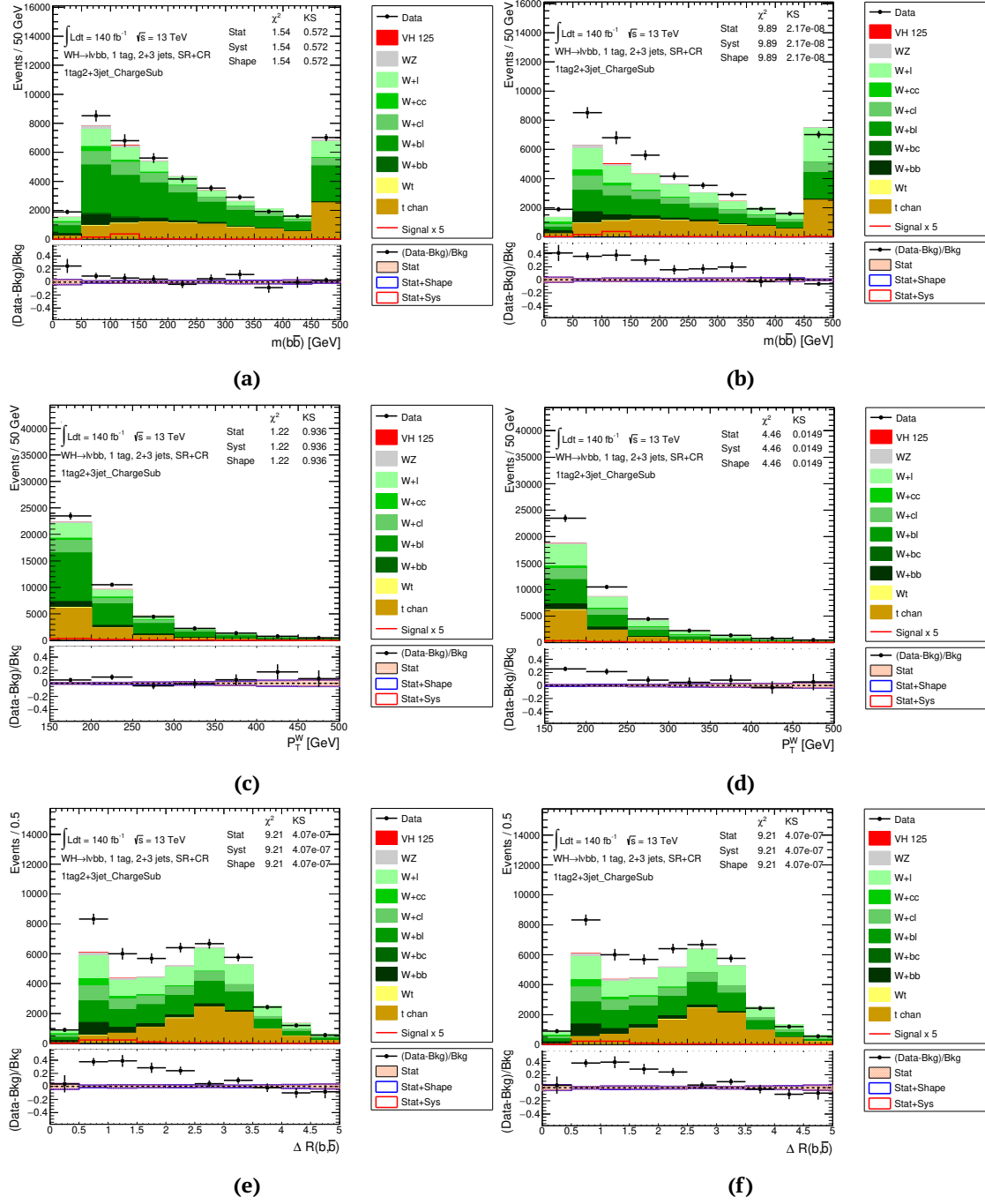


Figure B.93: The distributions of the m_{bb} (in (a) and (b)), p_T^V (in (c) and (d)) and ΔR_{bb} (in (e) and (f)) variables in the 1-lepton, 1-tag region, and inclusively in SR and CRs, p_T^V range, and jet categories for the $qq \rightarrow WH$ signal, the diboson WZ background, all flavour components of the W +jets background, the single-top t -channel, and the single-top Wt channel. In the first column, both signal and backgrounds are simulated using the nominal SHERPA 2.2.1 prediction, while on the second column they are simulated using the alternative MADGRAPH 5 prediction.

