Flavour tagging and measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel with the ATLAS detector

Ava C. A. Lee

supervised by

Dr. Tim Scanlon

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

at

University College London

Department of Physics and Astronomy

University College London

December 24, 2022

Declaration

I, Ava Lee, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

A measurement of the Standard Model Higgs boson decaying to a pair of bottom quarks ($H \to b\bar{b}$) and produced in association with a vector boson, VH, at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is presented. The measurement is performed on data corresponding to an integrated luminosity of 139 fb^{-1} that was collected using the ATLAS detector during Run-2 of the Large Hadron Collider. Studies have been conducted to improve the multivariate analysis, which is used to separate the $H \to b\bar{b}$ signal from the background. Uncertainties in the statistical fit used to measure the signal strength are explored. The observed significance of the $VH, H \to b\bar{b}$ signal is 6.7σ. The observed signal significances of the WH and ZH production processes are 4.0σ and 5.3σ, respectively, indicating strong evidence of WH production and the first observation of ZH production. In addition, measurements of the cross-sections of the WH and ZH production processes with $H \to b\bar{b}$ decay as a function of the gauge boson transverse momentum are consistent with Standard Model predictions. Studies involving enhancements to tracking and b-jet identification, which are key components of the $H \to b\bar{b}$ analysis, are also presented. These enhancements include investigating different track reconstruction improvements, enhancing the machine learning algorithms used to identify the different jet flavours, and exploring various automation techniques in the calibration of the b-jet identification efficiency.

Acknowledgements

First of all, I would like to thank my supervisor, Tim Scanlon, for his invaluable support and guidance throughout my PhD, and without whom this thesis would not have been possible. I express my gratitude to the members of the ATLAS Flavour Tagging, $H \to b\bar{b}$ and CTIDE groups. In particular, I thank Sébastien Rettie, Yanhui Ma, Elisabeth Schopf and Thomas Calvet for their assistance. I am also grateful to many members of the UCL CDT DIS and HEP groups who have made the past few years more enjoyable and worthwhile. Finally, I am indebted to my friends and family, especially to Mummeh, Pusheen and Bobby, for their continued support and patience through the ups and downs in this period of my life.

Impact statement

The Large Hadron Collider probes the fundamental structure of particles to advance our knowledge about how the Universe works on the most fundamental level. Although this knowledge does not always have a direct impact, application of this knowledge can have an untold and significant impact further in the future.

The work presented in this thesis has contributed to precision measurements of the Higgs boson decaying to a pair of bottom quarks, which is the most likely decay of the Higgs boson and makes it the ideal channel in the search for deviations in the production rate of the Higgs boson. Machine learning techniques demonstrated in the thesis are critical to the knowledge economy and are employed in a wide range of industries, such as healthcare, biology, finance, insurance, and retail. Furthermore, the work conducted at the Large Hadron Collider is widely publicised and provides numerous positive societal benefits, which involves generating excitement about STEM and attracting people into this research area.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The discovery of the Higgs boson in 2012 by the ATLAS [1] and CMS [2] experiments at the Large Hadron Collider (LHC) [3] was a major achievement in particle physics that completed the Standard Model (SM). Since the discovery, measurements of the properties of the Higgs boson have been conducted with increasing precision, which is crucial to finding potential divergences from the SM in the search for new physics. By the end of Run-1 of the LHC (2009 - 2013), the Higgs boson decaying to pairs of W bosons [4], Z bosons [5, 6], photons [7, 8], and τ-leptons [9] have been observed and were found to be consistent with SM predictions. Due to the higher luminosity achieved during Run-2 (2015-2018), it was possible to directly measure the coupling of the Higgs boson to the top quark, which is an up-type quark, in the observation of the Higgs boson produced in association with a pair of top quarks [10, 11].

The decay of the Higgs boson to a pair of bottom quarks ($H \to b\bar{b}$) can also be used to directly probe the coupling of the Higgs boson to down-type quarks. Furthermore, the $H \to b\bar{b}$ decay mode has a predicted branching ratio of 58%, hence it is the dominant decay mode of the Higgs boson and plays an important part in determining the extent to which the SM is accurate. The dominant production mode of the Higgs boson is via gluon fusion, however, for the $H \to b\bar{b}$ final state, the extremely large multi-jet $b\bar{b}$ background results in this production channel not being best suited to observe this decay. The associated production of the Higgs boson with a vector boson (VH) is used in the search for $H \to b\bar{b}$ as the leptonic channels of the vector boson provide clean signatures that can be used to reject the large hadronic background. As b-quarks do not exist as free particles and undergo fragmentation to form jets, the identification of jets that originated from b-quarks, known as b-tagging, is important in reconstructing the pair of quarks in the final state of Higgs boson decay.

The leptonic signatures of the $VH, H \to b\bar{b}$ channel were exploited at the Teva-

tron [12], where the results from the CDF and DØ experiments were combined to give the first evidence of the $VH, H \to b\bar{b}$ signal. At the LHC, the first evidence of $VH, H \to b\bar{b}$ was provided in 2017 and 2018 by the ATLAS [13] and CMS [14] experiments, respectively. The $VH, H \to b\bar{b}$ signal was first observed in 2018 by both experiments [15, 16] using Run-1 and Run-2 (up to 2017) data. For the observation by ATLAS, the total integrated luminosity of the data used was 79.8 fb^{-1}. For the full Run-2 period, the total integrated luminosity of the data recorded by ATLAS is 139 fb^{-1}, which was used to analyse the $VH, H \to b\bar{b}$ channel [17] in this thesis.

As the instantaneous luminosity produced by the LHC increases, the amount of data at the LHC is rapidly growing, which requires big data techniques to ensure that data is efficiently processed and analysed. These big data techniques include the application of machine learning in b-tagging and signal-background discrimination in analyses to benefit from the predictive power of classification algorithms. In addition, improving software algorithms and incorporating software engineering practices, such as in the form of automation, can improve performance and reproducibility, and provide a time-saving solution to scientific research, especially within large experimental collaborations. Improvements to the tracking algorithms have been studied in the tracking group, automation techniques have been implemented in the flavour tagging group, and machine learning tools have been studied in both the $VH, H \to b\bar{b}$ analysis and flavour tagging groups.

The author's direct contribution to the full Run-2 $VH, H \to b\bar{b}$ analysis results include producing the $VH, H \to b\bar{b}$ 1-lepton samples, maintaining the analysis software framework and produced the 1-lepton inputs to the statistical fit. The author was also the Electron 'Combined Performance' contact for the software framework for two years. The author conducted several multivariate analysis (MVA) studies in the 1-lepton channel, which are documented in Chapter 11, and proposed improvements that resulted in the final configuration of the Run-2 MVA that was adopted for the full results in all the analysis channels. The author investigated uncertainties relating to a dominant background, $t\bar{t}$, resulting in the final systematic scheme, and conducted numerous studies into the robustness of the statistical fit, some of which are presented in Chapter 13. The author automated the b-tagging efficiency calibration software, which is detailed

in Chapter 9 and helped other analysis groups automate their calibration software. Furthermore, the author analysed the effect of different track collections on a low-level b-tagging algorithm, as discussed in Chapter 7, and conducted optimisation studies for a high-level b-tagging algorithm, which are presented in Chapter 8.

Part I

Theoretical and experimental framework

Chapter 2

Theoretical framework

The Standard Model (SM) [18, 19] describes the fundamental particles of the Universe and their interactions. The model has been studied extensively in experiments with particle accelerators, such as the Large Hadron Collider (LHC), and is considered one of the triumphs of modern physics due to its remarkable agreement with experimental data.

2.1 The Standard Model

The fundamental particles of the SM are illustrated in Fig. 2.1. The interaction of particles is governed by the fundamental forces: strong, weak, electromagnetic, and gravitational. As the mathematical framework has not been developed to include the effect of the gravitational force in the SM, the SM only involves the strong, weak, and electromagnetic forces.

The mathematical framework of the SM is provided by quantum field theory, in which particles are described in terms of dynamical quantum fields and interactions are governed by the local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry. Furthermore, the SM is established from Lagrangian formalism.

Starting from the Lagrangian for the Dirac equation:

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

where ψ is the Dirac spinor field, γ^μ is the Dirac γ-matrices, m is the mass of the fermion, and $\bar{\psi} = \psi^\dagger \gamma^0$. The symmetry of the Dirac Lagrangian is required to be invariant under

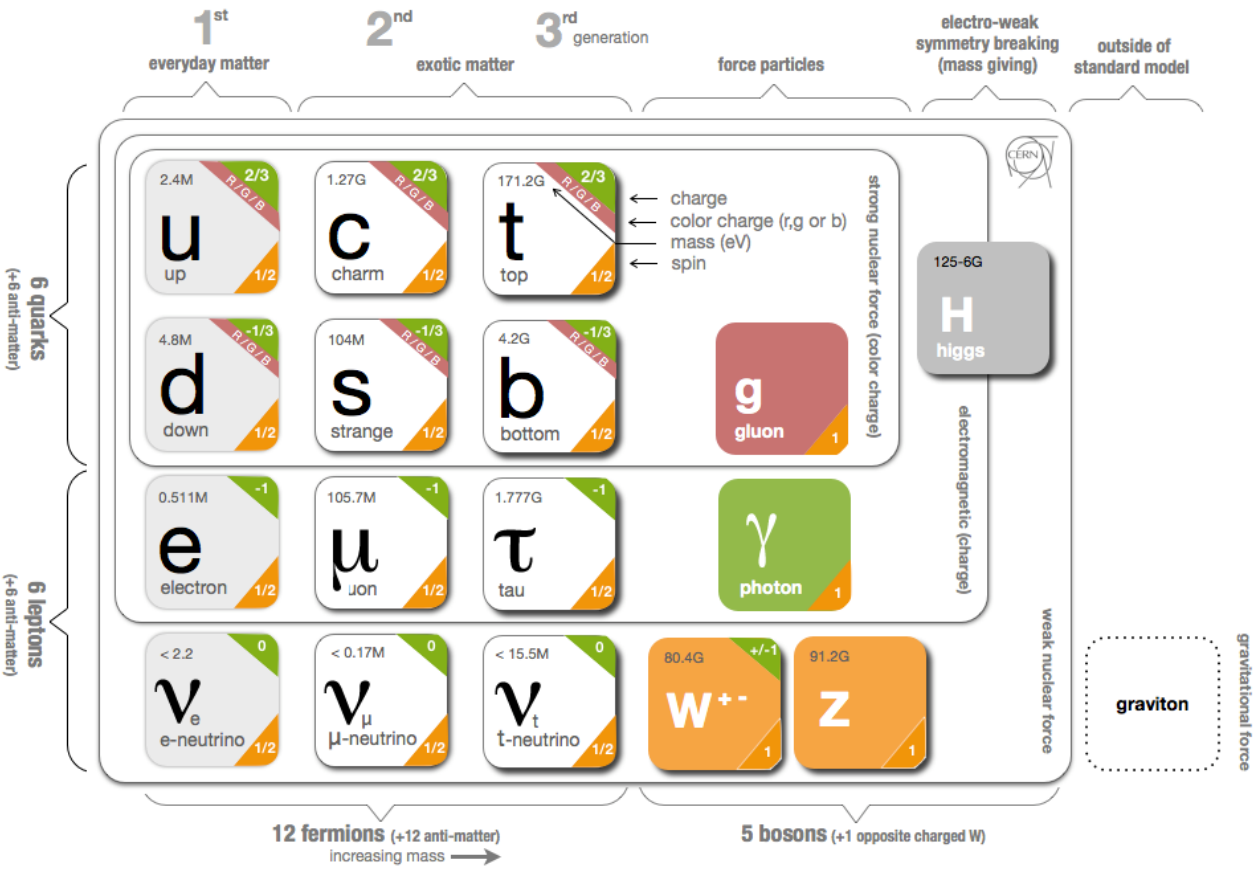


Figure 2.1: Infographic of the Standard Model particles [20]. The mediators of the different interactions are denoted by force particles. It can be seen that neutrinos interact only through the weak force, charged leptons interact via the electromagnetic and weak forces, and quarks can interact through all three of the fundamental forces in the Standard Model.

the $U(1)$ gauge transformation, which is defined as

$$\psi(x) \rightarrow e^{i\alpha(x)}\psi(x), \tag{2.2}$$

where $\alpha(x)$ is an arbitrary phase that is a function of space-time x. As $\partial_\mu\psi$ does not respect the $U(1)$ symmetry, ∂_μ is replaced with the covariant derivative:

$$D_\mu = \partial_\mu - ieA_\mu, \tag{2.3}$$

where A_μ is a vector field, to ensure the requirement of the $U(1)$ gauge invariance.

By incorporating the electromagnetic field strength tensor, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, into the Dirac Lagrangian, an additional term that represents the kinetic energy of the photon is defined, resulting in the Lagrangian of quantum electrodynamics (QED):

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.4}$$

where A_μ can now be regarded as the field of the massless photon.

The weak and electromagnetic (EM) forces are unified as the electroweak (EW) interaction through the invariance of the $SU(2)_L \times U(1)_Y$ gauge transformation in the Glashow-Salam-Weinberg model [21–23]. The weak isospin, $\boldsymbol{T} = \frac{1}{2}\boldsymbol{\sigma}$, which is expressed in terms of the three Pauli matrices $\boldsymbol{\sigma}$, are the generators of the $SU(2)_L$ transformation. Fermions with total weak isopin of $T = \frac{1}{2}$ are known as left-handed fermions, while right-handed fermions have total weak isospin of $T = 0$. For the $U(1)_Y$ transformation, the generator is the weak hypercharge, $Y = 2(Q - T)$, where Q is the EM charge. The gauge field B_μ is required to satisfy the $U(1)_Y$ gauge symmetry. To impose the $SU(2)_L$ gauge invariance, three gauge fields, W_μ^i for $i = 1, 2, 3$, are introduced. Linear combinations of the fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \tag{2.5}$$

give rise to the charged bosons $W^\pm$, which leaves the weak neutral field W_μ^3 to be accounted for. Out of the observed bosons in QED and the weak interaction, the photon and Z boson are both neutral. As the subscript L of the $SU(2)_L$ transformation indicates that the coupling of these gauge fields is only to left-handed fermions, W_μ^3 cannot be associated to the Z boson, which couples to both left-handed and right-handed fermions. Therefore, the W_μ^3 and B_μ fields are combined to produce the corresponding fields of the photon:

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W, \tag{2.6}$$

and the Z boson:

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{2.7}$$

where θ_W is the weak mixing angle.

The description of the strong interaction, quantum chromodynamics (QCD), can be

deduced from its local gauge invariance by replacing the $U(1)$ transformation with its $SU(3)_C$ gauge transformation of the three quark colour fields, which represent the three colour changes. The $SU(3)$ transformation contains eight generators, each of which is a 3×3 matrix. Local gauge invariance on the QCD Lagrangian is applied by the introduction of eight gauge fields (one for each generator), which correspond to the eight massless gluons that mediate the strong force.

Due to the local gauge invariance of the SM Lagrangian, the addition of mass terms for gauge bosons is forbidden, meaning that all gauge bosons are massless. While this is in agreement with the gauge bosons of QED and QCD, it is in contradiction to the gauge bosons of the electroweak interaction, where experimental observations have revealed large masses of W and Z bosons. Hence, a mechanism to explain the generation of mass for gauge bosons is introduced.

2.1.1 The Higgs Mechanism

In the Brout-Englert-Higgs mechanism [24, 25], spontaneous symmetry breaking is induced through particle interactions with the Higgs field, resulting in the masses of particles. Specifically, the gauge boson masses are generated from spontaneously breaking the $SU(2)_L \times U(1)_Y$ symmetry in the EW sector.

The Higgs field is constructed from four real scalar fields, ϕ_i for $i = 1, 2, 3, 4$, which are arranged in an isospin doublet with $T = \frac{1}{2}$ and $Y = 1$, known as the Higgs doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.8}$$

The Lagrangian associated with the Higgs field is

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \tag{2.9}$$

with the Higgs potential,

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{2.10}$$

where μ and λ are constants, and $\lambda > 0$. The vacuum state corresponds to the minimum of the potential and is the lowest energy state of the field ϕ. If $\mu^2 > 0$, the potential has

a minimum at $\phi = 0$. If $\mu^2 < 0$, the potential has minima at

$$\phi = \pm v = \pm \left| \sqrt{\frac{-\mu^2}{\lambda}} \right|, \tag{2.11}$$

indicating that the field has a non-zero vacuum expectation value v with two possible degenerate vacuum states. The symmetry of the Lagrangian is broken by the choice of the vacuum state, in the process known as spontaneous symmetry breaking. Setting $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3^2 = -\frac{\mu^2}{\lambda} = v^2$ yields the vacuum:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.12}$$

Fields can be expanded around this vacuum state as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix}. \tag{2.13}$$

Spontaneous symmetry breaking leads to a massive scalar boson and three massless Goldstone bosons associated to $\eta(x)$ and $\phi_1(x), \phi_2(x), \phi_4(x)$, respectively. In the unitary gauge, the fields are transformed such that the three degrees of freedom associated with the Goldstone bosons are absorbed by the W/Z bosons and the only scalar field that remains is the Higgs field $h(x)$ associated to the massive scalar boson, the Higgs boson H, giving the Higgs doublet:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.14}$$

The mass terms can be determined by replacing the partial derivative in the Lagrangian of (2.10) by the covariant derivative:

$$D_\mu = \partial_\mu + ig\boldsymbol{T} \cdot \boldsymbol{W_\mu} + ig' \frac{Y}{2} B_\mu, \tag{2.15}$$

where $\boldsymbol{W_\mu} = (W_\mu^1, W_\mu^2, W_\mu^3)$, g and g' are the coupling constants of $\boldsymbol{T}$ to $\boldsymbol{W_\mu}$ and Y to B_μ, respectively. The gauge bosons masses are obtained by terms that are quadratic

in the gauge boson fields ($W^{\pm}$, Z and h) in the Lagrangian evaluated with the Higgs doublet in the unitary gauge and the covariant derivative:

$$\begin{aligned}\mathcal{L}_{Higgs} = &\frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{g^2}{4}(W^+_\mu W^{-\mu})(v+h)^2 \\ &+ \frac{g^2+g'^2}{8}(Z_\mu Z^\mu)(v+h)^2 + \frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{16}(v+h)^4,\end{aligned} \tag{2.16}$$

where the field of the Z boson is defined as

$$Z_\mu = \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2+g'^2}}, \tag{2.17}$$

yielding the mass terms of the W boson:

$$m_W = \frac{gv}{2}, \tag{2.18}$$

Z boson:

$$m_Z = \frac{v\sqrt{g^2+g'^2}}{2}, \tag{2.19}$$

and the Higgs boson:

$$m_H = \sqrt{2\mu^2}, \tag{2.20}$$

and the photon remains massless.

2.1.1.1 Yukawa interaction

Fermions can also interact with the Higgs field through the Yukawa interaction, giving rise to the masses of fermions. The Lagrangian for fermions that satisfies the $SU(2)_L \times U(1)_Y$ symmetry is

$$\mathcal{L}_{fermions} = -g_f(\bar{L}\phi R + \bar{R}\phi^\dagger L), \tag{2.21}$$

where g_f is the Yukawa coupling of the fermion f to the Higgs field, L is a left-handed fermion doublet, R is a right-handed fermion singlet, and ϕ is the Higgs doublet. For leptons $l = e, \mu, \tau$ with neutrinos ν_l, the Lagrangian is expanded as

$$\mathcal{L}_{fermions} = -g_l \left[\begin{pmatrix} \bar{\nu_l} & \bar{l} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} l_R + \bar{l_R} \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L \right]. \tag{2.22}$$

After spontaneously symmetry breaking, the Lagrangian can be evaluated with the Higgs doublet in the unitary gauge:

$$\mathcal{L}_{fermions} = -\frac{g_l}{\sqrt{2}} v(\bar{l_L} l_R + \bar{l_R} l_L) - \frac{g_l}{\sqrt{2}} h(\bar{l_L} l_R + \bar{l_R} l_L). \tag{2.23}$$

Defining the Yukawa coupling between the lepton and the Higgs field as

$$g_l = \sqrt{2}\frac{m_l}{v}, \tag{2.24}$$

where m_l is the mass of the lepton, the Lagrangian becomes

$$\mathcal{L}_{fermions} = -m_l \, \bar{l} \, l - \frac{m_l}{v} \, \bar{l} \, l \, h, \tag{2.25}$$

where the first term provides the mass of the lepton and the second term represents the coupling between the lepton and the Higgs boson.

The masses of quarks are generated similarly with the left-handed quark doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L. \tag{2.26}$$

However, as the non-zero vacuum expectation value occurs in the lower component of the Higgs doublet, only the masses for the lower component of the quark doublets, known as down-type quarks, can be generated using (2.21). To acquire the masses for the upper component of the quark doublets, up-type quarks, a conjugate doublet in the unitary gauge is derived:

$$\phi_c = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} = -\frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix}. \tag{2.27}$$

The Lagrangian for up-type quarks is expressed as

$$\mathcal{L}_{fermions,up} = g_f(\bar{L}\phi_c R + \bar{R}\phi_c^\dagger L), \tag{2.28}$$

and is expanded for the ud doublet as

$$\mathcal{L}_{fermions,up} = -g_u \left[\begin{pmatrix} \bar{u} & \bar{d} \end{pmatrix}_L \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} u_R + \bar{u_R} \begin{pmatrix} -\phi^0 & \phi^{-*} \end{pmatrix} \begin{pmatrix} u \\ d \end{pmatrix}_L \right]. \tag{2.29}$$

After symmetry breaking, the Lagrangian becomes

$$\mathcal{L}_{fermions,up} = -\frac{g_u}{\sqrt{2}} v(\bar{u_L} u_R + \bar{u_R} u_L) - \frac{g_u}{\sqrt{2}} h(\bar{u_L} u_R + \bar{u_R} u_L), \tag{2.30}$$

with the Yukawa coupling $g_u = \sqrt{2} m_u / v$, where m_u is the mass of the up quark. Therefore, the Yukawa coupling of fermions to the Higgs field is generally given by

$$g_f = \sqrt{2} \frac{m_f}{v}, \tag{2.31}$$

indicating that the coupling strength of fermions to the Higgs boson is proportional to the mass of the fermion m_f.

2.2 pp collisions at the Large Hadron Collider

Proton-proton (pp) collisions at the LHC [26, 27] occur at the TeV energy scale to probe the SM by inducing hard scattering processes, which allows the partons (quarks and gluons) within the protons to interact and produce high mass particles, such as the Higgs boson. Due to asymptotic freedom, the coupling constant of the strong interaction, α_s, decreasing at high energy scales, perturbation theory can be used to calculate the cross section for these processes. However, perturbative corrections could result in integrals that are infinite. These integrals are treated using renormalisation, which introduces the renormalisation scale μ_R.

The cross section of a hard scattering process initiated by two protons with momenta P_1 and P_2 to give final state X, i.e. $pp \to X$, is

$$\sigma(pp \to X) = \sum_{i,j} \int dx_1 dx_2 \, f_{i,p}(x_1, \mu_F^2) f_{j,p}(x_2, \mu_F^2) \, \hat{\sigma}_{ij \to X}(p_1, p_2, \mu_R^2, \mu_F^2), \tag{2.32}$$

where $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$ are the momenta of the partons, expressed in terms

of the momentum fractions of the protons, x_1 and x_2, respectively. The functions $f_{i,p}$ and $f_{j,p}$ are the parton distribution functions (PDFs) that correspond to the probability of partons i and j with x_1 and x_2 to participate in the hard scattering process. The PDFs are fixed at the factorisation scale μ_F, which is the point of separation between perturbative and non-perturbative scales, i.e. hard and soft (low momentum transfer) processes. The partonic cross section of the process from the initial state of partons i and j to the final state X is denoted by $\hat{\sigma}_{ij\to X}$.

2.2.1 Higgs boson production and decay

The SM Higgs boson is produced from the different hard scattering processes of quarks (q) or gluons (g) in the pp collision. Fig. 2.2 illustrates the four main production modes at the LHC [28], which are summarised as follows:

- Gluon fusion (ggF): The ggF process proceeds through a loop diagram involving a virtual top quark as the Higgs boson does not couple with massless particles, such as gluons. The production channel has the largest cross section but in some decay channels is subject to significant background from multijet production, which makes it challenging to access.

- Vector boson fusion (VBF): The VBF production mode has the second largest cross-section and is associated with the scattering of two quarks that is mediated by the exchange of a W or Z boson, which emit the Higgs boson. Features of the VBF processes, including the strong suppression of QCD radiation in the centre of the detector, are exploited to distinguish them from the large multijet backgrounds.

- Associated production with a vector boson (VH): The VH production channel, where $V = W, Z$, is also known as Higgsstrahlung. The decays of the W or Z bosons are used to produce a more distinct signature for triggering and QCD rejection. This production channel is analysed in Part III.

- Associated production with a pair of top quarks ($t\bar{t}H$): The $t\bar{t}H$ production mechanism involves a Higgs boson being produced from the fusion of a pair of top

quarks $t\bar{t}$, and can be used to directly access the coupling of the Higgs boson to the top quark. However, the cross section of the $t\bar{t}H$ production mode is smaller than the other three production mechanisms, meaning that $t\bar{t}H$ processes are rarer and challenging to measure.

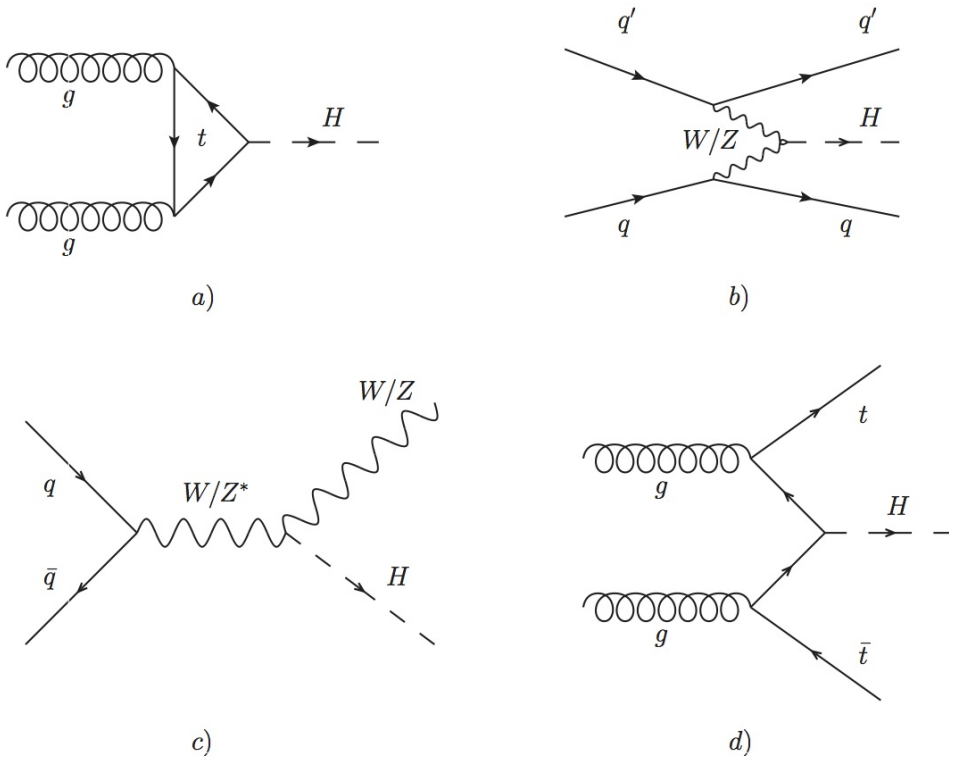


Figure 2.2: Feynman diagrams of main production modes of the Higgs boson at the LHC [29]: a) gluon fusion, b) vector boson fusion, c) associated production with a vector boson, d) associated production with a pair of top quarks.

The Higgs boson couples to all massive SM particles, as illustrated in Fig. 2.3, but as the coupling strength of the Higgs boson to another particle increases with the mass of the particle, the Higgs boson is likely to decay into more massive particles. The probability of the Higgs boson decaying into a certain final state X can be specified by the branching ratio (BR):

$$BR(H \to X) = \frac{\Gamma(H \to X)}{\Gamma_H}, \tag{2.33}$$

where $\Gamma(H \to X)$ is the width of the $H \to X$ decay mode, known as the partial decay width, and Γ_H is the total width of all possible decay modes of the Higgs boson. The predicted BRs for a SM Higgs boson with $m_H = 125$ GeV are displayed in Table 2.1,

where the largest BR is that of $H \to b\bar{b}$, which is the focus of the work presented in Part III.

Decay mode	Branching ratio
$H \to b\bar{b}$	57.8%
$H \to WW^*$	21.6%
$H \to gg$	8.6%
$H \to \tau^+\tau^-$	6.4%
$H \to c\bar{c}$	2.9%
$H \to ZZ^*$	2.7%
$H \to \gamma\gamma$	0.2%

Table 2.1: Predicted branching ratios for the a Higgs boson mass of 125 GeV.

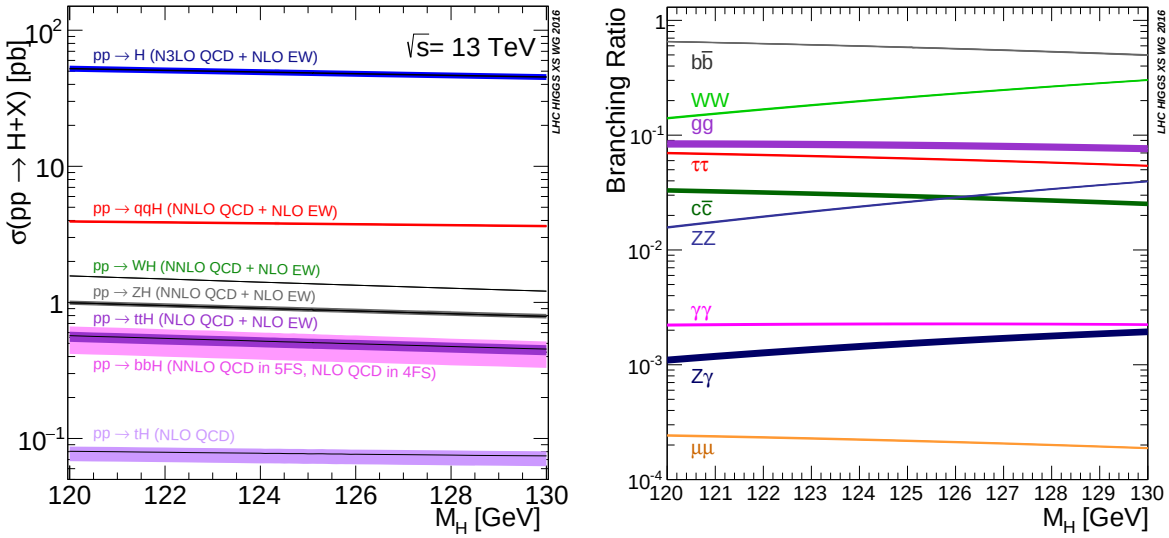


Figure 2.3: Expected production cross sections (left) and decay branching ratios (right) of the Standard Model Higgs boson at $\sqrt{s} = 13$ TeV as a function of the Higgs boson mass [30].

2.2.2 $VH, H \to b\bar{b}$ channel

The search for the $H \to b\bar{b}$ decay has been conducted in the VBF [31], VH [32], and $t\bar{t}H$ [33] production modes at the LHC. Although the VH and $t\bar{t}H$ production modes have much lower cross sections, their leptonic decay channels are exploited to provide clean signatures for triggering and to aid the rejection of the overwhelming QCD background. The cross sections of the leptonic channels in the VH mode are higher than those in the $t\bar{t}H$ mode, hence the VH mode is preferred in the search for $H \to b\bar{b}$.

The leptonic decay of the vector bosons results in three different signatures that can be favourably targeted at the LHC, as demonstrated in Fig. 2.4.

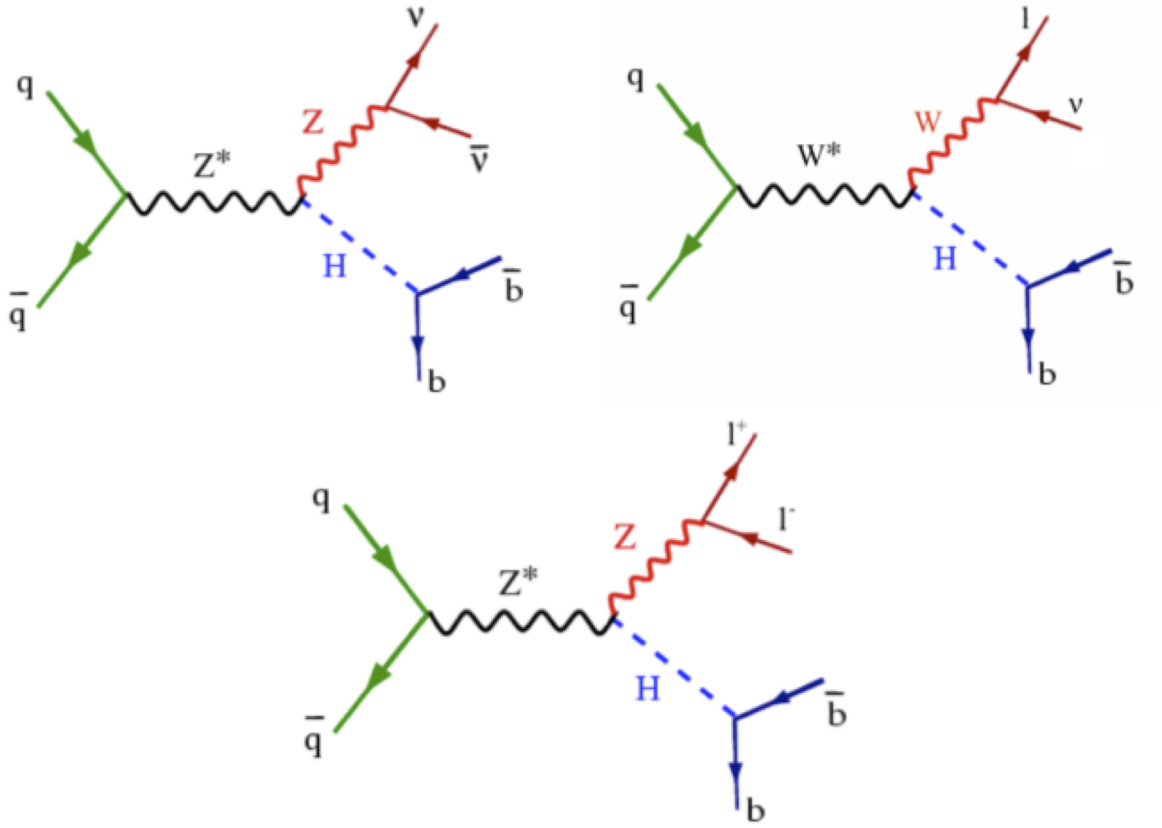


Figure 2.4: Feynman diagrams of the leptonic decay channels of the Higgs boson produced in association with a W or Z boson [34].

2.2.3 Monte Carlo simulations

The evolution of pp collisions, which is presented in Fig. 2.5, is simulated by Monte Carlo (MC) generators [35, 36] to model the processes generated in the collisions.

The processes in a pp collision start with incoming protons that are characterised by PDFs. A collision between the partons within the protons result in the hard scattering process at parton level that is described by matrix element (ME) calculations. Incoming and outgoing partons of the hard scattering process can emit radiation, known as initial state radiation (ISR) and final state radiation (FSR), respectively. These radiations can be in the form of gluons, which can radiate further emissions, generating parton showers of the incoming and outgoing partons. As the collision evolves, the distance between the partons increases until the strength of the strong interaction becomes large

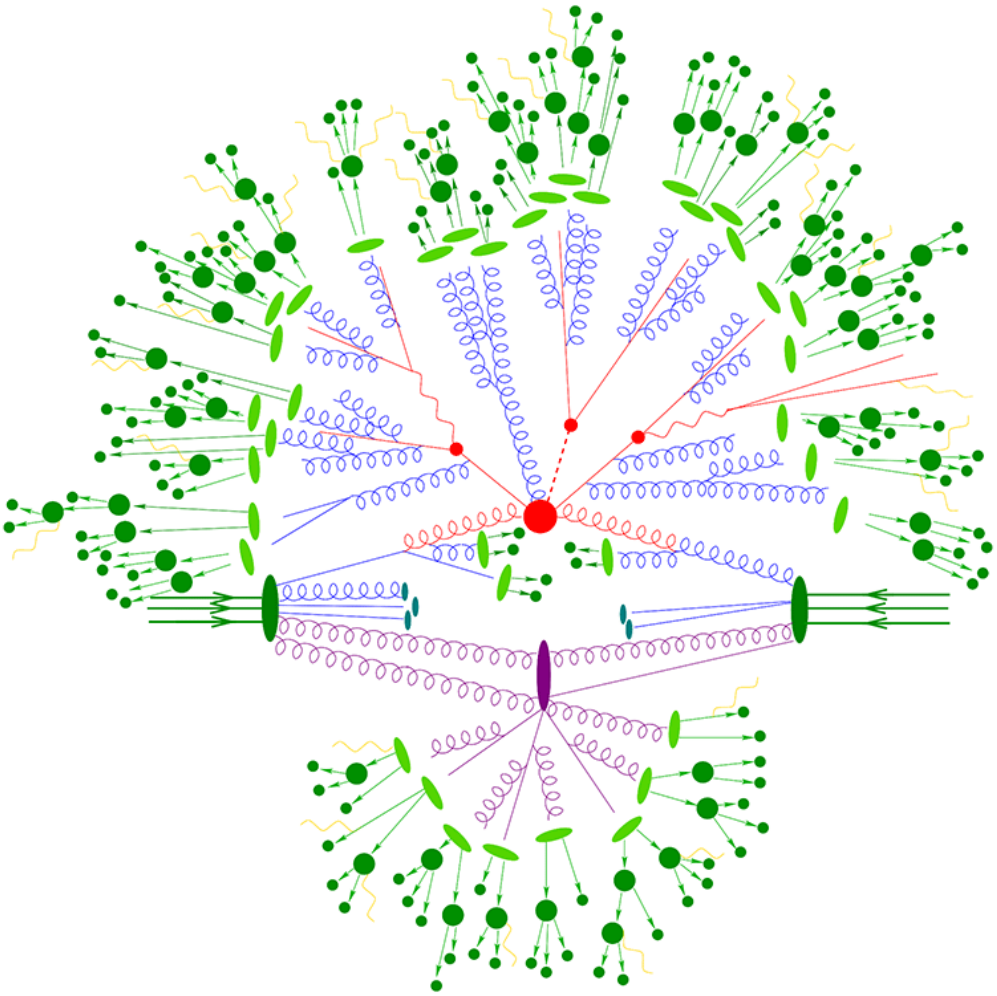


Figure 2.5: Processes in a *pp* collision [37]. Hard scattering is denoted by red, parton shower by blue, hadronisation by light green, underlying event by purple, and hadrons by dark green.

enough to confine the partons into colourless hadrons in the hadronisation process. So far, the processes in the evolution have been discussed using one parton from each of the two incoming protons, but as protons are composed of many partons, interactions between other parton pairs can occur. Partons that do not participate in the hard scattering process are considered as remnants of the scattering, which are represented by the underlying event (UE).

Chapter 3

The Large Hadron Collider and the ATLAS detector

The Large Hadron Collider (LHC) [3] at the European Organisation for Nuclear Research (CERN) in Switzerland is the world's largest particle accelerator. The LHC was designed to accelerate protons in opposite directions and collide them at a specific centre-of-mass energy that depends on the running period of the accelerator. The LHC operated at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV for the 2010-2012 data-taking period, known as Run-1. During the shutdown from 2013 to 2015, the LHC was upgraded to reach a higher centre-of-mass energy. The LHC operated at $\sqrt{s} = 13$ TeV for the 2015-2018 running period, known as Run-2, and is currently being upgraded to achieve up to $\sqrt{s} = 13.6$ TeV for Run-3, which is expected to start in 2022.

3.1 The Large Hadron Collider

The LHC is part of the CERN accelerator complex, which is illustrated in Fig. 3.1, and consists of a 27 km ring of superconducting magnets, and is located approximately 100 m underground.

Prior to entering the LHC ring, the protons are produced from the ionisation of hydrogen gas by an electric field. To ensure that the protons achieve their final energy of 6.5 TeV, the energy of the protons is gradually increased in a series of accelerators: Linear Accelerator to 50 MeV, Proton Synchrotron Booster (PSB) to 1.4 GeV, Proton Synchrotron (PS) to 25 GeV, and SPS (Super Proton Synchrotron) to 450 GeV.

The protons are then split and enter the two beam pipes of the LHC ring, where they circulate in opposite directions and are accelerated to their final energy. The protons collide at four interaction points around the LHC ring corresponding to the ALICE [39],

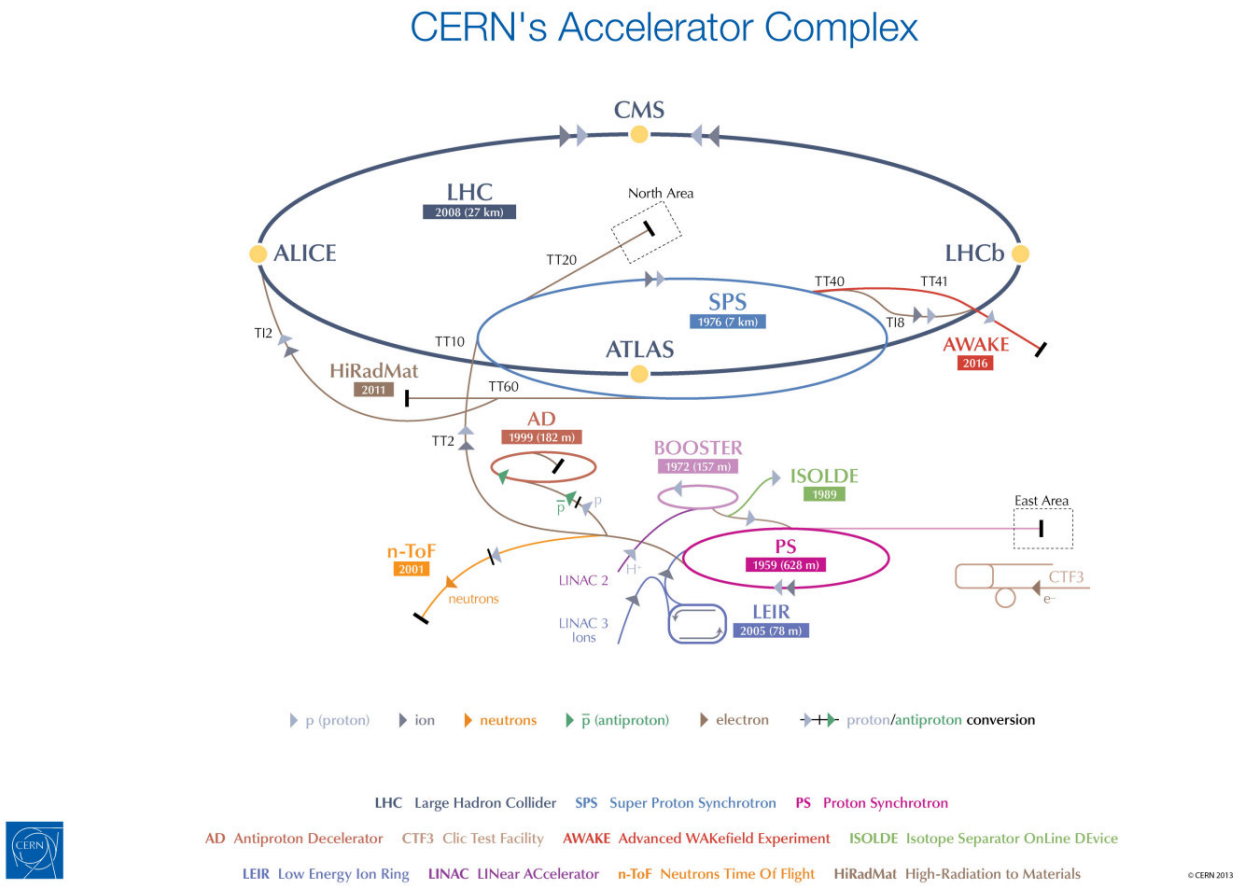


Figure 3.1: A diagram of the CERN accelerator complex [38].

ATLAS [40], CMS [41] and LHCb [42] detectors.

3.1.1 Luminosity and pile-up

Protons are formed into bunches with each bunch containing approximately 1.1×10^{11} protons. The separation of the bunches result in beams crossing at 25 ns intervals.

These beam properties can be used to define the instantaneous luminosity given by

$$\mathcal{L} = \frac{f n_b N_1 N_2}{2\pi\sigma_x\sigma_y}, \tag{3.1}$$

where f is the beam revolution frequency, n_b is the number of bunches in one beam, N_1 and N_2 are the respective numbers of protons in bunch 1 and 2, and σ_x and σ_y are the mean beam width in the x and y direction, respectively. The integrated luminosity over a certain time period, L, is used to define how much data has been delivered.