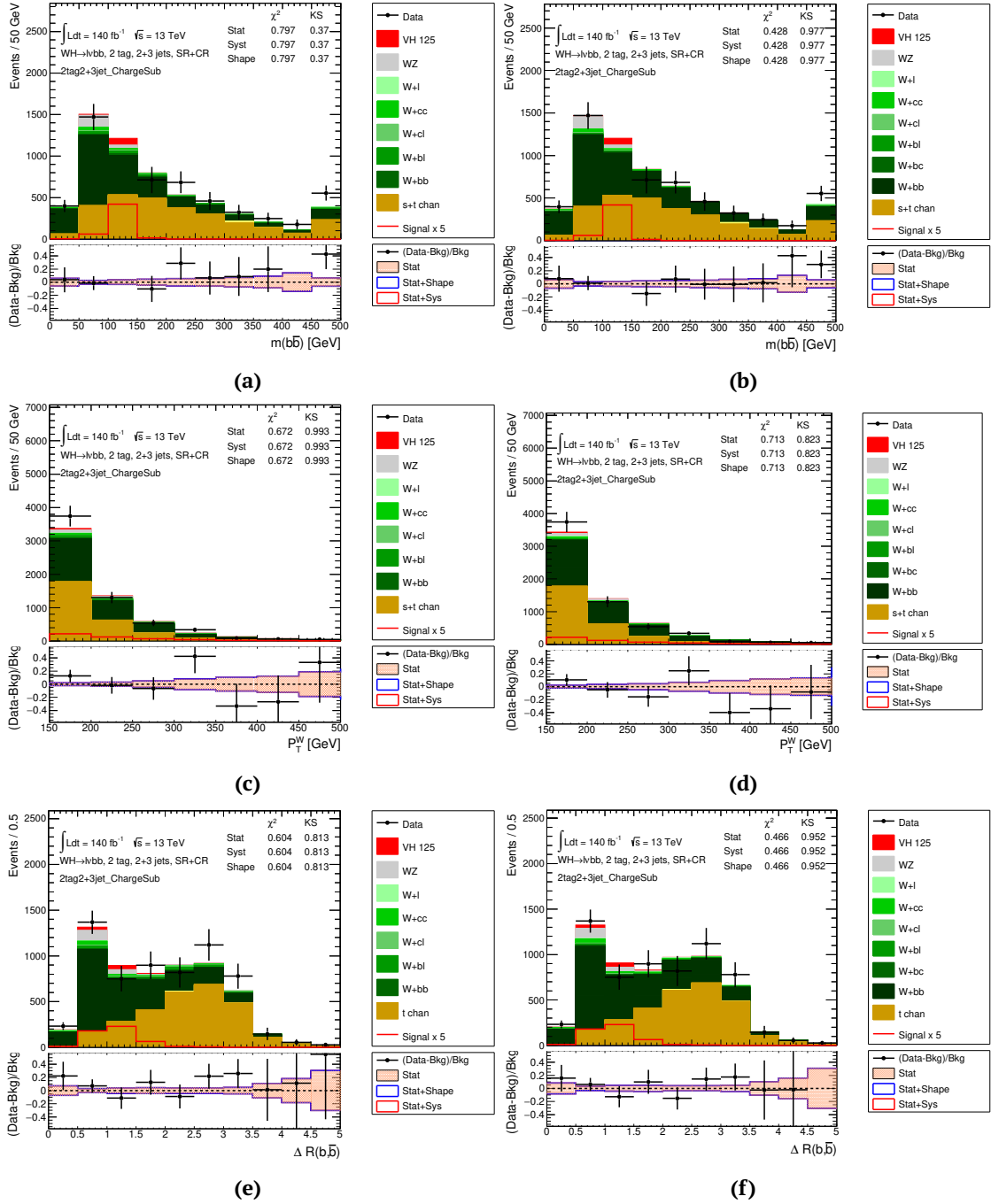


Figure B.94: The distributions of the m_{bb} (in (a) and (b)), p_T^V (in (c) and (d)) and ΔR_{bb} (in (e) and (f)) variables in the 1-lepton, 2-tag region, and inclusively in SR and CRs, p_T^V range, and jet categories for the $qq \rightarrow WH$ signal, the diboson WZ background, all flavour components of the W +jets background, the single-top t -channel, and the single-top Wt channel. In the first column, both signal and backgrounds are simulated using the nominal SHERPA 2.2.1 prediction, while on the second column they are simulated using the alternative MADGRAPH 5 prediction.

C Additional material for the $VH, H \rightarrow b\bar{b}$ fit results

C.1 $VH, H \rightarrow b\bar{b}$ Signal strength decorrelation studies

In Figure C.95 (a), the combined signal strength from the unconditional fit to the data is compared to the case where a signal strength parameter is left floating, independently, in the medium ($75 < p_T^V < 150$ GeV), high ($150 < p_T^V < 250$ GeV) and extreme ($p_T^V > 250$ GeV) p_T^V regions. A very good compatibility of 98% is found between the combined and the individual signal strengths measurements in each p_T^V category. Furthermore, the combined signal strength is compared to individual signal strength parameters left floating independently according to the p_T^V region and lepton channel (Figure C.95 (b)), the jet category (Figure C.95 (c)) and the lepton channel and jet category (Figure C.95 (d)). The compatibilities between the combined and the independent measurements are 99%, 30% and 90%, respectively.