An event corresponds to the two bunches crossing within a detector. However, in the same or adjacent bunch crossing, a significant number of collisions can occur, which is referred to as pile-up. Pile-up can be categorised as in-time and out-of-time. In-time pile-up is additional pp collisions occurring in the same bunch crossing as the collision of interest, while out-of-time pile-up is additional pp collisions occurring in bunch crossings just before and after the collisions of interest. The amount of pile-up is quantified by the mean number of interactions per bunch crossing,

$$\mu = \frac{\mathcal{L}\sigma}{f n_b}, \tag{3.2}$$

where $\mathcal{L}$ is the instantaneous luminosity, σ is the inelastic cross section, f is the beam revolution frequency, and n_b is the number of bunches. Fig. 3.2 shows the mean number of interactions per bunch crossing and the integrated luminosity collected by the ATLAS detector during Run-2 of the LHC.

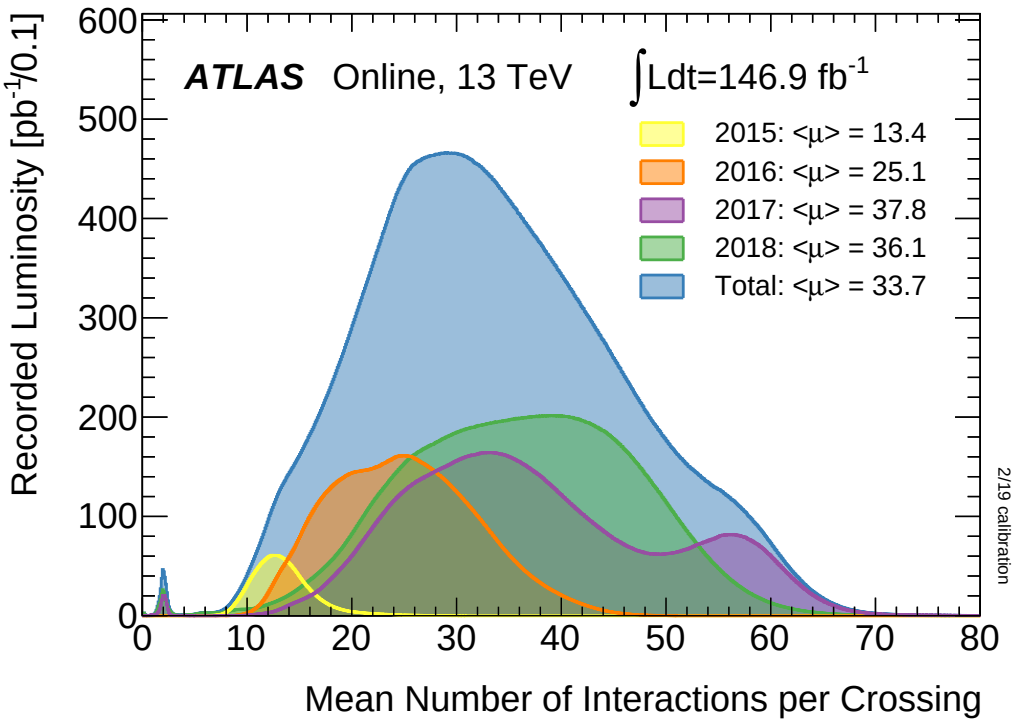


Figure 3.2: Mean number of interactions per bunch crossing collected by the ATLAS detector during Run-2 of the LHC [43]

.

3.2 The ATLAS Detector

The ATLAS (A Torodial LHC ApparatuS) [40] detector, as illustrated in Fig. 3.3, is formed of a cylindrical barrel around the beam pipe with the proton-proton (pp) collisions occurring at the centre of the detector. The detector is composed of three main components. The innermost component is the Inner Detector, which is surrounded by the electromagnetic and hadronic calorimeters, which are surrounded by the muon spectrometer.

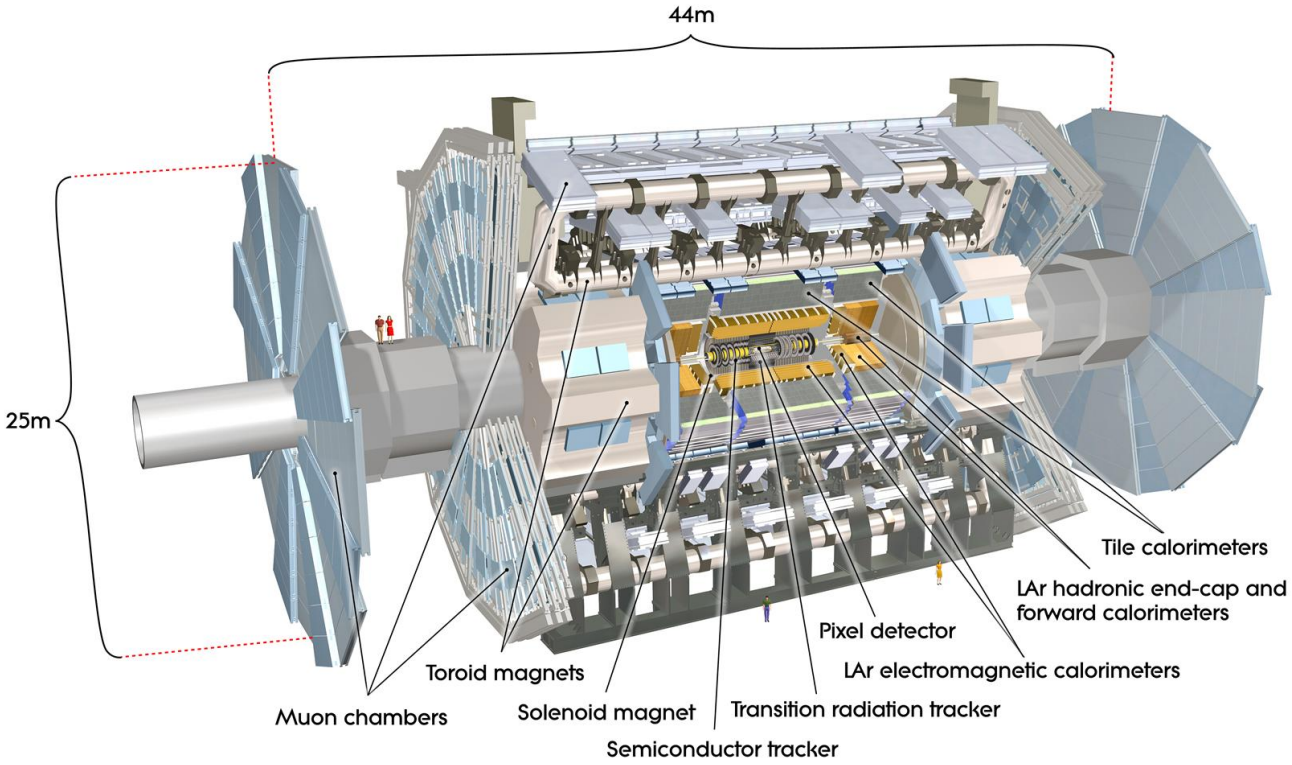


Figure 3.3: Cut-away view of the ATLAS detector [44].

A right-handed coordinate system is employed by the detector with the interaction point, which is at the centre of the detector, as the origin. The x-axis increases towards the centre of the LHC ring, the positive y-axis points upwards, and the z-axis is along the beam pipe. The transverse $(x - y)$ plane uses spherical coordinates to describe the angle around the beam axis, ϕ, and angle from the beam axis to the y-axis, θ. A more

commonly used quantity involving θ is the pseudorapidity:

$$\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3.3}$$

Distances between objects are defined in the $\eta - \phi$ plane and are given by the angular separation:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.4}$$

3.2.1 Inner Detector

The Inner Detector [45] (ID) is submerged in a magnetic field of 2T to measure the trajectories (tracks), charge, and momenta of charged particles within $|\eta| < 2.5$. The expected tracking momentum resolution of the ID as a function of the transverse momentum (p_T in GeV) of the track is

$$\frac{\sigma_{p_T}}{p_T} = 0.05\%\ p_T \oplus 1\%. \tag{3.5}$$

Fig. 3.4 illustrates the ID, which is composed of three sub-detector layers at different distances to the beam axis.

- **Pixel detector:** The pixel detector is positioned closest to interaction point to provide high precision measurements close to the primary vertex (PV), which is the point of a hard scattering event. Thus, this sub-detector is vital to reconstruction of vertices and track impact parameters (IPs), which are defined as the signed distance of closest approach of the track to the interaction point in the transverse (d_0) and longitudinal (z_0) plane. The sign of the IP is positive if the track crosses the jet axis in front of the PV, otherwise the sign is negative. The transverse and longitudinal IP significances are defined with respect to their uncertainties (σ) as $|d_0|/\sigma_d$ and $|z_0|/\sigma_z$, respectively. The pixel detector is formed of four layers of silicon pixels in the barrel and three disks of sensors in the end-caps. These layers typically provides four hits per track in the barrel region. The pixel sizes in these layers are $50 \times 400\ \mu m^2$, offering a resolution of 10 μm and 115 μm in the $r - \phi$ and z direction, respectively.

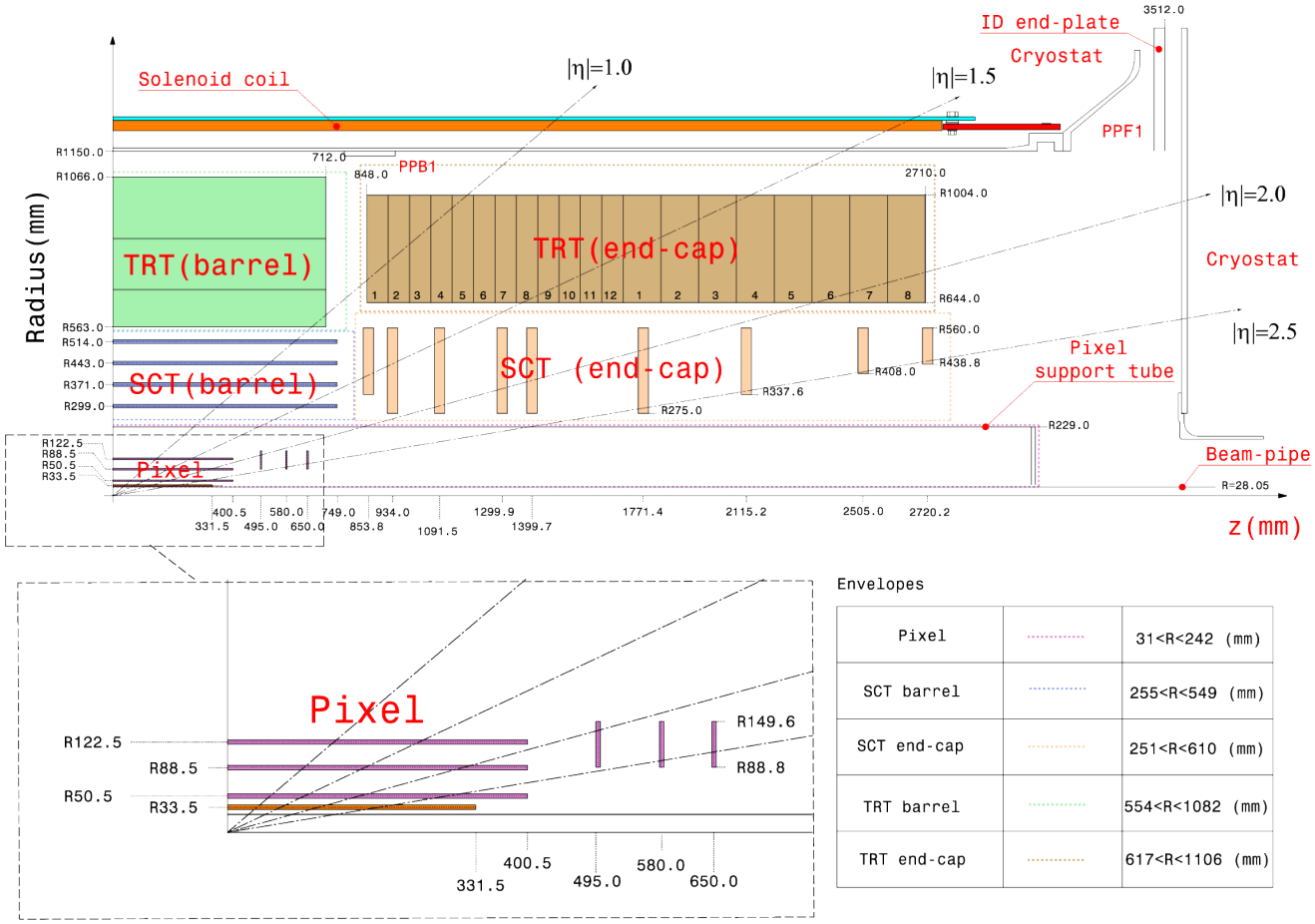


Figure 3.4: A $r - z$ cross-section view of the whole layout (top) of the Inner Detector [45], with a magnified view of the pixel detector (bottom).

During the shutdown before Run-2, the innermost layer of silicon pixels, known as the Insertable B-Layer (IBL) [46], was added between the beam pipe and the first layer to increase tracking robustness against data loss in the first layer, preserve tracking performance in the case of higher pile-up as a consequence of higher luminosity, and improve the quality of impact parameter reconstruction for tracks. The pixel sizes in the IBL are slightly smaller at $50 \times 250\ \mu m^2$, giving a higher resolution of 8 μm in the $r - \phi$ direction and 40 μm in the z direction.

- **Semiconductor Tracker (SCT):** The SCT is formed of four double-sided layers of silicon microstrips in the barrel region and two sets of nine disk layers in the end-caps, typically providing eight hits per track in the barrel region. The overall intrinsic resolution of the SCT is 17 μm in the $r - \phi$ direction, and 580 μm in the r direction for the disks and z direction for the barrel.

- **Transition Radiation Tracker (TRT):** The TRT covers $|\eta| < 2.0$ and is composed of straw tubes filled with xenon and argon gas and a tungsten wire in the centre of the tubes. Charged particles traversing the TRT ionise the gas mixture and this produces electrons, which are collected by the wire. This tracker provides up to 33 hits in the barrel and 38 in the end-caps. The TRT has an intrinsic resolution of 130 μm per straw tube, providing information only in the $r - \phi$ direction.

  The space between the straw tubes are filled with polymer fibres and foils to provide different refractive indices, therefore particles passing through can radiate photons that are absorbed by the xenon gas to give a larger signal. As lighter particles, such as electrons, are more likely to radiate photons than heavier hadrons and pions, the transition radiation helps to distinguish electrons from other particles. This also provides electron ID information based on the number of hits above a higher energy deposit threshold corresponding to transition radiation.

3.2.2 Calorimeters

Calorimeters [40] absorb particles to measure the energy deposition and direction. The calorimeters are composed of alternating layers of an absorber material and an active material. Signals are created in the layers of the active material. Through interactions with the absorber material, particles lose energy and create secondary particles.

The average length that a particle can travel through the material before losing a characteristic amount of energy depends on the material type and energy of the particle, but is generally much larger for hadrons than for electrons and photons. Hence, the absorber material and thickness of the calorimeters are chosen such that electrons and photons deposit all their energy in the electromagnetic calorimeters (ECAL) and hadrons in the hadronic calorimeters (HCAL). Fig. 3.5 shows the inside of the calorimeter system.

- **Electromagnetic calorimeters:** Electrons interact primarily via Bremsstrahlung and photons interact via pair production, producing a shower of electrons and photons. Liquid argon is used as the active material, which is ionised to measure the energy of the shower. Lead is used for the absorber material. The active and absorber material are layered in accordion geometry to provide full ϕ coverage.

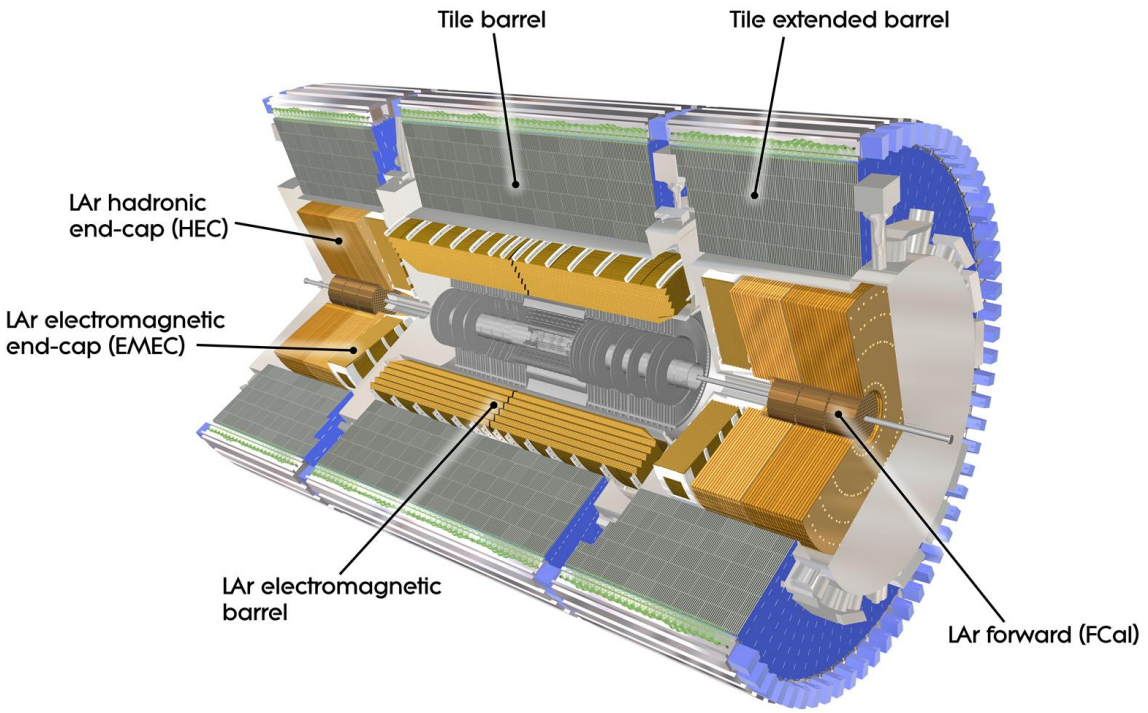


Figure 3.5: A cut-away view of the calorimeter system [47].

The evolution of an electromagnetic (EM) shower is characterised by the material's radiation length, X_0, which is the distance an EM interacting particle travels before losing all but $1/e$ of its energy, To obtain a large fraction of EM showers, ECAL ($|\eta| < 3.2$) is at least 22 X_0 thick and at least 24 X_0 thick in the barrel ($|\eta| < 1.475$) and end-caps ($1.375 < |\eta| < 3.2$) regions, respectively. The design energy resolution of the ECAL as a function of energy (E in GeV) and its uncertainty σ_E is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\%. \tag{3.6}$$

- **Hadronic calorimeters:** Hadrons can interact with the absorber material by the strong force as well as the EM force, resulting in a mixture of electrons, photons and hadrons in the showers. The barrel has polystyrene scintillating tiles as the active medium and uses steel for the absorber material. The hadronic end-cap calorimeter (HEC) and forward calorimeter (FCal) cover $1.5 < |\eta| < 3.2$ and $3.1 < |\eta| < 4.9$, respectively. The HEC has a copper absorber and the FCal has three layers of absorber composed of copper and tungsten, with both calorimeters

adopting LAr as the active material.

The average distance travelled by a hadron before undergoing an inelastic nuclear interaction is defined by the interaction length of the material, λ. The total thickness of HCAL is at least 9.7 λ thick in the barrel and at least 10 λ thick in the end-caps. The expected energy resolution of the HEC as a function of energy (E in GeV) is

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\%, \tag{3.7}$$

and of the FCal is

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\%. \tag{3.8}$$

3.2.3 Muon Spectrometer

Muons are able to pass through the calorimeters with minimal interactions, which allows them to be detected and their properties to be measured in the Muon Spectrometer (MS) [40], which covers $|\eta| < 2.7$. The MS, as illustrated in Fig. 3.6 is immersed in a strong magnetic field to measure muon tracks, charge, and momentum. In addition, parts of the MS are used as a hardware-based trigger, where the Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) are installed in the barrel and end-cap regions, respectively.

The MS consists of Monitored Drift Tubes (MDTs) in the barrel and Cathode Strip Chambers (CSCs) in the innermost end-cap layer for precision tracking. MDTs work similarly to the straw tubes in the TRT of the ID, except the tubes are filled with argon gas, and the time taken for the electrons to be collected by the wire is used to determine the position of the muon. CSCs are composed of argon gas and numerous tungsten wires between two copper strips, and work similarly to the MDTs, except positive ions are produced from the ionisation and are collected by the copper strips. The CSCs are used instead of MDTs due to their higher rate capability. The expected tracking momentum resolution of the MS as a function of the track p_T, at $p_T = 1$ TeV, is

$$\frac{\sigma_{p_T}}{p_T} = 10\%. \tag{3.9}$$

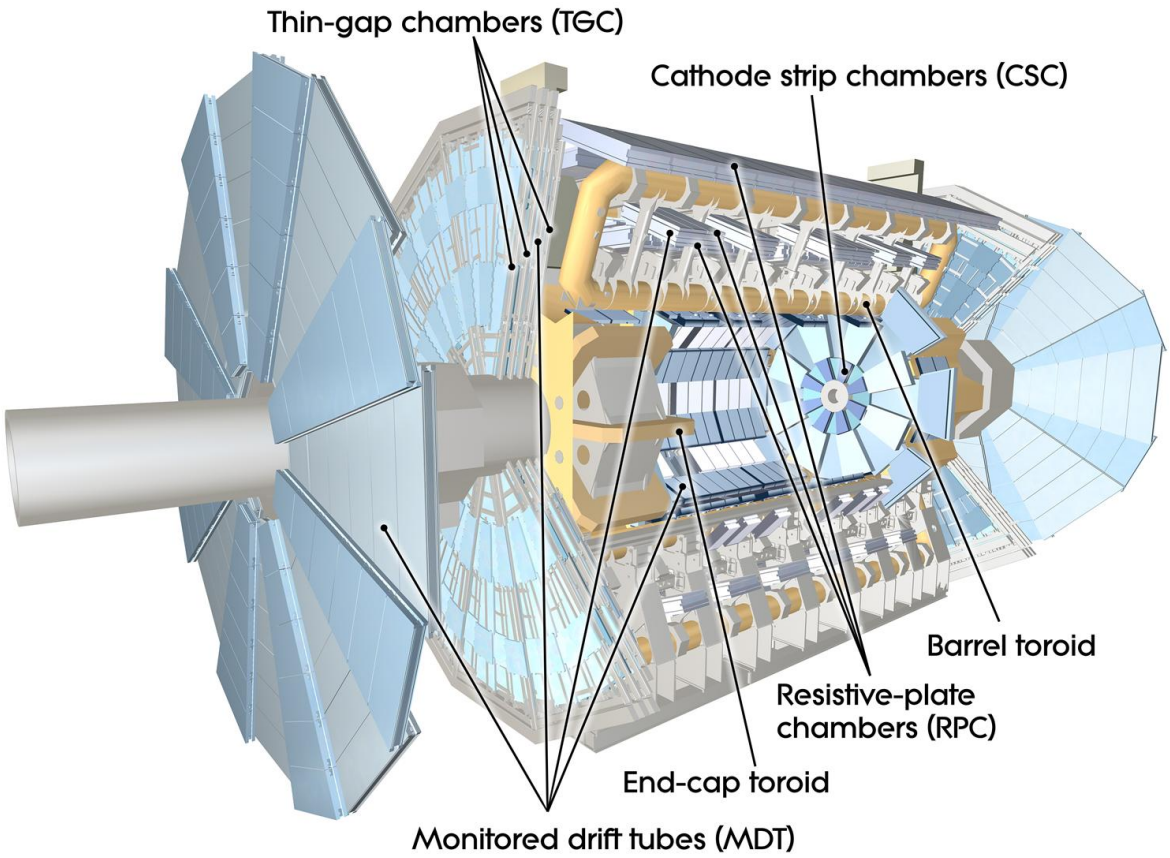


Figure 3.6: A cut-away view of the muon spectrometer [48].

3.2.4 Trigger and Data Acquisition System

At $\sqrt{s} = 13$ TeV, collisions happen at a rate of 40 MHz, which is too high to record every event. Therefore, a trigger system [49] is used to select events of interest, whilst maintaining an acceptable trigger rate to reduce the recording rate from 40 MHz to 1 kHz. The trigger system employs a Level 1 (L1) hardware trigger and a Higher Level Trigger (HLT) software trigger, with a trigger menu to define the selection criteria for the L1 and HLT triggers.

The L1 trigger only uses information from the calorimeters and the muon spectrometer to identify objects with high transverse momentum or events with a large energy imbalance in the total transverse momentum. The HLT trigger uses a farm of computers to run offline-like algorithms to reconstruct events. The L1 trigger reduces the event rate from 40 MHz to 100 kHz, while the HLT trigger reduces the event rate from 100 kHz to 1 kHz.

Chapter 4

Object reconstruction

During event reconstruction, physics objects are identified and reconstructed from the signals recorded by the detector. Following the reconstruction of objects, it is possible that the same signature in the detector can be used for reconstructing different objects, therefore overlap removal [50] is undertaken between objects, similar to the procedure discussed in Ref. [50].

4.1 Tracks and Primary Vertices

Tracks are reconstructed from the energy deposits produced by charged particles traversing the ID [51]. The energy deposits are clustered independently in the pixel detector and SCT layers forming space-points in each layer, which are measurements corresponding to hits in the sub-detectors.

Track reconstruction starts with the formation of a track seed using three space-points from the pixel and/or SCT layers. A combinatorial Kalman filter is used to create track candidates by adding space-points that are consistent with the initial trajectory from the seed. However, ambiguities arise from the possibility of the additional space-points being associated to multiple track candidates. To solve this ambiguity, a base score is assigned to each track based on the p_T of the track candidate and χ^2 of its fit. If clusters are assigned to a track, then its score is increased by a weight to reflect the expected cluster multiplicities in the sub-detectors and their resolution. The number of expected hits which are missing, known as holes, reduces the track score as it represents poorly reconstructed tracks. The scores are used to rank the tracks, then the descending ordered list is fed into the ambiguity solver, which imposes the following requirements on the tracks: $p_T > 400$ MeV, $|\eta| < 2.5$, minimum of 7 hits, maximum of one hole in the pixel detector, maximum of two holes in the combined pixel and SCT detectors.

Another requirement is that clusters can be shared by a maximum of two tracks, with preference given to higher scored tracks, and a track can have a maximum of two share clusters. Tracks that have too many shared clusters are used to create stripped-down track candidates, which then re-enter the track reconstruction procedure as new inputs to the ambiguity solver. Accepted track candidates are combined with measurements from the TRT to give the reconstructed track. The tracking reconstruction efficiency is dependent on a number of factors including the type of particle, pseudorapidity, and p_T of the particle. For example, at $p_T = 5$ GeV and $|\eta| < 2.5$, the track reconstruction efficiency ranges from 97-99% for muons, 80-95% for pions, and 70-95% for electrons [52].

Alternatively, in MC, tracks can be reconstructed using truth information to identify the hits a particle created in the detector, and using them to directly fit the particle trajectories without the track finding and ambiguity solving stages. These tracks are referred to as pseudo tracks, and they represent the best tracking possible within the ATLAS detector given the minimum hit requirements. The quality of matching to a given truth particle is quantified by the truth match probability [53]:

$$TMP = \frac{10 \times N^{Pixel}_{common} + 5 \times N^{SCT}_{common} + 1 \times N^{TRT}_{common}}{10 \times N^{Pixel}_{reco} + 5 \times N^{SCT}_{reco} + 1 \times N^{TRT}_{reco}}, \tag{4.1}$$

where, for a given sub-detector layer i, N^i_{common} is the number of hits created by the truth particle in the detector that are common to the reconstructed track and N^i_{reco} is the number of hits on the reconstructed track.

A set of the reconstructed tracks combined with information about the position of the beam spot in the detector are used to reconstruct PVs [54]. The x- and y-coordinates of the vertex seeds are determined from the centre of the beam spot, while the z-coordinate of the vertex seeds is calculated as the mode of the z-coordinates of the points of closest approach of the tracks in the vicinity of the seed. Then, an iterative χ^2 minimisation is conducted to find the optimal vertex position. In each iteration, the tracks are weighted based on their compatibility with the vertex seed and the vertex position is recomputed using the weighted tracks. Tracks that are incompatible with the vertex at more than seven standard deviations are removed from the vertex candidate. The iterative procedure is repeated until all reconstructed tracks are associated to a vertex. The primary vertex is given by the reconstructed vertex with highest p_T^2 sum of

all associated track. Other reconstructed vertices are considered as pile-up vertices. To ensure tracks selected are consistent with the PV, further requirements are enforced on the tracks: $d_0 < 2$mm and $|z_0 \sin\theta| < 3$ mm, where θ is the polar angle of the track. The primary vertex reconstruction efficiency is expected to exceed 99%.

4.2 Leptons

As neutrinos are not directly detected and τ-leptons are not used for the work presented in Parts II and III, leptons referred to in this thesis consists of only electrons and muons and are represented by the ℓ symbol.

4.2.1 Electrons

Signals in the ID and ECAL are used to reconstruct electrons [55]. Seed clusters of size equivalent to 3×5 cells in $\eta \times \phi$ in the ECAL are identified by requiring energy of the seed cluster to be greater than 2.5 GeV. Reconstructed tracks are matched to the ECAL clusters. If no match is found, then the ECAL cluster is considered a photon candidate. If several tracks are matched to the EM cluster, tracks are ranked based on the angular separation between the track and cluster, the number of pixel hits, and the number of holes. The highest ranked track corresponds to the electron track. The electron track also has to be consistent with the PV by requiring $d_0/\sigma_d < 5$ (Section 3.2.1) and $\Delta z_0 \sin\theta < 0.5$ mm, where Δz_0 is the longitudinal IP corrected for the reconstructed PV.

An identification procedure is conducted by combining the electron track and cluster information, such as the number of hits in the ID layers and the ratio of the cluster energy to the p_T of the track, in a multivariate analysis to reduce the background from fake prompt electrons, which can originate from the conversion of a photon to an electron-positron pair, semi-leptonic decays of heavy-flavour hadrons, and jets misidentified as electrons. The discriminant is used to define different quality working points based on the electron efficiency and the hadron jet rejection. The different working points are Loose, Medium, and Tight, corresponding to approximately 95%, 92%, and 87% electron efficiencies with hadron jet rejections of 99.7%, 99.8%, and 99.9%, respectively,

in the [60, 80] p_T range.

To further suppress the fake prompt electrons background, the isolation working points are defined with respect to the p_T of the electron candidate using the following two variables:

- $p_T^{varcone0.2}$ corresponds to the sum of the p_T of all tracks within a varying cone of $\Delta R = \min(0.2, 10 \text{ GeV}/p_T)$ around the electron track candidate,
- $E_T^{topocone0.2}$ refers to the sum of the p_T of the EM clusters within a cone of $\Delta R = 0.2$ around the cluster corresponding to the electron candidate.

The isolation working points relevant for work presented in this thesis are:

- LooseTrack uses $p_T^{varcone0.2}$ to target a fixed 99% efficiency across the entire electron p_T range,
- HighPtCalo obtains a p_T dependent efficiency of around 95% by requiring $E_T^{topocone0.2} < \max(0.015 p_T, 3.5 \text{ GeV})$.

Electron efficiencies in simulation are compared to those in data for $Z \to e^+e^-$ and $J\psi \to e^+e^-$ samples to derive scale factors for the reconstruction, isolation and identification procedures described above. The calibration accounts for the imperfect detector simulation and physics modelling of the electron.

4.2.2 Muons

Muon tracks are reconstructed independently from the signals in the ID and the MS. In the ID, the muon tracks are reconstructed using the same procedure described in Section 4.1, while in the MS, the tracks are reconstructed using a χ^2 fit on the hits in the muon chambers [56]. Additionally, muon tracks have to be consistent with the PV by the requirement of $d_0/\sigma_d < 3$ and $\Delta z_0 \sin\theta < 0.5$ mm. Four types of reconstructed muons are formed from the tracks, along with information from the calorimeter:

- Combined muons are formed from a fit combining the muon tracks that were reconstructed separately in the ID and MS.
- Segment-tagged muons have a track in the ID extrapolated to the MS that matches at least one track segment in the MDT or CSC chambers. These muons arise when muons cross only one layer of the MS chambers, which is either due to low p_T or as they fall in regions with reduced MS acceptance.

- Calorimeter-tagged muons have a muon track in the ID that is matched to an energy deposit in the calorimeter. These muons are used to recover acceptance in the MS regions that are only use for cabling to the ID and calorimeters.

- Extrapolated muons have a muon track reconstructed from the MS only, but whose estimated energy loss and IPs are consistent with a track originating from the interaction point.

An identification procedure is performed to distinguish prompt muons from fake prompt muons, which can arise from the semi-leptonic decay of a heavy-flavour hadron and the in-flight decays of a hadron. Cuts on several discriminating variables, such as the χ^2 of the combined fit and the charge and momentum of the muon, are used to define quality working points: Loose, Medium, and Tight with approximate efficiencies of 98%, 96%, and 92% with fake muon rejections of 99.2%, 99.8%, and 99.9%, respectively, in the 20 GeV $< p_T <$ 100 GeV range.

Similar to electrons, isolation working points are defined with respect to the p_T of the muon candidate to discriminate muons from the fake prompt muons. For muons, the following isolation variables are defined:

- $p_T^{varcone0.3}$ corresponds to the sum of the p_T of all tracks within a varying cone of $\Delta R = \min(0.3, 10 \text{ GeV}/p_T)$ around the muon track candidate,

- $E_T^{topocone0.2}$ refers to the sum of the p_T of the calorimeter clusters within a cone of $\Delta R = 0.2$ around the cluster corresponding to the muon candidate,

- $p_T^{cone0.2}$ corresponds to the sum of the p_T of tracks within a cone of $\Delta R = 0.2$ around the muon track candidate.

The isolation working points [57] relevant to the thesis are:

- LooseTrack uses $p_T^{varcone0.3}$ to ensure a fixed 99% efficiency across the whole muon p_T range,

- HighPtTrack imposes a requirement of $p_T^{cone0.2} < 1.25$ GeV to provide around 95% efficiency, depending on the p_T of the muon.

Scale factors are obtained from the evaluation of muon efficiencies in simulation and data using $Z \to e^+e^-$ and $J\psi \to e^+e^-$ samples to correct mismodelling in simulation.

4.3 Jets

Jets are collimated streams of hadrons produced by the hadronisation of quarks and gluons, which cannot exist as free particles due to colour confinement.

The reconstruction of jets starts with the formation of clusters with calorimeter cell signals that are topologically connected [58]. The energy significance, S, of a cell is defined as the ratio of the energy deposit to the background noise of the cell. If a cell registers $S > 4$, it is clustered together with its neighbouring cells. If a neighbouring cell has $S > 2$, then its neighbours are also added to the cluster. The merging of the cells are applied to further neighbours until there are no more neighbouring cells with $S > 0$.

Jets are reconstructed from the clusters with the anti-k_T algorithm [59], which uses the distance measures:

$$d_{ij} = \min\left(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2}\right)\frac{\Delta R(i,j)^2}{R^2}, \tag{4.2}$$

$$d_{iB} = \frac{1}{k_{T,i}^2}, \tag{4.3}$$

where $\Delta R(i,j)$ is the angular distance between two reconstructed clusters i and j, R is the radius parameter defining the size of the resulting jet, and $k_{T,i}$ and $k_{T,j}$ are the transverse momentum of clusters i and j, respectively. The algorithm compares the distance between the two clusters (d_{ij}) to the distance between cluster i and the beam (d_{iB}). If $d_{ij} < d_{iB}$, the clusters are combined. The combined cluster becomes cluster i in the next iteration to be compared to another cluster j. The iteration stops when d_{ij} $>= d_{iB}$ with the final cluster used as the reconstructed jet.

Reconstructed jets are calibrated to have the same jet energy scale (JES) as truth jets, which are jets that are reconstructed from stable truth particles produced in the scatter [60]. Several corrections are applied in the calibration, which include changing the origin of the jet to the PV and subtracting the average p_T contribution from pile-up. The final stage of the JES calibration accounts for the remaining differences between data and MC, where the corrections are derived by comparing the p_T of the reconstructed jet in data to a balancing object whose p_T is determined to a better resolution, such as a Z-boson or a photon. The jet energy resolution (JER) of the calibrated jet, defined as

σ_{p_T}/p_T, is determined using the width of the Gaussian fit of the asymmetry between transverse momentum of the Z+jet or γ+jet and that of the reconstructed jet [61].

Jets produced in pile-up interactions can also be reconstructed. These pile-up jets are suppressed using cuts on the Jet Vertex Tagger (JVT) [62] discriminant, which is a multivariate combination of two variables related to the p_T fraction of tracks associated to the jet that comes from the PV. The JVT discriminant > 0.59 selection, which is used for the analysis in Part III, is 92% pure in hard-scatter jets and has a 2% fake rate from pile-up jets.

The flavour of a jet is defined by the hadrons and leptons inside the jet. If a b-hadron (but no c-hadron) is found inside the jet, then it is labelled as a b-jet. If c-hadron (but no b-hadron) is found inside the jet, then it is labelled as a c-jet. If a τ-lepton is found (but no b- and c-hadrons) is found inside the jet, then it is labelled as a τ-jet. If neither b- or c-hadrons or τ-leptons are found, then it is labelled as a light-jet.

4.4 Missing transverse momentum

The conservation of momentum requires the vectorial sum of the momentum in the transverse plane of an event to be zero. However, this assumption does not hold if neutrinos are produced as they only interact weakly, hence the probability of them interacting with the detector is negligible. An energy imbalance attributed to the neutrinos defines the missing transverse momentum, $\boldsymbol{E_T^{miss}}$, which is calculated from its components in the transverse plane [63]:

$$\boldsymbol{E_{x(y)}^{miss}} = \boldsymbol{E_{x(y)}^{electron}} + \boldsymbol{E_{x(y)}^{photon}} + \boldsymbol{E_{x(y)}^{tau}} + \boldsymbol{E_{x(y)}^{muon}} + \boldsymbol{E_{x(y)}^{jet}} + \boldsymbol{E_{x(y)}^{soft}}, \tag{4.4}$$

where $\boldsymbol{E_{x(y)}^{object}}$ is the x (y) component of the contributions from the reconstructed objects. The soft term contribution, $E_{x(y)}^{soft}$, includes the p_T sum of all tracks matched to a PV but not associated to any of the reconstructed objects. The scalar total missing transverse momentum is given by

$$E_T^{miss} = \sqrt{(\boldsymbol{E_x^{miss}})^2 + (\boldsymbol{E_y^{miss}})^2}. \tag{4.5}$$

Chapter 5

Machine Learning at the LHC

Machine learning (ML) can be broadly defined as the study of computational algorithms that learn from data and improve with experience without explicitly being programmed. As the growth of data increases rapidly, the use of ML to extract information is crucial in many fields including high energy physics, where ML is an established tool in many physics analyses [64].

The two main types of ML systems are supervised and unsupervised learning. In supervised learning, the training data provided to the algorithm includes both the inputs and the desired outputs, which the training attempts to replicate. The training typically involves an optimisation problem in which the objective is to minimise a loss function. In unsupervised learning, only the inputs are known, hence the algorithm draws inference from the data. As only supervised learning is used for the work presented in this thesis, only this type of learning will be detailed in this chapter.

Problems within supervised learning are categorised into classification if the output is discrete or regression if output is continuous. The type of classification can be binary, such as in the distinction of signal and background, or multiclass, such as in discriminating between b-, c-, or light-flavour jets.

5.1 Boosted decision trees

A boosted decision tree (BDT) [65] is a forest of decision trees with boosting algorithms. BDTs are widely used in physics analyses to separate signal from background in a supervised classification configuration, and are described in more detail below.

In decision trees, as illustrated in Fig. 5.1, the tree structure starts at the initial node, known as the root node. Each node in the tree recursively splits the data into two branches using a threshold value on an event variable until a stopping criterion is

fulfilled. The terminal nodes, known as leaves, determine the outputs. The predicted class of an event is determined by the signal purity in a leaf, $p_s = p = s/(s+b)$, where s and b are the sums of weights of signal and background events, respectively. The events in a leaf with $p > 0.5$ are predicted as signal events.

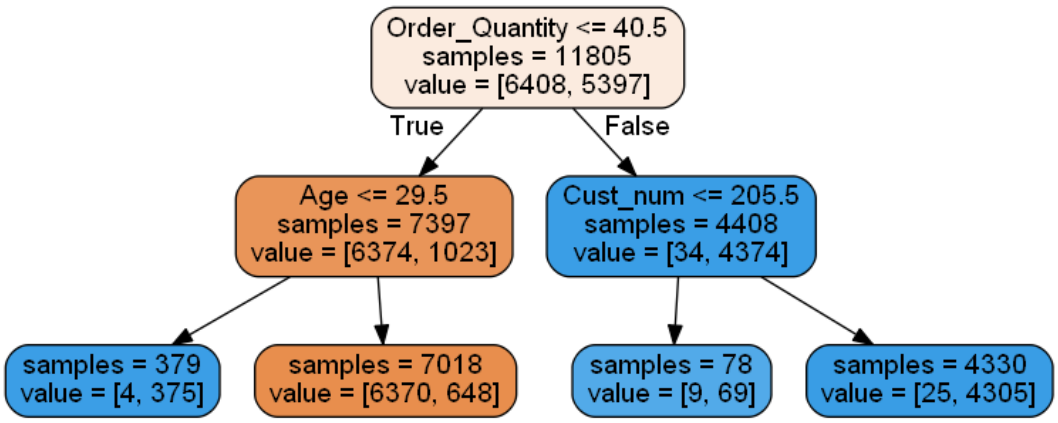


Figure 5.1: Decision tree [66].

The stopping criterion can be defined by any of the following:

- Minimum leaf size: Number of events of each node has to be above a certain proportion, e.g. 5%, of the total number of training events to ensure sufficient statistics in each leaf.
- Maximum tree depth: Number of layers of nodes from the root node to the furthest leaf node is set to a maximum, e.g. 4, to ensure the tree is not computationally expensive and to reduce overtraining.
- Perfect separation: If all events in a node belong to one class.

As all variables and their threshold values are considered at each node, the algorithm has to choose a variable for splitting that provides the best separation between the two classes. A splitting measure to describe the impurity of a node is minimised to determine the optimal variable and threshold value for splitting at that node. A commonly used impurity function is the Gini index:

$$\text{Gini} = 1 - \sum_{i=s,b} p_i^2 = 2p(1-p), \tag{5.1}$$

which involves the subtraction of the sum of squared probabilities of each class i, p_i, from unity.

5.1.1 Boosting algorithms

Classifiers that are error-prone and unstable are known as weak learners, such as decision trees. Boosting algorithms [67] are used to combine a sequence of weak learners into a more stable classifier with a smaller variance. Each weak learner in boosting learns from the mistakes made by a previous weak learner, thus giving rise to a gradual improvement in the performance of the weak learners, which will result in a strong learner. The boosting techniques explored in this thesis are adaptive and gradient boosting.

5.1.1.1 AdaBoost

In adaptive boosting (AdaBoost) [65, 67, 68], the objective is to minimise the exponential loss function. The exponential loss function can be minimised analytically, which gives rise to the weighted misclassification rate, such that events that were misclassified in iteration m, or the m^{th} weak learner, have their weights increased in the subsequent iteration $m+1$ using a boost weight.

The boost weight is specific to each iteration and is defined as

$$\alpha_m = \beta \log\left(\frac{1-\bar{\epsilon_m}}{\bar{\epsilon_m}}\right) \tag{5.2}$$

using the weighted misclassification rate of the iteration:

$$\bar{\epsilon_m} = \frac{\sum_j w_j^m}{\sum_i w_i^m}, \tag{5.3}$$

where j is restricted to the list of misclassified events and the sum over i covers all events. The learning rate of the AdaBoost algorithm, β, controls the strength of the boosting. Hence, for each event i with weight w_i that was misclassified in a given iteration m, its weight in the next iteration is

$$w_i^{m+1} = w_i^m \exp(\alpha_m \bar{\epsilon_m}). \tag{5.4}$$

As the iterations proceed, the weights of misclassified events increase and subsequent trees focus their training on the events that are difficult to classify.

The output of the BDT for an event i is given by

$$y_i(\boldsymbol{x}_i) = \frac{1}{M}\sum_{m=1}^{M}\alpha_m h_m(\boldsymbol{x}_i) \tag{5.5}$$

for M total number of iterations, i.e. trees, a set of input variables $\boldsymbol{x}$ and the prediction of a tree $h_m(\boldsymbol{x})$, which has values of +1 and -1 for signal and background, respectively.

5.1.1.2 GradBoost

In gradient boosting (GradBoost) [67], a numerical method involving gradient descent is used to minimise any loss function. TMVA [68], which is used for the work in this thesis, uses a binominal log-likelihood loss function:

$$L(F, y) = \log(1 + e^{-2F(\boldsymbol{x})y}), \tag{5.6}$$

where $F(\boldsymbol{x})$ is the response of the classifier, and y represents the set of true class values.