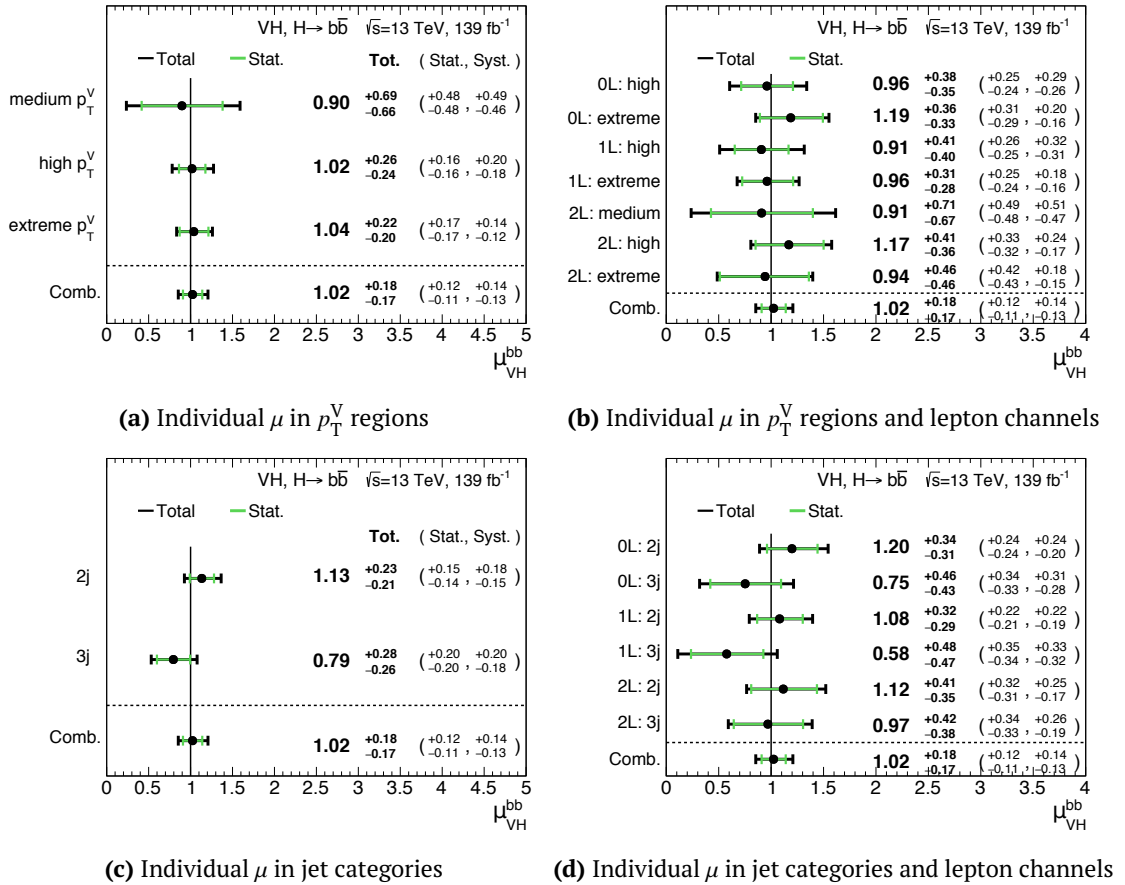


Figure C.95: Figure (a) shows the comparison between the μ value obtained from a fit to the three p_T^V regions combined, with the case where a signal strength is left floating independently for each medium, high and extreme p_T^V region. Figures (b), (c) and (d) show the cases where a signal strength parameter is left floating independently for each p_T^V region and lepton channel, each jet category and each jet category and lepton channel, respectively.

Acknowledgements

First of all, I would like to thank Prof. Dr. Karl Jakobs for giving me the opportunity of being a member of this amazing group. Much of my professional growth is due to the formative and stimulating environment I found in Freiburg. Thanks to Dr. Christian Weiser for patiently guiding me through these four challenging years. I really appreciated your willingness to share your knowledge and experience with me, even with a very busy schedule, because “*there must always be time to talk about physics*”. Thank you Chris, for being so patient and helpful and for solving every problem I had. You are amazing!

Thanks to the people that helped me in my everyday work. Valerio, for giving me the guidance and advice I needed from day one of my PhD and for always being there, no matter if day or night, weekday or weekend. Thank you because every time I felt lost in my work, the discussions we had (over coffee, or even better, over a crêpe) gave me the motivation to keep going. You are a great teacher.

Thanks to Stephen, for our long discussions, physics-related or otherwise, and for helping me with all technical issues I faced during my PhD.

Thanks to my friends in the group. Thanks Hannah for always giving me a word of support and for believing in me. Thanks to Frank, Théo, Arturo, Leena, Cedric, Sven, Julian, Philipp, Ralf, Moritz, Krzysztof, Liv, Frederik, Marc and Ines. I feel so lucky to have had you as colleagues.

Thank you to all the members of the $H \rightarrow b\bar{b}$ group I had the pleasure to work with. Thank you for the fruitful discussions and for the incredible commitment you have put in every day to achieve these beautiful results. Thank you to Francesco, for everything you taught me about b -tagging and for the help you gave me in all my “quasi-stable studies”. Thanks also to the whole Trackless b -tagging team, it was very inspiring to work on such a challenging topic with you.