The gradient descent procedure involves minimising the loss by updating a set of parameters in each iteration. In gradient boosting, the procedure involves updating the weak learners in each iteration instead, such that each weak learner that is added to the classifier minimises the residual loss, which is defined as $y - F(\boldsymbol{x})$.

The algorithm is initialised by fitting the first weak learner to the training set $\{(x_i, y_i)\}_{i=1}^{n}$. In each iteration m and for each event i, the gradient is calculated as

$$\delta_{im} = -\left[\frac{\partial L(F(\boldsymbol{x}_i), y_i)}{\partial F(\boldsymbol{x}_i)}\right]_{F=F_{m-1}}, \tag{5.7}$$

where each weak learner $h_m(\boldsymbol{x})$, apart from the first, is fitted to the residuals of the previous overall classifier F_{m-1}, i.e. the weak learner is trained on the set $\{(x_i, \delta_{im})\}_{i=1}^{n}$. The classifier is updated as

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \nu h_m(\boldsymbol{x}), \tag{5.8}$$

where ν is the learning rate of GradBoost, also known as shrinkage. The final classifier response $F_M(\boldsymbol{x})$ yields the predicted class values for all events.

5.2 Deep neural networks

Neural networks [67] are composed of layers of artificial neurons, or nodes, with nodes in each layer linked to nodes in other layers via connection weights, w. A neural network (NN) typically consists of an input layer, one or more hidden layers, and an output layer, as illustrated in Fig. 5.2. An NN is described as shallow if it has one hidden layer, and deep if it has more than one hidden layer.

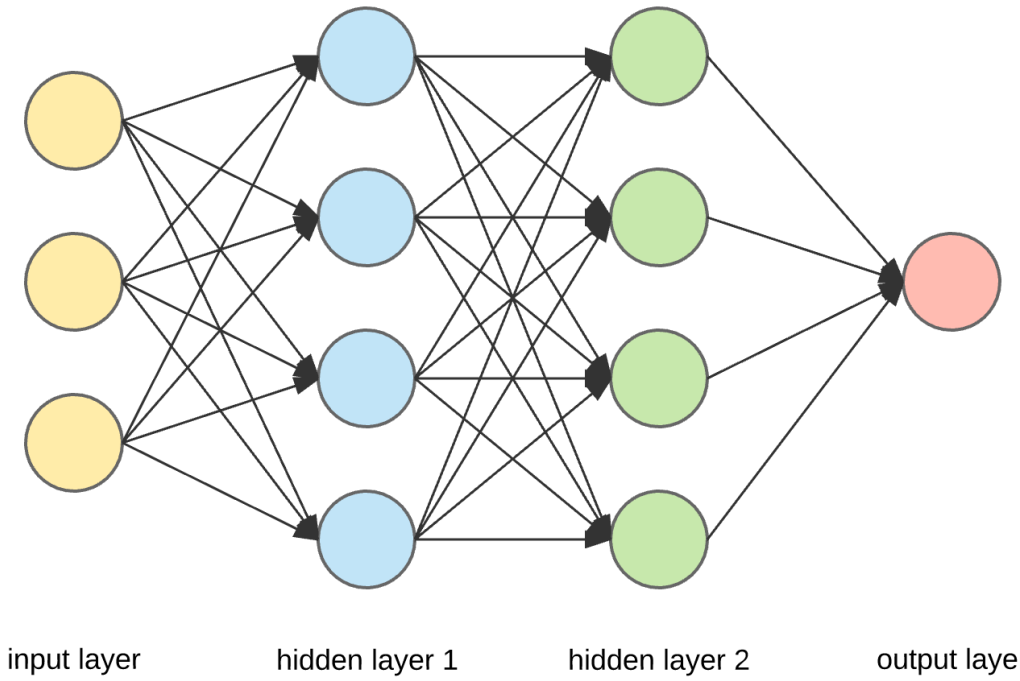


Figure 5.2: Deep neural network with three input variables corresponding to three input nodes [69].

Each node in the input layer corresponds to an input variable d. A hidden layer is formed of nodes that are not directly observed and is created from linear combinations of nodes from the previous layer, and represents derived features. If the previous layer is the input layer defined by D nodes (corresponding to the number of input variables) and the current hidden layer contains M nodes, then the value of each node in the current layer can be obtained from the elements of the matrix

$$Z = XW, \tag{5.9}$$

where X is an $N \times D$ matrix for N total number of events, W is the weight matrix of

size $D \times M$ that contains the connection weights between two nodes. With $l = 1$ as the input layer and $l = t$ as the output layer, the value of node j in a certain layer l can be expressed as

$$z_j^l = \sum_{m=1}^{M_{l-1}} w_{mj}^l a_m^{l-1} + b_j^l, \tag{5.10}$$

where w_{ij}^l is the weight between the incoming node i and node j, b_j^l is the bias with an input value of 1 for node j that is included to add flexibility to the NN model, and $a_i = \sigma(z_i)$ where $\sigma(z)$ is the activation function that is applied to every hidden layer. By defining b_j^l as w_{0j}^l corresponding to the fixed activation value of $a_0^{l-1} = 1$, the equation becomes

$$z_j^l = \sum_{m=0}^{M_{l-1}} w_{mj}^l a_m^{l-1}. \tag{5.11}$$

An output function, $g(T)$, transforms the nodes in the output layer space, T, for $k = 1, \ldots, K$ in K-class classification. For $K > 2$, such as in the case of b-, c- and light-jet classes, the softmax function:

$$g_j(T) = \frac{e^{T_j}}{\sum_{k=1}^{K} e^{T_k}} \tag{5.12}$$

is applied to each node j in the output layer to obtain the probability of each class.

To optimise the weight parameters of the NN, the quality of the model fit to the training data is quantified using a loss function. For classification, the loss function is typically defined by the cross-entropy loss:

$$\mathcal{L} = -\sum_{k=1}^{K} y_k \log(\hat{y_k}), \tag{5.13}$$

where y is the actual output class and $\hat{y_k}$ is the computed probability of class k from the function g_k. Due to the high cost of training the NN, the loss function is minimised using stochastic gradient descent (SGD), which updates the weight of a random subset of the data (minibatch) as

$$w_{t+1} = w_t - \alpha \frac{\partial \mathcal{L}}{\partial w_t}, \tag{5.14}$$

where t is the current time step and α is the learning rate. Variations of SGD involve adapting the gradient and the learning rate to speed up the convergence of the NN. For

example, the Adam [70] optimiser uses the exponential moving averages of the gradient (i.e. the momentum) and of the squared gradient to adapt the gradient and the learning rate, respectively.

In a training epoch of the NN, all training events are passed forward once and passed backward once. The backward pass stems from the backpropagation algorithm [71], which provides an analytical solution to calculating the gradients used to update the weights. By implementing the chain rule, the gradient of the loss function is written as

$$\frac{\partial \mathcal{L}}{\partial w_{ij}^{l}} = \frac{\partial \mathcal{L}}{\partial z_{j}^{l}} \frac{\partial z_{j}^{l}}{\partial w_{ij}^{l}}, \tag{5.15}$$

where

$$\frac{\partial z_{j}^{l}}{\partial w_{ij}^{l}} = a_{i}^{l-1}. \tag{5.16}$$

As $z_{j}^{l+1} = z_{j}^{l+1}(z_{j}^{l})$, the multivariate chain rule can be applied to the hidden layers to give

$$\frac{\partial \mathcal{L}}{\partial z_{j}^{l}} = \sum_{m=1}^{M_{l+1}} \frac{\partial \mathcal{L}}{\partial z_{m}^{l+1}} \frac{\partial z_{m}^{l+1}}{\partial z_{j}^{l}}, \tag{5.17}$$

which shows that the gradient in layer l is dependent on the gradient in the next layer $l+1$. Each node in layer $l+1$ is given by

$$z_{m}^{l+1} = \sum_{n=0}^{M_{l}} w_{nm}^{l+1} \sigma(z_{n}^{l}), \tag{5.18}$$

which leads to the derivative:

$$\frac{\partial z_{m}^{l+1}}{\partial z_{j}^{l}} = w_{jm}^{l+1} \sigma'(z_{j}^{l}). \tag{5.19}$$

Therefore, the gradient can flow backwards in the hidden layers by calculating the gradient of the activation function, with the algorithm initialised by computing the gradient of the loss function for the output layer. Overall, backpropagation provides an efficient way of minimising the loss function in an NN.

5.2.1 Activation functions

Activation functions are applied to introduce nonlinearity between the hidden layers, allowing the NN to learn complex patterns in the data. The activation functions typically used in NNs are the ReLU and Leaky ReLU functions, which are shown in Fig. 5.3.

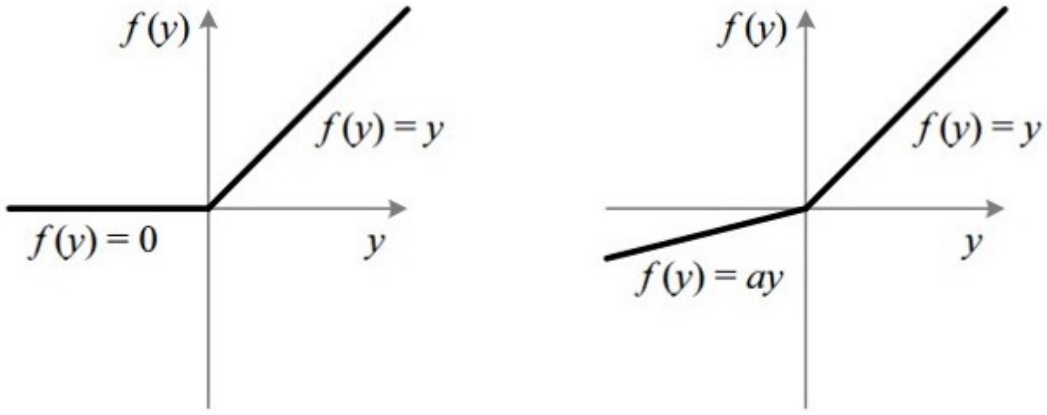


Figure 5.3: ReLU (left) and Leaky ReLU (right) functions [72].

The ReLU function $f(y) = \max(0, y)$ is the most common activation function of choice in NNs as it has a much larger range of non-zero gradients and its derivatives are much faster to compute compared to those of the sigmoid function $f(y) = 1/(1 + e^{-y})$, which is the historical default activation function of choice.

However, the ReLU function can lead to the 'dying ReLU' problem, in which some nodes of the NN effectively die in the sense that they can only output zero due to the zero gradients for $y < 0$. To deal with this problem, a variation of the ReLU function, the leaky ReLU function $f(y) = \max(\alpha y, y)$, is used. The leaky ReLU function contains a hyperparameter α that defines the slope of the function for $y < 0$, which can also be conceptualised as how much the function 'leaks', allowing non-zero gradients for $y < 0$.

Part II

b-jet identification

Chapter 6

b-jet identification

The identification of jets originating from b-quarks, known as b-tagging, is crucial in the majority of ATLAS analyses, such as the $VH(H \to b\bar{b})$ analysis, since b-jets, which are jets that contain b-hadrons, are present in many interesting physics signatures.

An important property of b-hadrons is the relatively long lifetime (most common lifetime is approximately 1.5 $\times 10^{-12}$ s [73]), which allows it to travel far enough to form the displacement of a decay point, i.e. a displaced vertex, with respect to the primary vertex (PV). Other properties of a b-hadron include a large mass of around 5 GeV and a large fraction of momentum retained from the initial b-quark. Due to the CKM transition matrix values, the decay of a b-quark into a c-quark is highly favoured. As c-hadrons can also have a relatively long lifetime, this results in a distinctive decay chain. These properties are illustrated in Fig. 6.1.

In ATLAS, there are low-level and high-level b-tagging algorithms [75], which are designed to discriminate b-jets against the much larger c-jets and light-flavoured jets (light-jets), which arise from gluons and u, d, and s quarks, background. The efficiency to tag jets of flavour j is defined as

$$\epsilon_j = \frac{\text{Number of jets of flavour } j \text{ tagged by the algorithm}}{\text{Total number of jets of flavour } j}, \tag{6.1}$$

which is used to quantify the performance of the algorithms used in the identification of the flavour quark j that the jet originated from. Specifically, the performance of b-tagging algorithms is quantified by the b-jet tagging (b-tagging) efficiency, ϵ_b. The rejection, $1/\epsilon_j$, of c- or light-jets is often compared against the b-jet tagging efficiency to evaluate the overall performance of the b-tagging algorithm, in the form of receiver operating characteristics (ROC) curves [76].

The performance of the high-level algorithms is also evaluated using b-tagging

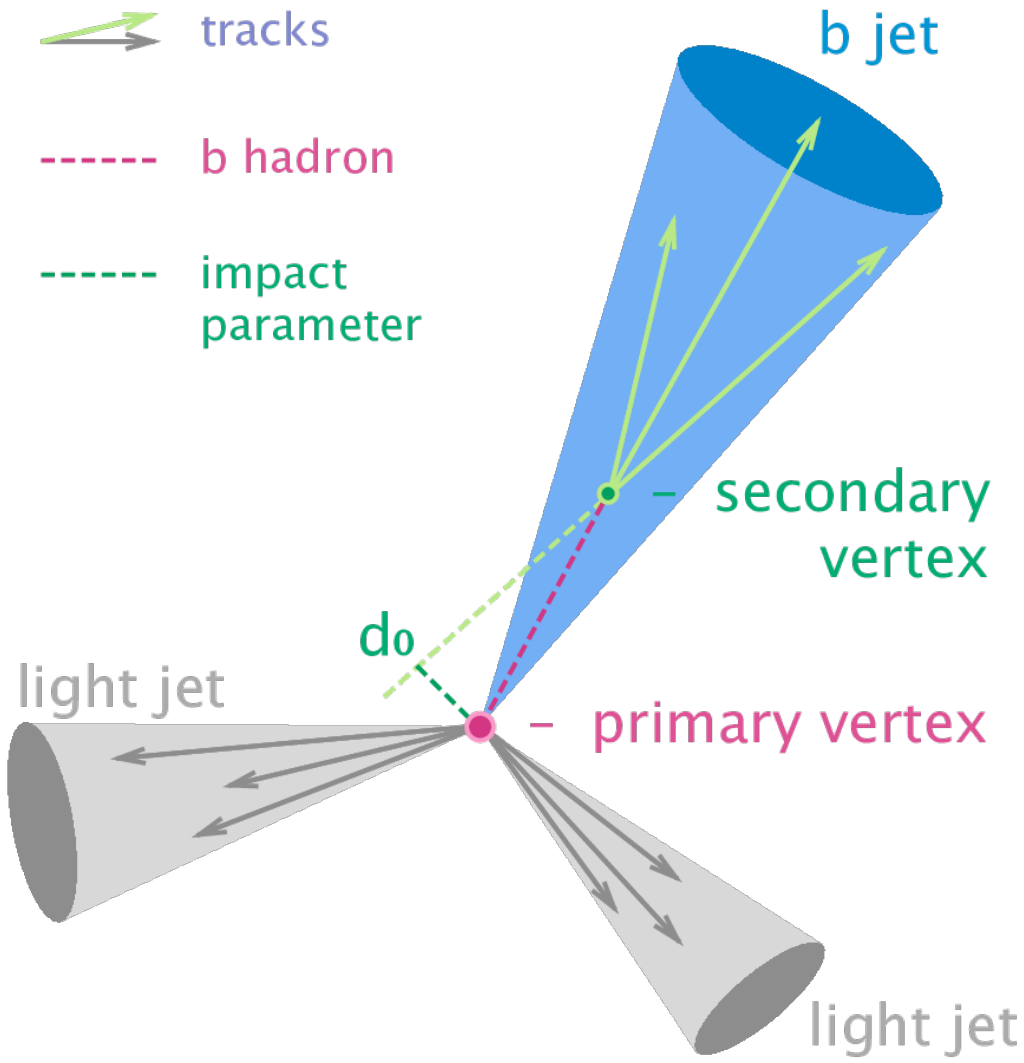


Figure 6.1: The properties of b-jets that are used in b-tagging [74].

efficiency working points (WPs), which are selections applied on the b-tagging algorithm discriminant to target specific b-tagging efficiencies. The single-cut WPs corresponding to 60%, 70%, 77%, and 85% b-tagging efficiencies form the binned distribution of the output of the b-tagging algorithm, known as the pseudo-continuous b-tagging discriminant. The probability density of each bin is known as the b-jet tagging probability, which is used to determine the b-tagging efficiency for any single-cut WP, i.e. $\epsilon_b = X\%$, by summing the b-jet tagging probabilities in the range $[X\%, 0\%]$.

6.1 Low-level algorithms

Low-level b-tagging algorithms (b-taggers) are designed to distinguish b-jets from c- and light-jets by exploiting the distinct properties of b-jets.

6.1.1 Impact Parameter based

Due to the long lifetime of b-hadrons, the tracks produced from b-hadron decay products tend to have large IPs (Section 3.2.1), and hence can be separated from the tracks originating from the PV. This property is employed by the impact parameter based b-taggers [77], IP2D and IP3D. The transverse IP is used by IP2D, while both the transverse and longitudinal IPs are used by IP3D.

While tracks from b- and c-hadrons tend to have positive IPs, tracks associated with light-jets tend to have an approximately symmetrical distribution of IPs centred around zero. The transverse and longitudinal IP significances are defined with respect to their uncertainties (σ) as d_0/σ_d and z_0/σ_z, respectively. Probability density functions (PDFs) of the signed IP significance for b-, c-, and light-jets are evaluated. These PDFs are used to compute the log-likehood ratio (LLR) discriminants between the different jet flavours.

6.1.2 Secondary Vertex based

Secondary vertex (SV) based algorithms [78] aims to reconstruct the SV of the b-hadron decay. First, two-track vertices in a jet are reconstructed, then the algorithms combine all tracks from these vertices, excluding vertices that are likely to have come from long-leved particles, such as K_s and Λ, photon conversions, or hadronic interactions with the detector material. An iterative procedure is used to reconstruct the SV by removing the least compatible tracks until a good χ^2 of the vertex fit is achieved. Discriminating variables constructed from the properties of the reconstructed SV, such as the invariant mass of the reconstructed SV, are used to distinguish b-jets from c- and light-jets.

6.1.3 Decay chain multi-vertex reconstruction (JetFitter)

As $|V_{cb}|^2 \gg |V_{ub}|^2$ [73], the preferred full b-hadron decay chain starts from the PV to the SV of the b-hadron decay then to the tertiary vertex of the c-hadron decay. The decay chain multi-vertex reconstruction algorithm, known as JetFitter [79, 80], attempts to reconstruct this b-hadron decay chain by assuming that the three vertices lie on the same flight path.

JetFitter is initialised with a PV and a b-hadron flight axis, initially represented by the jet axis, with each track that crosses the flight axis representing a displaced vertex. A Kalman filter is used to reconstruct the flight path on which the primary, secondary, and tertiary vertices lie. A set of discriminating variables are derived from the reconstructed secondary and tertiary vertices, such as the mass, number of tracks, and fraction of the jet energy in both types of vertices.

6.2 High-level algorithms

High-level b-taggers combine the output from low-level b-taggers to make use of the correlations between the different variables to further discriminate b-jets against the backgrounds.

6.2.1 Multivariate (MV2)

The first category of multivariate algorithms, MV2 [75], uses boosted decision trees (BDTs) trained on jets from $t\bar{t}$ events to output a single b-tagging discriminant that separates b-jets (signal) from c- and light-jets (background). In addition to the output variables from the low-level b-taggers, the jet p_T and η are provided to the BDT to take advantage of their correlations with the output variables. The performance of the c-jet rejection versus the light-jet rejection can be varied by changing the fractions of c- and light-jets in the background training sample. The name of the MV2 taggers are defined to indicate the c-jet and light-jet fractions, for example in MV2c10, the background is composed of approximately 10% c-jets and 90% light-jets, which is used in the analysis in Part III.

6.2.2 Deep learning (DL1)

A deep neural network, known as DL1 [75], is used as the new default flavour identification algorithm in ATLAS. The algorithm treats the jet flavours equally in training, and hence the algorithm can be used for both b- and c-tagging. In addition to the MV2 input variables, DL1 uses JetFitter c-jet tagging variables to enhance its c-jet rejection power. The algorithm has three class outputs corresponding to probabilities of the b, c and light jet flavours, denoted by p_b, p_c, and p_u, respectively. The final discriminant for b-tagging is determined by the combining the three outputs into a single discriminant that is defined as a function of the effective fraction of c-jets in the background training sample (f_c):

$$\mathrm{DL1}(f_c) = \log\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u}\right) \tag{6.2}$$

which allows the c-jet fraction to be tuned.

Chapter 7

Track reconstruction improvements

Track reconstruction, as detailed in Section 4.1, is a key component in b-tagging algorithms, as well as being crucial for the reconstruction of all objects in the detector. As the centre-of-mass energy increases at the LHC, an increasingly large amount of events with dense environments are generated. These events involve high p_T jets and are interesting in the searches for new physics, including probing high p_T Higgs bosons. In the core of high p_T jets, the separation between the particles becomes comparable to the size of the clusters in the ID layers, which can reduce the reconstruction efficiency of tracks. Therefore, enhancing the tracking performance in dense environments is vital. However, how such improvements are fed through to flavour tagging is unclear, hence it is important to understand which elements of the tracking have the biggest impact on b-tagging, thus efforts can be focused on improving those elements.

Different track collections are defined to study various elements of the tracking, so their impact upon flavour tagging can be studied:

- **Nominal:** Standard ATLAS reconstructed tracks.
- **Pseudo:** Tracks are reconstructed using reconstructed hits assigned to each track using truth information from simulation (i.e. ideal pattern recognition), which replaces all steps of the track reconstruction procedure except for the final track fit.
- **Ideal:** Tracks that are created from pseudo tracks but using reconstructed hits that have their positions altered to the true position in simulation. Thus, these tracks are expected to have the same reconstruction efficiency as pseudo tracks, but with more precisely reconstructed parameters.
- **Fakes-removed (FR):** Same as the nominal collection, but fake-tracks are removed. A fake-track is a track that corresponds to a random combination of hits

in the detector, which do not correspond to the trajectory of a truth particle. For the results presented in this chapter, tracks with truth match probability (TMP) < 0.75 are removed.

- **Pseudo-replaced (PR):** Replace tracks in the nominal collection with the corresponding pseudo track. This collection has the same tracking efficiency as the nominal collection, but the reconstruction will be improved by the correct assignment of all hits to the track.

To compare the performance of the different tracking improvements and understand their impact on b-tagging, the track collections were evaluated on b-tagging algorithms. The effects of the improvements were tested on low-level b-taggers, such as JetFitter, to ensure that the different track collections are performing as expected.

7.1 Simulated samples

The different track collections were generated using simulated samples, which were passed through the full ATLAS detector simulation [81] with GEANT4 [82] for the interaction of particles and Evtgen [83] for the decay of heavy-flavour hadrons. The samples used for the studies in this chapter are

- $\boldsymbol{Z'}$**:** The Z' sample involves a hypothetical BSM gauge boson with the following artificial BRs: b-jets (30%), c-jets (30%), light-jets (30%), τ (5%), and e (5%). The samples are simulated using Pythia 8 [35] for the underlying event with the NNPDF2.3LO [84] PDF set.

- $\boldsymbol{t\bar{t}}$**:** The MC samples for the non-all hadronic (semi-leptonic and di-leptonic) $t\bar{t}$ processes are generated using Powheg [85] with the top-quark mass set at 172.5 GeV, interfaced to Pythia 8 with the NNPDF3.0NNLO [86] and NNPDF2.3LO PDF sets for the matrix element calculation and the simulation of parton shower, respectively.

7.2 JetFitter

The impact of the different track collections was investigated to ascertain the elements in tracking that cause the largest improvements in JetFitter. Pseudo track replacement was applied separately to heavy-flavour tracks (HF), fragmentation (FRAG) tracks and both types of tracks to further deduce how JetFitter responds to the tracking. The HF and FRAG tracks are nominal tracks that have an associated truth track with TMP $>$ 0.75. If a track originates from a b- or c-hadron, then it is labelled as a HF track, else it is labelled as a FRAG track. The overall impact of the different track collections is determined from the ROC curves of JetFitter.

7.2.1 Distributions of output variables

The set of output variables from the JetFitter algorithm [80] (Section 6.1.3) is summarised in Table 7.1. These variables are also used as input variables by DL1.

Variable	Description
m	Invariant mass of tracks from displaced vertices
f_E	Fraction of the charged jet energy in the SVs
$\Delta R(p_{\text{jet}}, p_{\text{vtx}})$	ΔR between jet axis and vectorial sum of momenta of all tracks attached to displaced vertices
S_{xyz}	Significance of the average distance between PV and displaced vertices
N_{TrkAtVtx}	Number of tracks from multi-track displaced vertices
N_{2TrkVtx}	Number of two-track vertex candidates prior to decay chain fit
$N_{\text{1-trk vertices}}$	Number of single-track displaced vertices
$N_{\geq\text{2-trk vertices}}$	Number of multi-track displaced vertices

Table 7.1: Output variables of the JetFitter algorithm.

Distributions of the most important output variables between the different track collections for different jet flavours are demonstrated in Fig. 7.1, 7.2, and 7.3. In all distributions, the FR collection displayed the most similar behaviour to the nominal collection, which is expected as the two track collections are the same except for the removal of fake tracks, which make up only approximately 8.2% of the tracks. It can be seen from the distribution of m that as the track collections change from nominal to pseudo to ideal that the mass of b-jets increases and the mass of light-jets decreases, providing a better discrimination between the two jet flavours. From the f_E distribution,

pseudo and ideal collections result in lower jet energy fraction in the SVs for c- and light-jets, and for b-jets, but to a lower extent. While the number of multi-track displaced vertices in the pseudo and ideal collections decreases for c- and light-jets, and increases for b-jets. These observations imply that pseudo and ideal collections are less likely to contain fake displaced vertices, and more accurately reconstruct the attributes of the b-jets, which will help improve the b-tagging performance. Based upon how the different track collections impact the distributions of the variables, it suggests that replacing the reconstructed tracks with pseudo tracks (PR collection) has a bigger impact than either removing the fake tracks (FR collection) or enhancing the efficiency (pseudo collection).

7.2.2 Performance

The effect of the different track collections on the performance of JetFitter is presented in Fig. 7.4. As expected, the ideal track collection offered the best performance for both c-jet and light-jet rejections, which is closely followed by the pseudo collection, demonstrating that the correct hit assignment is vital for b-tagging performance, although using the ideal hit positions is less important as the performance between the pseudo and ideal collections is not much different. The PR and FR + PR collections yielded similar performance to the pseudo collection, suggesting that the correct reconstruction of tracks is more important than obtaining the maximum possible tracking efficiency. PR does not suffer from the lower efficiency of b-tracks compared to the pseudo collection. The FR collection also showed similar performance to the nominal collection, indicating that JetFitter is not very sensitive to the presence of fake tracks.

Further investigation into applying pseudo tracks separately to heavy-flavour tracks (HF), fragmentation (FRAG) tracks and both types of tracks in Fig. 7.5 revealed that applying pseudo tracks to FRAG only yielded nearly the same performance in light-jet rejection compared to applying pseudo tracks to HF + FRAG, signalling that the quality of FRAG tracks is essential for light-jet rejection. This is likely a consequence of HF tracks being tagged irrespective of the quality of the b-decay tracks, implying that the dominant mechanism for the worse performance stems from misreconstructed tracks in light-jets appearing to have lifetimes and consequently being used to construct displaced decay vertices. In c-jet rejection, the quality of both HF and FRAG tracks are important,

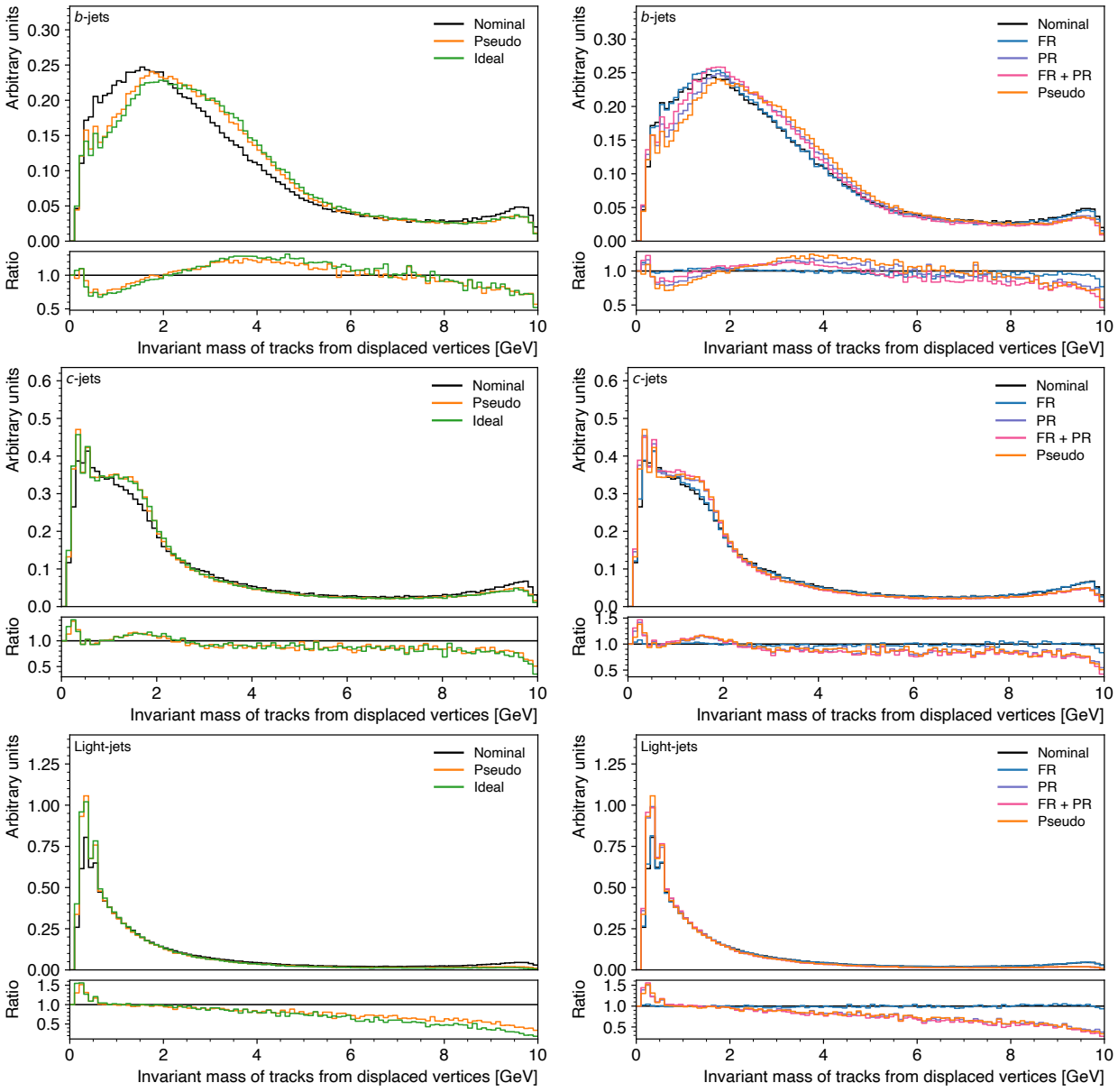


Figure 7.1: Distributions of the invariant mass of tracks from displaced vertices evaluated on the Z' sample using JetFitter with replacing the nominal tracking with ideal track collections (left) and altering the nominal tracking with enhanced tracking (right), separated into b- (top), c- (middle) and light- (bottom) jets.

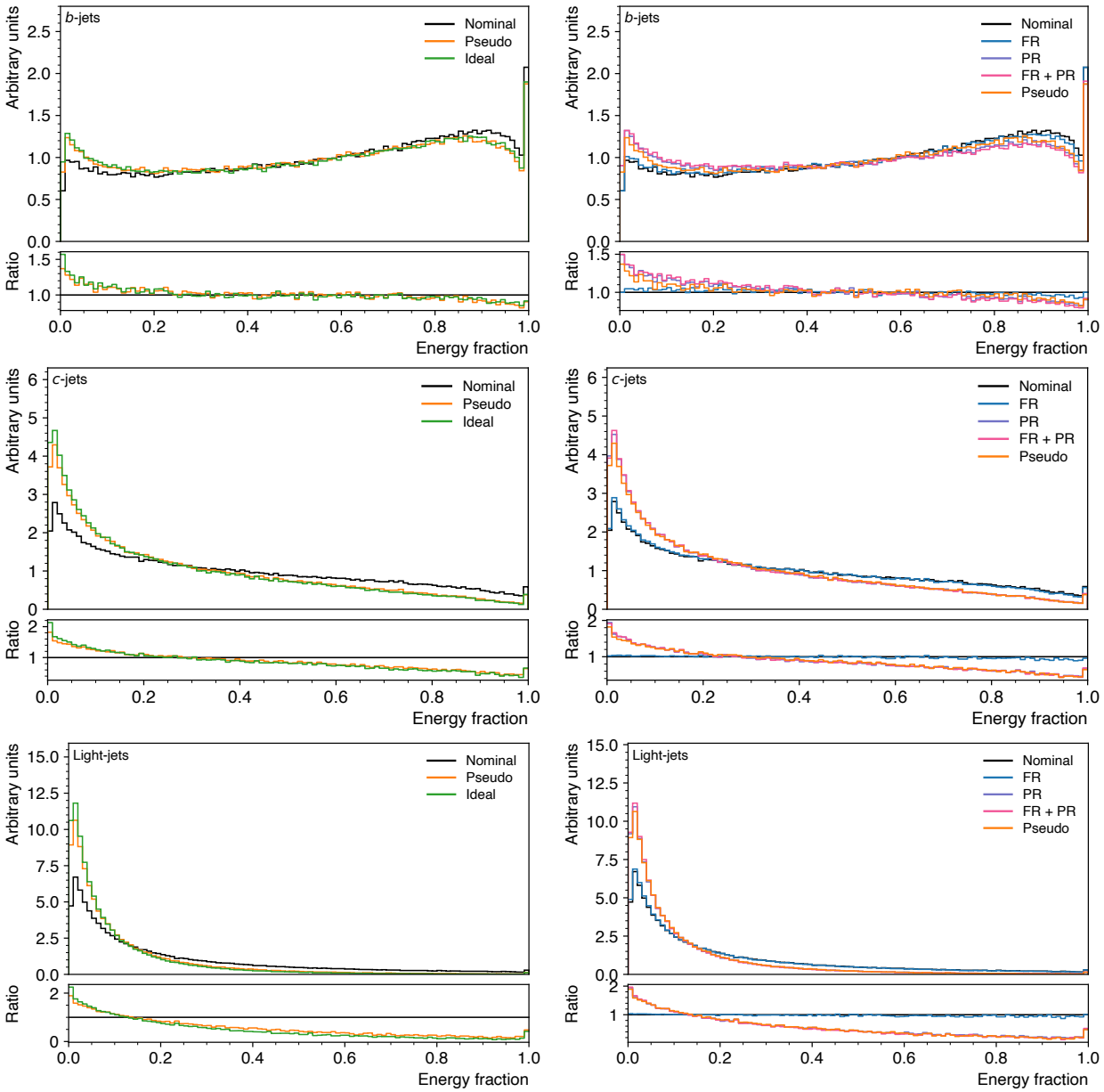


Figure 7.2: Distributions of the jet energy fraction in secondary vertices evaluated on the Z' sample using JetFitter with replacing the nominal tracking with ideal track collections (left) and altering the nominal tracking with enhanced tracking (right), separated into b- (top), c- (middle) and light- (bottom) jets.

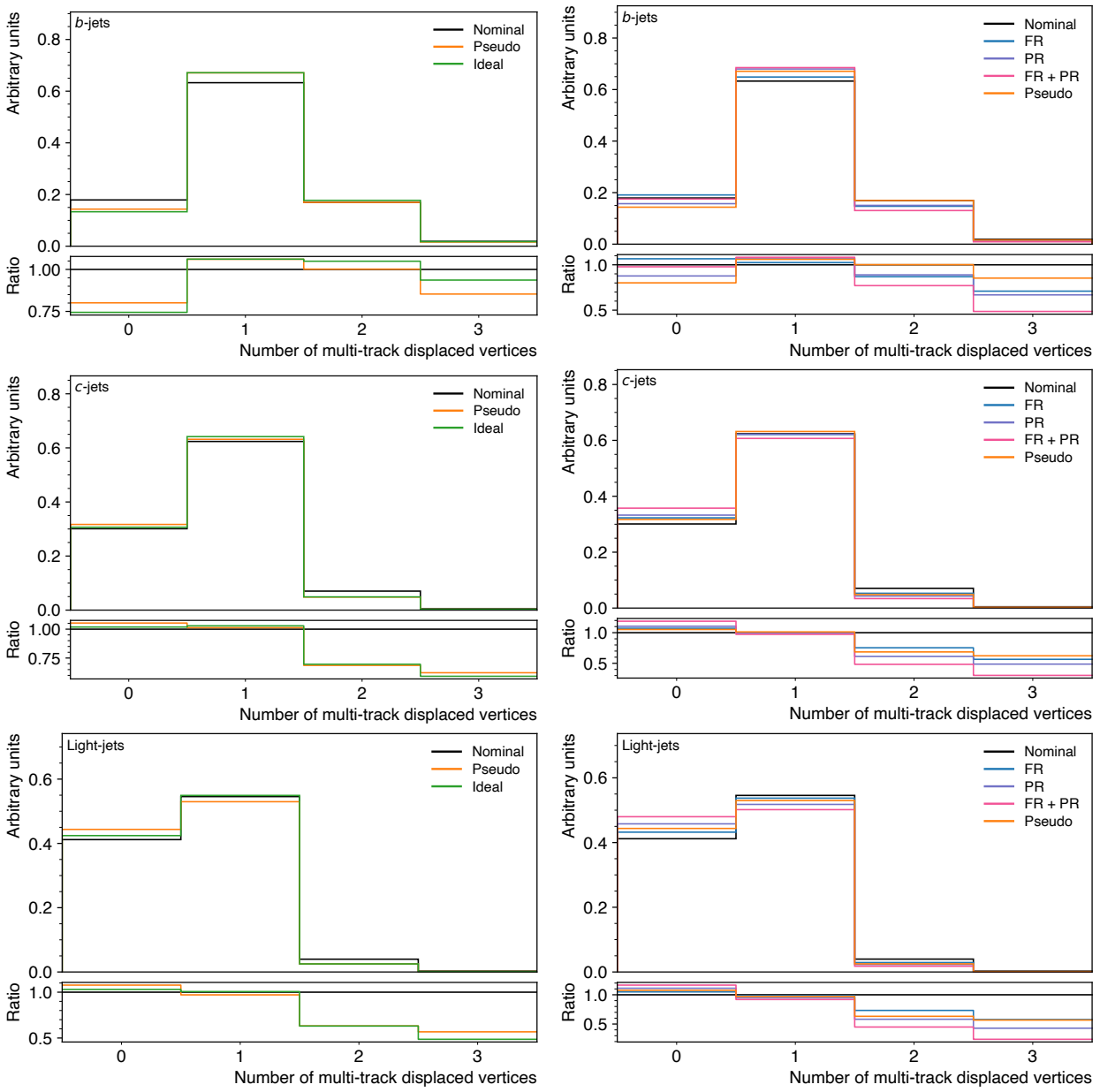


Figure 7.3: Distributions of the number of multi-track displaced vertices evaluated on the Z' sample using JetFitter with replacing the nominal tracking with ideal track collections (left) and altering the nominal tracking with enhanced tracking (right), separated into b- (top), c- (middle) and light- (bottom) jets.

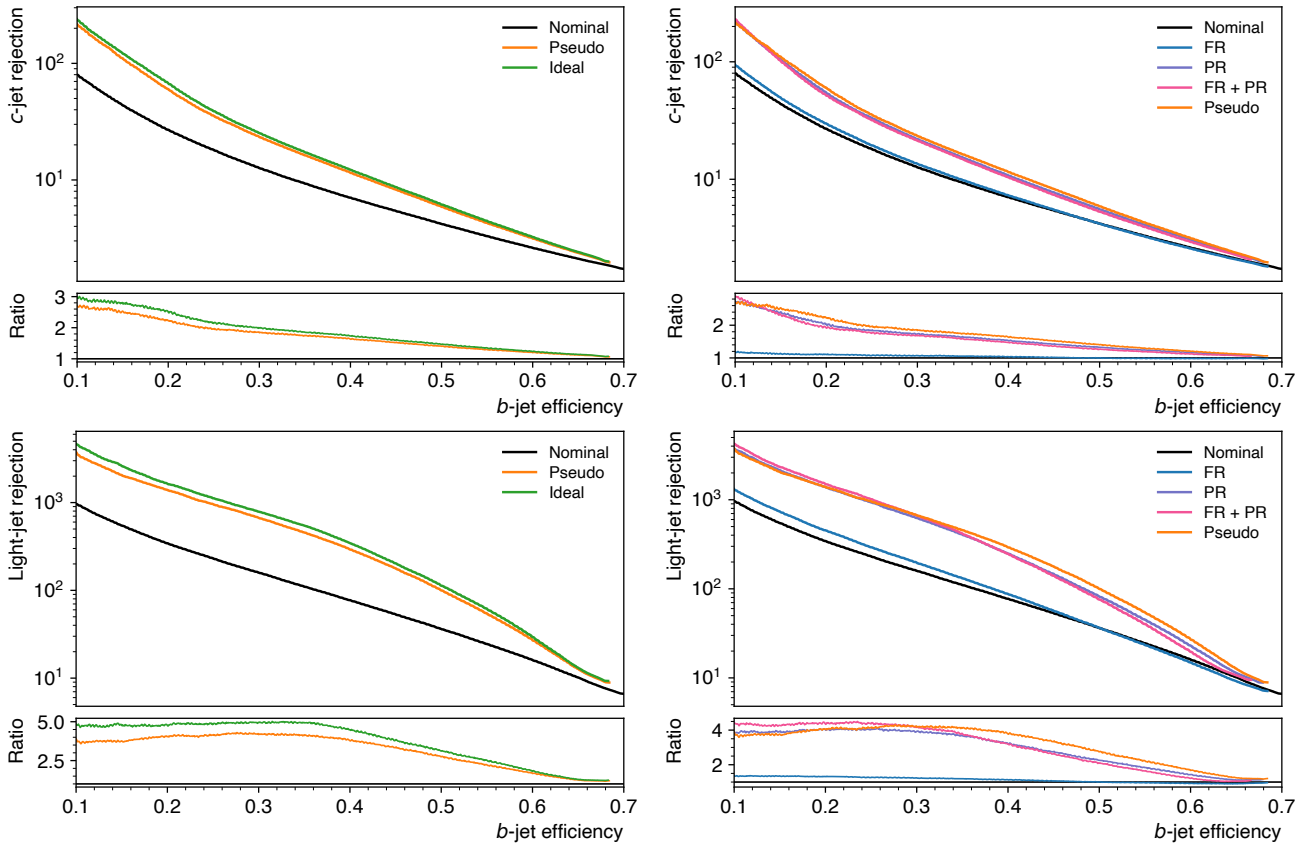


Figure 7.4: ROC curves of JetFitter evaluated on the Z' sample using different track collections for c-jet rejection (top) and light-jet rejection (bottom).

which is expected as c-jets are also categorised as HF tracks. This indicates that c-jets are being incorrectly tagged due to poorly reconstructed FRAG tracks as well as poorly reconstructed c-tracks, therefore it is important to improve both types of tracks.