Thank you Savona. When I got on the plane that 31 January four years ago, I was afraid that I wouldn’t feel at home in Freiburg, that it would have been difficult to make friends. I have never been so wrong! You are my family from all over the world. Thank you to Vlad for the sensibility you put in everything you do. To Stefano and Valentina for always welcoming me into your home as I were family, and to Edo for the belly laughs you give me! To Lisa and your always being so proactive, to Danny for sharing with me the same love for Rome, and to Thor and your determination. To Manfredi, for all the moments we shared in Freiburg and in Saint-Genis, and to Fabio for all the coffee breaks together. To Sun Won, for the sweet attentions you pay to each one of us. To Veronika, for sharing with me both the good and the difficult moments of this experience. Coffee is better with you! And to Michi, looking forward for a Kölsch in Wien! Thank you Costanza, for always listening to me during one of the most challenging periods of my PhD, and for feeding

me with pizza, curcuma bread, and many other Pino's creations! Thanks to Giac(h)omino for your cheerfulness. I really enjoyed our time as flatmates! Thank you Simone, leaving Rome to move to Freiburg was very difficult, but having a friend to share the moment with made it easier. Thank you Riccardo, for being the beating heart of this wonderful group. Your genuineness and energy continue to inspire me. And thanks to Maria, your cheerfulness and spontaneity always brings a smile to my face. Thank you to Giulia and Tazio, because your decision to welcome me into your home four years ago marked the beginning of a really precious friendship. To Giulia, I will never stop surprising me for how similar we are! Thank you for sharing with me every joy and pain of this journey, and for giving me all the support and understanding I needed, whenever I needed it. To Tazio, I never imagined that a nickname as ugly as *Sgargiul* could become so much a part of me. Thank you because every laugh you've given me has made both the good and the bad days better. Thank you Peter, for cheering me up every time I needed it and for always having the right word and the useful advice to keep me going. My time at CERN would not have been the same without you.

Grazie a tutte le persone con le quali ho condiviso la mia esperienza al CERN e a Ginevra. Grazie Bibi, per la tua dolcezza e la tua simpatia. Ci sono quelli che amano le tue battute brutte e quelli che mentono. Grazie Giuliano, Giulia, Arianna, Daniele, Valentina, Frankie, Camilla, Adele, Marco, Serena, Alessandra, Filippo, Maria Giovanna, Mastro, Betto, Grazia, Luigi, Umberto, Francesco, Fefe, Marco, Chicco, Gioacchino e Nello. Grazie al gruppo del Cinefellow, per i tanti film visti insieme e le passeggiate in montagna.

Grazie agli amici di una vita, per avermi aspettato a braccia aperte ogni volta che sono tornata a Roma. Grazie Eleonora, per la tua dolcezza, la tua incredibile forza, e per il tuo essere sempre presente da più di 20 anni. Grazie Federica e Eugenia, perchè dopo tanti anni e anche se distanti, non ci siamo mai perse. Grazie Elisa, per le nostre lunghissime chiacchierate, per i tuoi preziosi consigli, e per il tuo costante supporto. Grazie Matteo, Eleonora, Giulio, Simone, Stefano e Veronica, le serate giochi con voi hanno alleggerito i miei mille pensieri!

Grazie mamma e papà, per aver sempre supportato le mie scelte, anche quando queste mi hanno portato lontano da voi. Grazie per avermi fatto sentire la vostra vicinanza ogni giorno di questi quattro lunghi anni. Grazie Chiara, per avermi trasmesso un po' della tua forza e della tua tenacia. Grazie Andrea, per farmi sentire sempre a casa. E grazie ad entrambi, per avermi regalato la gioia più grande, il mio piccolo Alessandro. In quest'anno così difficile, sei stato il mio pensiero felice e vedere il tuo sorriso ha reso tutto più semplice. Questa tesi è dedicata a te, chissà che un giorno potremo leggerla insieme!

Ed infine, grazie a te Luca. Che hai condiviso con me tutte le gioie di questa esperienza e hai portato sulle tue spalle le mie insicurezze. Se sono arrivata fino in fondo a questo percorso lo devo anche alla forza con la quale ogni giorno mi hai spinto a credere in me stessa. E ora che siamo finalmente insieme, non vedo l'ora di condividere con te tutto quello che verrà!