The results revealed that reducing the number of fake displaced vertices from c- and light-jets will improve b-tagging from the perspective of the JetFitter performance, while removing fake tracks based on the TMP criterion does not affect the JetFitter performance. In addition, it is clear that the quality of FRAG tracks is important to light-jet rejection in JetFitter, as it is crucial to improve the performance of the reconstruction of light-jets, which appears to be more important than improving the efficiency to reconstruct tracks from b-hadron decays in terms of the performance of JetFitter. Whereas the quality of both HF and FRAG tracks are important in c-jet rejection, thus having better reconstructed c-tracks and FRAG tracks should improve the performance of JetFitter. Overall, this suggests, based upon the JetFitter b-tagging tool, that the largest impact upon b-tagging will come from enhancing the hit selection

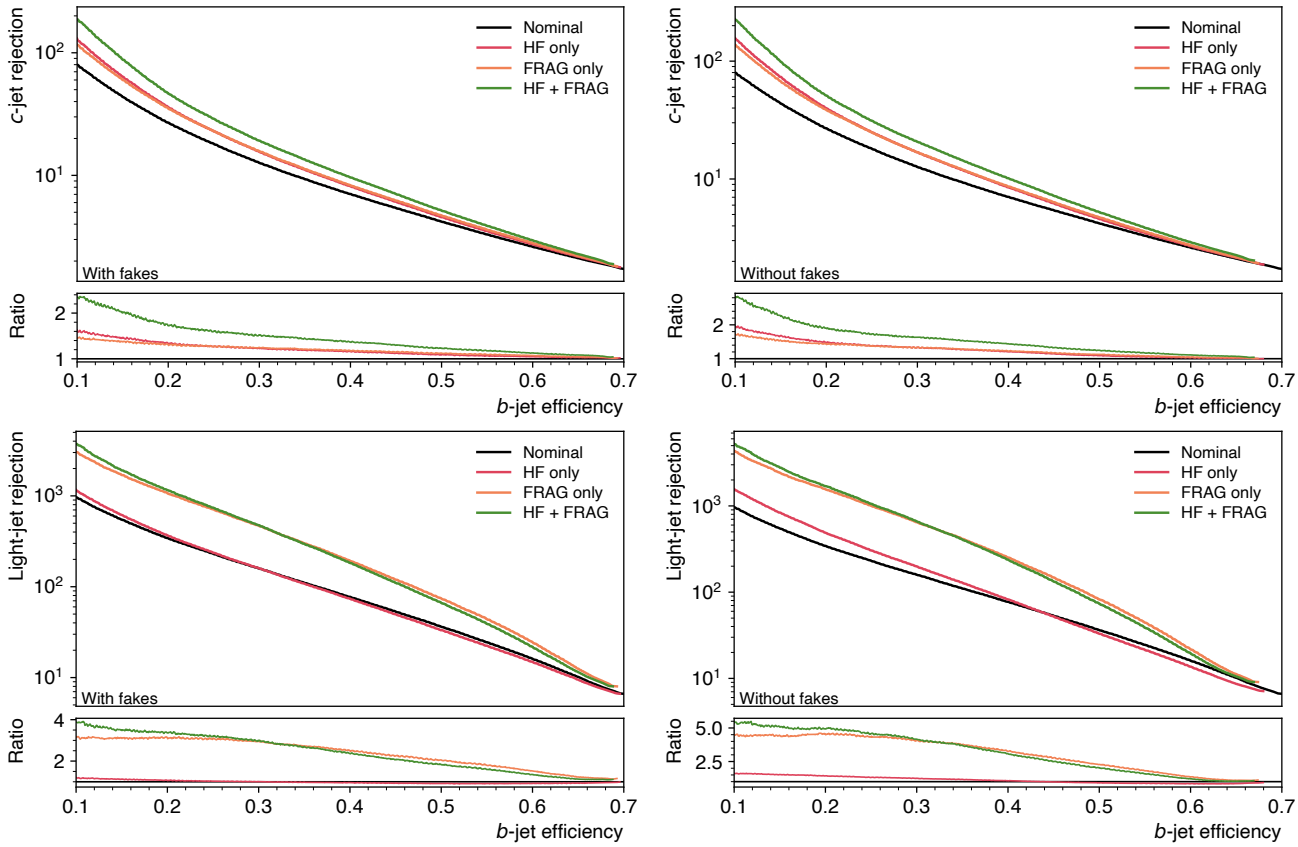


Figure 7.5: ROC curves of JetFitter evaluated on the Z' sample from applying pseudo tracks to tracks with (left) and without (right) fake tracks for c-jet rejection (top) and light-jet rejection (bottom). The pseudo tracks were applied separately to heavy-flavour tracks (HF), fragmentation tracks (FRAG), and both types of tracks.

for fragmentation tracks.

7.3 Conclusion

Different track collections were evaluated on the JetFitter b-tagging algorithm. In general, the results were consistent with the expectation that pseudo and ideal collections will offer better performance than the nominal collection.

The performance of the nominal and FR collections showed that JetFitter was not sensitive to fake tracks as defined by the TMP criterion. Applying pseudo tracks separately to HF, FRAG, and HF + FRAG tracks revealed that the quality of FRAG tracks is important in light-jet rejection, while the quality of both HF and FRAG tracks are crucial in c-jet rejection. These findings imply that, for this version of these algorithms, improving the efficiency with which we reconstruct tracks from b-

hadron decays is not that important, although further studies should be undertaken to understand if retuning the low-level taggers to use the additional reconstructed tracks could alter this conclusion. Other potential future work involves verifying on retuned higher-level taggers, such as DL1, if the same conclusions hold.

Chapter 8

Improving multivariate b-tagging algorithms

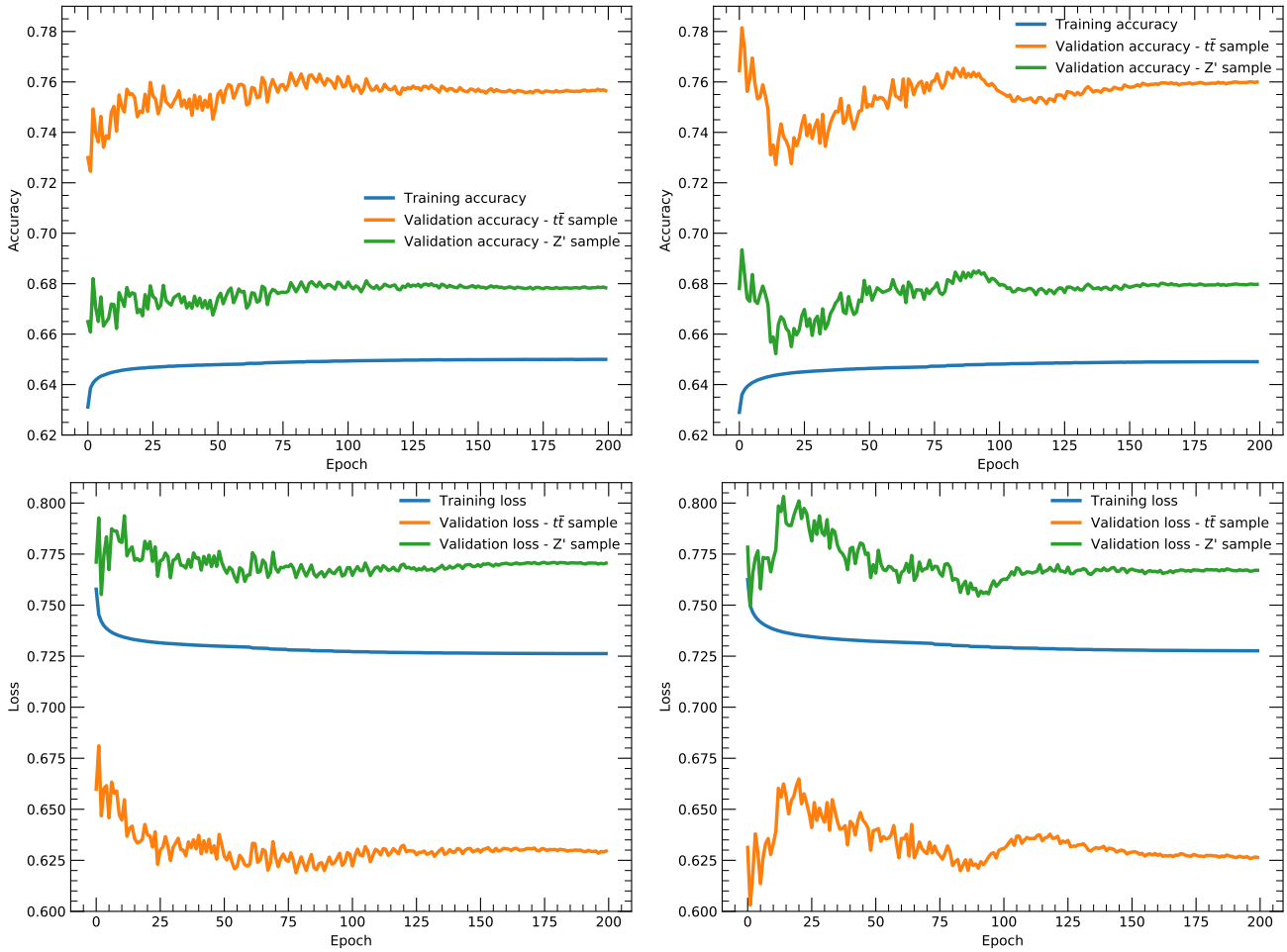
Optimisation studies into DL1 were conducted to see if further improvements to b-tagging can be achieved. As the activation functions and learning rate can have a significant impact on the training performance of neural network (NN) models, these were explored in the optimisation of DL1.

8.1 Activation functions

Activation functions are essentially components that allow NNs to learn, and thus are crucial to the accuracy and convergence of NNs [87]. The default activation function used in the DL1 architecture is the ReLU function. This was compared to the Leaky ReLU function, which aims to solve the 'dying ReLU' problem (Section 5.2.1).

The impact of the different activation functions on the training and validation accuracy and loss of DL1 is illustrated in Fig. 8.1. It can be observed that the validation accuracy and loss reach an optimum at around 87 epochs for both activation functions, hence the training of DL1 was stopped at this number of epoch to prevent overtraining.

The models were evaluated to obtain the b-jet, c-jet, and light-jet efficiencies. Fig. 8.1 compares the ROC curves from the models, and shows that using the Leaky ReLU function did not offer an improvement over the default ReLU function. However, this could be due to the choice of the hyperparameter α in the Leaky ReLU function, which was set to 0.3 in this optimisation study. Thus, an optimisation of α could result in a better performance over the ReLU function.



(a) ReLU (b) Leaky ReLU

Figure 8.1: Training and validation accuracy (top) and loss (bottom) during training of DL1 using the ReLU (left) and Leaky ReLU (right) activation functions. The training sample is a hybrid of the $t\bar{t}$ and Z' samples.

8.2 Cyclical learning rate

The cyclical learning rate [88] involves increasing and decreasing the learning rate over a number of cycles. It aims to improve the generalisation and performance of NN models by taking advantage of the loss function topology. The difficulty in minimising the loss is often due to saddle points rather than poor local minima, which slows down the convergence of the NN. Furthermore, sharp minima can often lead to poor generalisation. Hence, increasing the learning rate occasionally allows the model to move out of sharp minima. While this temporarily decreases the loss, ultimately it allows for the model to converge to a better minimum.

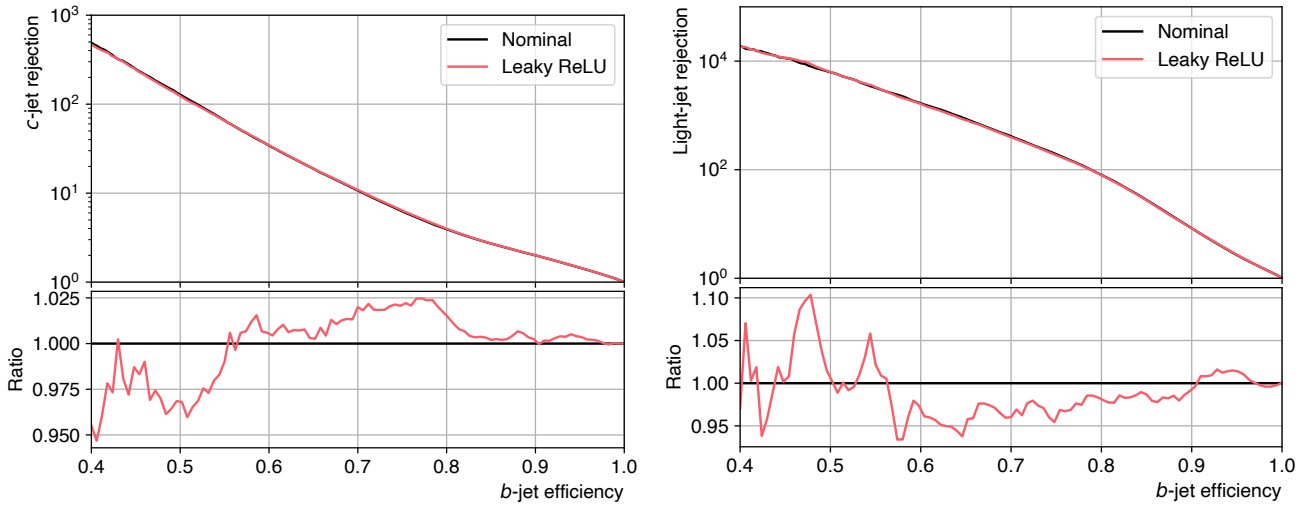


Figure 8.2: ROC curves of DL1 using the ReLU and Leaky ReLU activation functions for c-jet rejection (left) and light-jet rejection (right).

The cycle length is determined by a multiple of the number of iterations in an epoch, which is equivalent to the training set size divided by the batch size. The triangular policy, where the learning rate varies linearly between the maximum and minimum learning rate boundary, and the exponential range policy, which is the triangular policy with boundary values declining by an exponential factor, were tested on DL1. The policies are presented in Fig. 8.3.

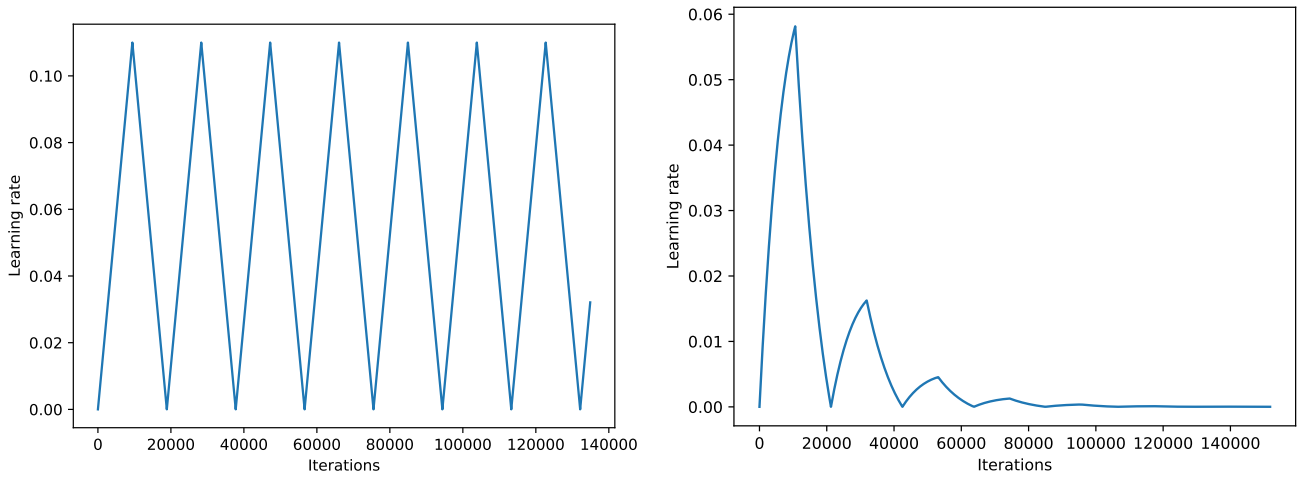
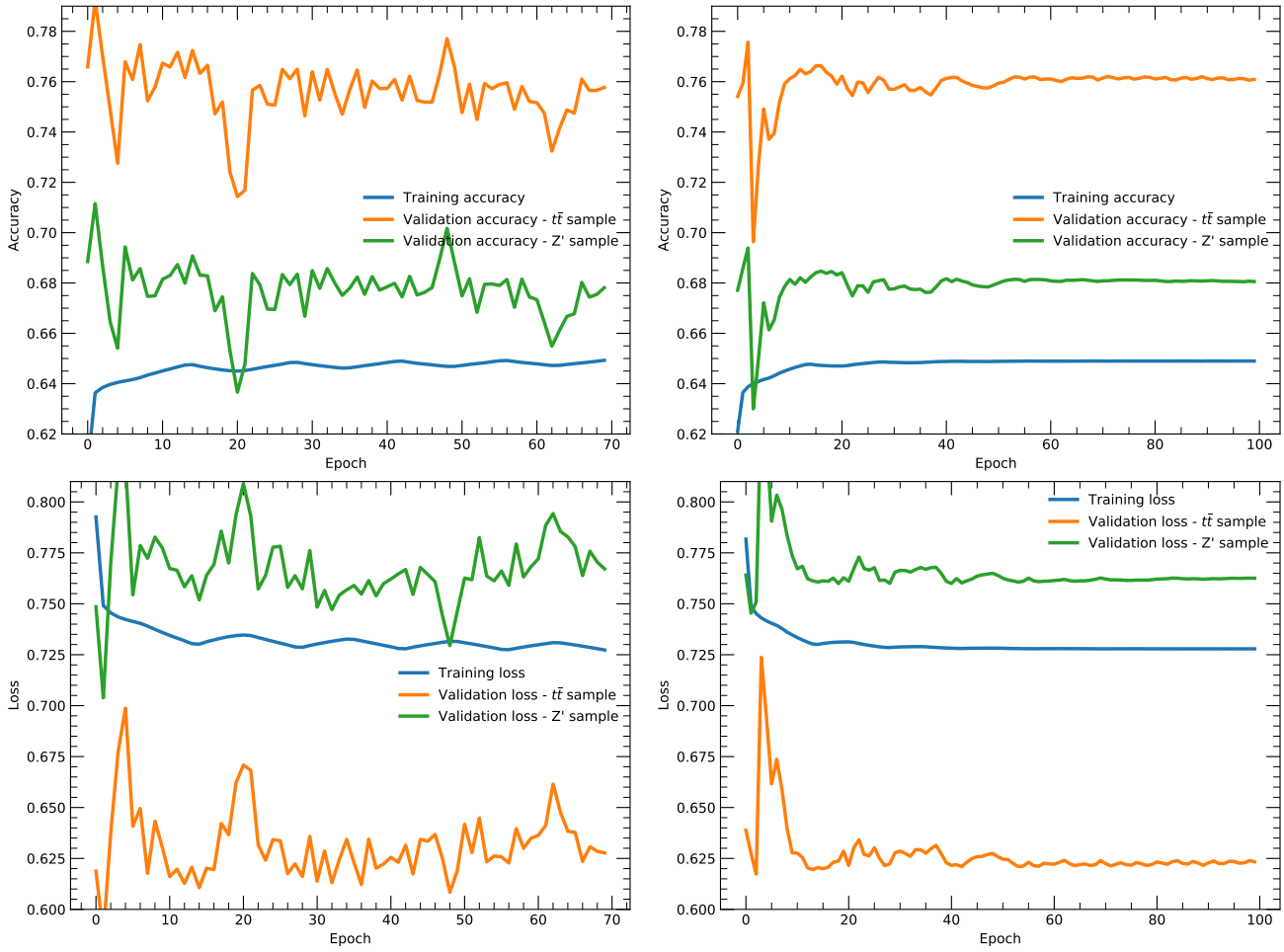


Figure 8.3: Learning rate over one epoch of DL1 using a cyclical learning rate with the triangular (left) and the exponential range (right) policies.

Fig. 8.4 demonstrates the impact of the different policies on the accuracy and loss, and reveals that an optimum is achieved for the validation accuracy and loss at around 48 and 15 epochs for the triangular policy and the exponential range policy, respectively.

The ROC curves from the different policies compared to the default constant learning rate of 0.01 are displayed in Fig. 8.5. Overall, the exponential range policy resulted in a comparable performance to DL1, with the triangular policy resulting in a slightly worse performance.



(a) Triangular

(b) Exponential range

Figure 8.4: Training and validation accuracy (top) and loss (bottom) during training of DL1 using a cyclical learning rate with the triangular (left) and exponential range (right) policies. The training sample is a hybrid of the $t\bar{t}$ and Z' samples.

8.3 Summary

Activation functions and cyclical learning rates were explored in the optimisation of DL1, however both investigations offered very similar performance to the default DL1 model. Though, both studies involved hyperparameters, which may significantly affect the results.

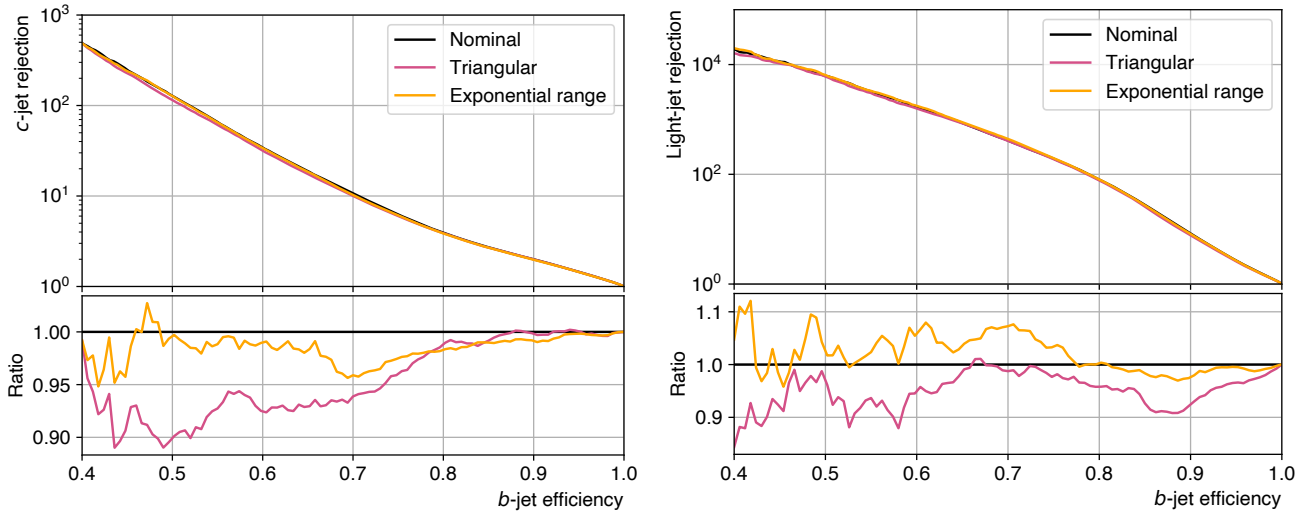


Figure 8.5: ROC curves of DL1 using a cyclical learning rate with the triangular and exponential range policies for c-jet rejection (left) and light-jet rejection (right).

Chapter 9

Automation of the b-tagging efficiency calibration

Due to imperfect physics and detector modelling in simulation, the b-jet tagging efficiency predicted by b-tagging algorithms in simulation does not fully represent the performance of b-tagging algorithms in data. Consequently, data-to-simulation scale factors are determined using measurements of the b-tagging performance in data, referred to as calibrations, to correct the b-tagging efficiency in simulated samples.

The calibration software was automated to reduce the person-power needs for b-tagging calibration and to improve the reproducibility [89] of the calibration results. Large collaborations in particle physics, such as CERN, commonly involve many different institutions who use their own high-performance computing clusters to develop software for analyses. While code sharing has been made easier with collaborative version control software, such as GitLab [90], these codes are often inherited from past analyses and come with undocumented assumptions and dependencies [91], making the reproducibility of results hard to achieve. Automation can address these issues by packaging computational environments and ensure that results are robust to changes by enforcing the frequent testing of new versions of the software.

9.1 Calibration of the b-tagging efficiency

The b-jet tagging efficiency is calibrated using di-leptonic top-quark ($t\bar{t}$) events [75] as the branching fraction of $t \to Wb$ is close to 1 [92], indicating that a sample of high b-jet purity can be obtained. Events with two different flavour oppositely-charged leptons ($e\mu$) are selected as the non-$t\bar{t}$ contributions in the $e\mu$ channel are small ($\sim$15%). Furthermore, exactly two reconstructed jets are required in the event selection to greatly

reduce $t\bar{t}$ events with additional light-jets.

Events are classified according to the p_T of the jets to measure the b-tagging efficiency as a function of p_T. In addition, events were further classified as a function of the b-tagging discriminant of the two jets (w_1, w_2).

A binned log-likelihood function is used to extract the conditional probabilities for a b-jet with a certain p_T to obtain a b-tagging discriminant w that falls within a certain pseudo-continuous b-tagging discriminant bin W, i.e. $\mathcal{P}_{w\in W|f=b,p_T}$. The conditional probability is related to to the b-tagging efficiency (ϵ_b) of a given p_T and working point (WP) by

$$\epsilon_b\left(p_T, \mathrm{WP}\right) = \sum_{w\in \mathrm{WP}} \mathcal{P}_{w\in W|f=b,p_T}. \tag{9.1}$$

For each working point corresponding to the target efficiencies in simulation $\epsilon_b^{MC} =$ 60%, 70%, 77%, and 85%, the b-tagging efficiency as a function of jet p_T is calibrated by scaling the efficiency of each p_T bin in simulation to that in data.

9.2 Automation techniques

Several automation techniques involving the use of container technology (e.g. Docker [93] and Singularity [94]) and automated pipelines (e.g. GitLab continuous integration and delivery (CI/CD) pipeline) were investigated to determine the best approach to achieve the automation of the calibration.

A container holds packaged, self-contained and deployable parts of the software, such as libraries and configurations [95]. Since containers share the kernel with the host machine, this allows them to be more lightweight than virtual machines [96]. Both Docker and Singularity are built on Linux Containers [94], which enables the running of a set of processes to be isolated from the rest of the system. An automated pipeline automates the execution of a chain of jobs and can be constructed using CI/CD software, such as GitLab CI/CD, or by defining workflows in containers.

9.2.1 Docker

Docker is an open source platform for developing, deploying and running applications within containers. A Docker image is made up of a series of layers. The image is built by Docker using instructions from a Dockerfile, which contains all the commands, dependencies, libraries, tools and other files needed to execute the code. A running instance of a Docker image is known as a Docker container. When a new container is created, a new writable layer is added on top of the underlying layers. All changes made to the running container are stored to this writable top layer.

However, there are security concerns surrounding Docker. To use Docker, root access is required, which can lead to privilege escalation in the running of the containers [94]. From a computing security perspective, a machine is considered to be compromised if any user is able to run arbitrary code as the root user. As a result, Docker is not supported by large collaborations, such as CERN.

9.2.2 Singularity

Singularity was developed by the Lawrence Berkeley National Laboratory as an open source initiative to provide containers and reproducibility in scientific computing, bringing a secure means to create and deploy reproducible environments across high-performance computing (HPC) clusters [94]. Similar to Docker, Singularity uses containers to encapsulate computational environments and all software dependencies needed to run a defined workflow. These containers can also run without root access and can address the security concerns of Docker. Singularity also allows Docker images to be pulled from a Docker registry, therefore enabling the use of Docker indirectly.

9.2.3 GitLab CI/CD

GitLab is a web-based platform developed on the Git version control system to help organisations manage Git repositories collaboratively. Recently, GitLab has become a platform that provides a single application for the entire software development and operations lifecycle.

One of the key features of GitLab is their CI/CD pipelines, which allows the

ATLAS software stack to adopt these two software engineering best practices [97]. While CI is a practice in which software is built and tested each time a developer pushes code to the version control repository, CD is an approach in which CI and automated deployment capabilities allow the software to be developed and deployed rapidly. Pipelines incorporating these practices provide a means of testing code and ensuring code changes are reliable, and can therefore be made frequently in a robust manner. While these pipelines can be built using Docker and Singularity, GitLab offers a simple way of constructing these pipelines through configuring a YAML [98] file and a Runner, which is an isolated virtual machine that communicates with GitLab through an application programming interface. The YAML file contains a set of instructions that informs the Runner on how to run the pipeline.

9.3 Pipeline for calibration

After investigating the various automation techniques, it was decided that GitLab CI/CD pipelines would provide the best approach to automating the calibration software. These pipelines are well-supported at CERN, and unlike Docker, there are no major security concerns. While Singularity would provide a good way of automating software with complex workflows and large files, it is not as developed at CERN and has configuration settings that will require root access. GitLab CI/CD pipelines mainly require one file to setup the automation, making it easy to manage the implementation of the automation. Due to the usability and support of this platform, it is generally the best approach to automating software in ATLAS.

To construct the GitLab CI/CD pipeline, the calibration software was broken down into the following stages:

1. **Selection:** This stage corresponded to the event selection of the b-tagging efficiency calibration measurement. The final selection was applied on lists containing the locations of the data and simulated samples and their systematic uncertainties to produce histograms for each sample.
2. **Combine histograms:** The histograms of the selection stage were merged according to the event categorisation procedure.

3. **Likelihood fit:** The fitting procedure using the binned log-likelihood function was performed on the combined histograms to calculate the b-tagging efficiencies as a function of jet p_T for the data and simulated samples, which were used to obtain the data-to-MC scale factors.
4. **Validation:** A χ^2 goodness-of-fit comparison to reference histograms was used to validate the histograms from the likelihood fit.
5. **Website:** Once the histograms were validated, a website was produced to present the results in a readable format, to provide an easily accessible record of the result.

However, the selection stage required accessing files on the Worldwide LHC Computing Grid (WLCG) [99], which could be not integrated into GitLab CI/CD pipeline at present, hence this stage was not included in the automation of the software.

9.4 Results

The pipeline was implemented successfully, as shown by the status of the GitLab CI/CD pipeline in Fig. 9.1.

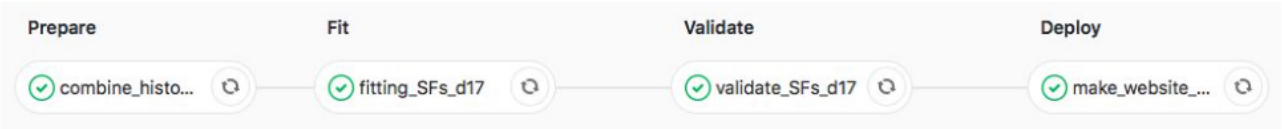


Figure 9.1: GitLab CI/CD pipeline for the calibration of b-tagging efficiency software for data recorded in 2017 (d17).

The data-to-MC scale factors of the b-tagging efficiency for the MV2c10 tagger and 70% efficiency WP, since this is used for the $VH \to bb$ analysis, are presented in Fig. 9.2. The results produced were consistent with the official results at the time [100] that the work in this chapter was undertaken.

9.5 Summary

Different automation techniques were studied to determine the best method to automate the calibration of the b-jet tagging efficiency. The GitLab CI/CD pipelines were chosen as the best approach to automating the calibration software due to reasons relating

to security, support, and usability. The automation of the calibration software was implemented successfully, and helped inform the automation of other calibration software. This will improve the robustness and reproducibility of these calibrations in the future, reducing person-power needs, facilitating rapid testing of any code changes, and allowing new people to easily carry out these calibrations.

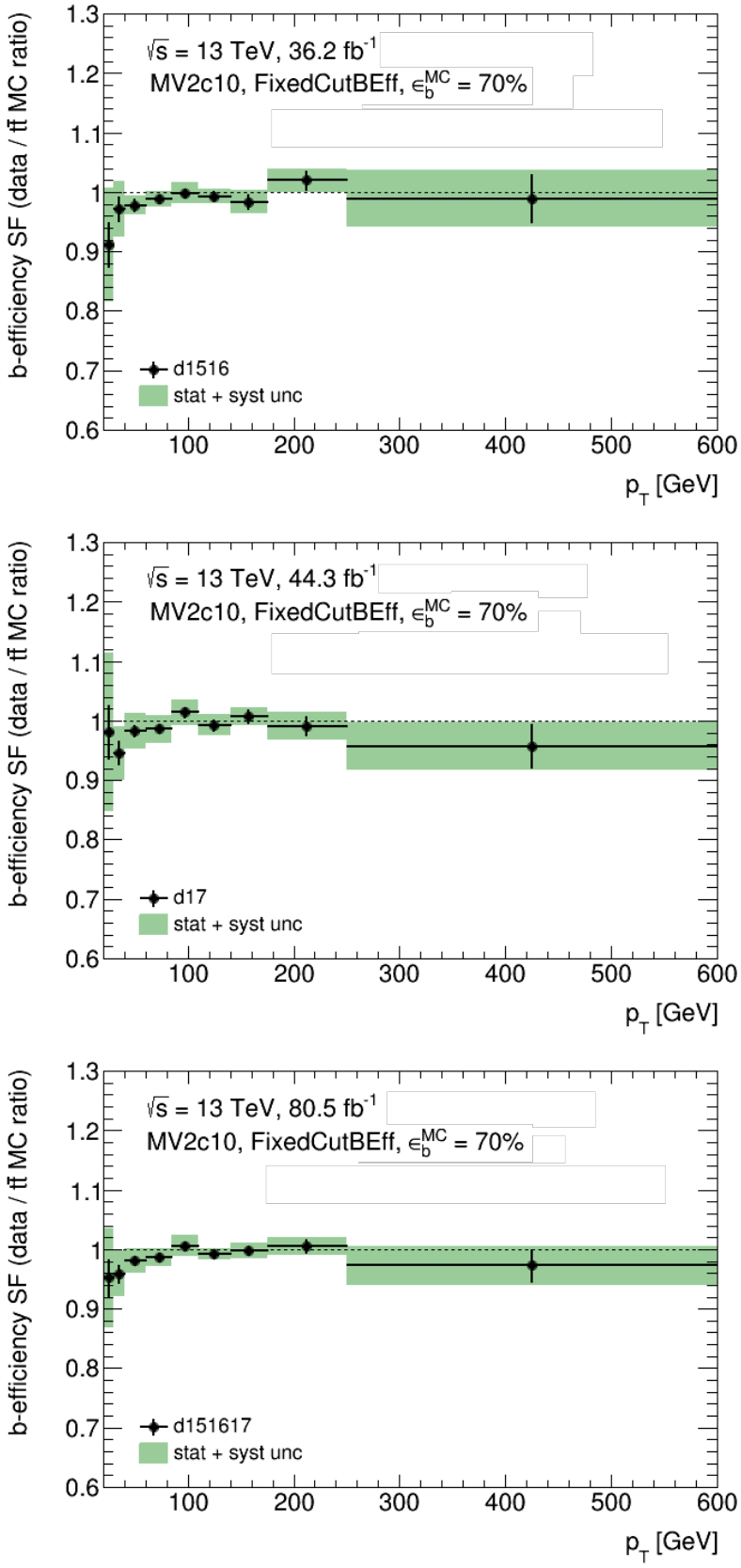


Figure 9.2: Data-to-MC scale factors for the MV2c10 b-tagging efficiency at the 70% working point for the 2015+2016 (top), 2017 (middle), and 2015+2016+2017 (bottom) datasets. The green band corresponds to the MC statistical and systematic uncertainties using summation in quadrature. The vertical error bars represent data statistical uncertainties only.

Part III

$VH, H \to b\bar{b}$ analysis

Preamble

The work in the following chapters relates to the full Run-2 $VH, H \to b\bar{b}$ analysis that is documented in Ref. [17], where some of the tables and figures in this part have originated from.

Chapter 10

Event selection and categorisation

Events are categorised into the 0-, 1- and 2-lepton channels corresponding to the $ZH \to \nu\nu b\bar{b}$, $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell\ell b\bar{b}$ signatures, respectively. The categorisation depends on the number of electrons and muons that pass the object selection detailed in this chapter. In all lepton channels, the Higgs boson candidate is formed by requiring events to have two b-tagged jets, giving rise to the 2-jet and 3-jet categories.

Monte Carlo (MC) simulations are used to model most signal and background events except for the $t\bar{t}$ background in the 2-lepton channel (Section 10.4.1), and the multi-jet background in the 1-lepton channel (Section 10.4.2).

10.1 Simulated samples

All simulated processes are normalised using the most accurate theoretical cross section predictions available and are produced using the ATLAS detector simulation [81] with GEANT4 [82] simulating the interactions of particles within the detector. The different processes simulated using MC generators are described in Section 2.2.3.

- **Signal:** All qq-initiated production processes are simulated using the Powheg [85] generator with the MinLO [101] procedure and the NNPDF3.0NLO [86] PDF set applied for the matrix element (ME) calculation. This generator is interfaced to the Pythia 8 [102] for the parton shower (PS), hadronisation, and underlying event (UE) modelling.

  The simulated samples for the gg-initiated ZH contribution are generated by Powheg with the NNPDF3.0NLO PDF set applied for the ME calculation, interfaced to Pythia 8 for the simulation of PS, hadronisation, and UE.

  Next-to-leading order (NLO) EW corrections are expected to have a substantial

impact on the p_T^V distribution, hence the VH differential cross section is computed for EW corrections at NLO as a function of p_T^V, which is applied to the simulated sample.

- **V+ jets:** The V+jets simulated samples are produced with SHERPA 2.2.1 [103] interfaced to the NNPDF3.0NNLO PDF set for both the ME calculation. The PS, hadronisation, and UE processes are also modelled by SHERPA 2.2.1.

  The V+jets events are classified based on the flavour components of the two jets used to reconstruct the Higgs boson candidate, giving the V+bb, V+bc, V+bl, V+cc, V+cl and V+ll categories. The main background contribution is V+hf, which is composed of the V+bb, V+bc, V+bl and V+cc events.

- **$t\bar{t}$:** The simulated samples for the $t\bar{t}$ processes are generated using POWHEG with the NNPDF3.0NLO PDF set applied for the ME calculation, interfaced to PYTHIA 8 for the simulation of PS, hadronisation, and UE.

- **Single-top:** The simulated samples for the single-top processes are produced independently for the t-, s-, and Wt-channels using POWHEG with the NNPDF3.0NLO PDF set applied for the ME calculation, which is interfaced to PYTHIA 8 for the PS, hadronisation, and UE modelling.

  A diagram removal (DR) [104] scheme is used to remove all diagrams in the Wt-channel that overlap with $t\bar{t}$ production.

- **Diboson:** The diboson processes are simulated using SHERPA 2.2.1 interfaced with the NNPDF3.0NNLO PDF set for the ME calculation. The PS, hadronisation, and UE are generated using SHERPA 2.2.1 as well.

10.2 Object selection

Data used in the $VH, H \to bb$ analysis were collected at a centre-of-mass energy of 13 TeV during Run-2 of the LHC, corresponding to a total integrated luminosity of 139 fb^{-1}. Events are selected only if they are of good quality and all relevant detector components were known to be operating as expected.

- **Leptons:** Lepton candidates are selected using the identification and isolation quality working points, outlined in Sections 4.2.1 and 4.2.2.

  Electron candidates are selected with the Loose identification and LooseTrack isolation working points, which are referred to as loose electrons. Tight electrons are selected with Tight identification and HighPtCalo isolation working points. Selected electrons are required to have $p_T > 7$ GeV and to be within the ID acceptance ($|\eta| < 2.47$).

  Similarly, loose muons are selected with Loose identification and LooseTrack isolation working points. For tight muons, Medium identification and HighPtTrack isolation working points. Selected muons are required to have $p_T > 7$ GeV and to be within the MS acceptance ($|\eta| < 2.7$).

- **Jets:** Jets are reconstructed using the anti-k_T algorithm with $R = 0.4$ (Section 4.3). Jet candidates with $p_T < 120$ GeV and $|\eta| < 2.5$ that are not associated with the PV of the hard interaction are required to have a JVT discriminant > 0.59 to suppress pile-up jets. Additional requirements on jets to maximise the signal sensitivity are $p_T > 20$ GeV in the central region ($|\eta| < 2.5$) and $p_T > 30$ GeV in the forward region ($2.5 < |\eta| < 4.5$).

- ***b*-tagging:** The 70% working point on the MV2c10 discriminant (Section 6.2.1) is used to b-tag exactly two jets in the central region ($|\eta| < 2.5$) to maximise the expected signal significance. At least one b-tagged jet is required to have $p_T > 45$ GeV.

  Due to the b-tagging requirement, a substantial amount of selected events are rejected, which significantly increases the statistical uncertainty of the simulated datasets. Hence, a probability weighting method (truth-tagging) is used, in which events are weighted according to the b-tagging efficiency of all jets in the event.

10.3 Event categorisation

Events without and with additional non-b-tagged jets are split into the 2-jet and 3-jet categories, respectively. In the 0-lepton (0L) and 1-lepton (1L) channels, only

one additional untagged jet is allowed. In the 2-lepton (2L) channel, any number of additional untagged jets is allowed as this increases the expected signal significance. The full event selection for the three lepton channels is summarised in Table 10.1.

The transverse momentum of the reconstructed V-boson (p_T^V) corresponds to E_T^{miss} in 0L, the vectorial sum of E_T^{miss} and the p_T of the charged lepton in 1L, and the transverse momenta of the two lepton system in 2L. The analysis focusses on the two high-p_T^V regions defined by 150 GeV $< p_T^V <$ 250 GeV and $p_T^V >$ 250 GeV, as the signal-to-background ratio increases for large p_T^V values. In the 2L channel, there is an additional medium-p_T^V region defined by 75 GeV $< p_T^V <$ 150 GeV.

Single lepton triggers are used for the electron sub-channel in 1L and for both the electron and muon sub-channels in 2L. In the electron sub-channels, the single electron trigger threshold increased from 24 GeV in 2015 to 26 GeV in 2016-2018. In the muon sub-channel, the single muon trigger threshold increased from 20 GeV in 2015 to 26 GeV in 2016-2018.

For the 0L and the 1L muon sub-channel, the same E_T^{miss} trigger is used, whose threshold varied from 70 to 110 GeV during the 2015-2018 data-taking period. The E_T^{miss} trigger is used in the 1L muon sub-channel because the E_T^{miss} trigger works more efficiently than single muon triggers in this analysis phase space, due to the limited coverage of the muon trigger system in $|\eta| < 2.5$.

Events with loose leptons are rejected in 0L. Additional requirements in 0L are imposed on the scalar sum of the transverse momenta of jets (H_T), which removes a small part of the phase space where the trigger efficiency depends slightly on the number of jets in the event. Cuts on the azimuthal angular separation of E_T^{miss}, jets and the missing transverse momentum that only considers reconstructed tracks in the ID and matched to the PV (p_T^{miss}) are applied to reject multi-jet events with a large E_T^{miss}, which originate from mismeasured jets that are aligned with the fake E_T^{miss}.

In 1L, events are required to have one tight electron with $p_T >$ 27 GeV, one tight muon with $p_T >$ 25 GeV, and no loose leptons. An additional selection of $E_T^{miss} >$ 30 GeV is applied in the electron sub-channel to reduce the background from multi-jet production.

In 2L, events are required to have two same-flavour loose leptons, where one of the

leptons must have $p_T > 27$ GeV and $|\eta| < 2.4$. The requirement of the invariant mass of the lepton pair (m_{ll}) to be compatible with that of the Z-boson further suppresses the multi-jet and $t\bar{t}$ backgrounds in this channel. As the charge misidentification rate is negligible in the electron sub-channel, ee events are not required to have opposite-sign charges.

10.3.1 Signal and control regions

Events are further categorised into signal regions and control regions. During the Run-2 analysis, the signal and control regions were redefined by selecting on the ΔR between the two b-tagged jets ($\Delta R(b_1, b_2)$) as a function of p_T^V. Lower and upper selections on $\Delta R(b_1, b_2)$ are applied to give two control regions (CRs), known as the low and high CRs. The high CR is defined from the 1L signal samples using cuts at 95% and 85% of the signal for the 2-jet and 3-jet events, respectively, to create a region enriched in $t\bar{t}$ and single top events in the 1L channel and in Z+hf events in the 2L channel. The low CR is defined from the 1L diboson samples, as the same regions are used for the diboson cross-check analysis, using cuts at 10% of the diboson background for both 2-jet and 3-jet events, resulting in a region enriched in W+hf events in the 1L channel and in Z+hf in the 2L channel.

10.4 Data-driven background estimations

10.4.1 $t\bar{t}$ in the 2-lepton channel

The dominant contribution to the $t\bar{t}$ background in this channel is the di-leptonic decay channel, which is separated into the same-flavour ($ee/\mu\mu$) and opposite-flavour ($e\mu$) categories. While the same-flavour region has a large Z+jets background, the $e\mu$ region has a very high purity of top background (99.5-99.9% depending on the jet categories) and therefore provides a very pure top CR, referred to as the $e\mu$CR.

By deriving a scale factor between the number of events ($\mathcal{N}$) in the signal region (SR) and the $e\mu$CR (CR) using simulation predictions (MC), distributions of a variable

Selection	0-lepton	1-lepton		2-lepton
		e sub-channel	μ sub-channel	
Trigger	E_T^{miss}	Single lepton	E_T^{miss}	Single lepton
Leptons	0 loose leptons	1 tight electron	1 tight muon	2 same-flavour loose leptons $\mu\mu$ with opposite-sign charges
	$p_T > 7$ GeV	$p_T > 27$ GeV	$p_T > 25$ GeV	$p_T > 27$ GeV
E_T^{miss}	> 150 GeV	> 30 GeV	-	-
m_{ll}	-	-		81 GeV $< m_{ll} <$ 101 GeV
Jets	2 or 3 jets			2 or ≥ 3 jets
	$p_T > 20$ GeV for $\lvert\eta\rvert < 2.5$			
	$p_T > 30$ GeV for $2.5 < \lvert\eta\rvert < 4.5$			
b-jets	Exactly 2 b-tagged jets			
	$p_T > 45$ GeV for leading b-jet			
H_T	> 120 GeV for 2 jets, > 150 GeV for 3 jets	-		-
min[$\Delta\phi(E_T^{miss}$, jets)]	$> 20°$ for 2 jets, $30°$ for 3 jets	-		-
$\Delta\phi(E_T^{miss}, bb)$	$> 120°$	-		-
$\Delta\phi(b_1, b_2)$	$< 140°$	-		-
$\Delta\phi(E_T^{miss}, p_T^{miss})$	$< 90°$	-		-
	-	-		75 GeV $< p_T^V <$ 150 GeV
	150 GeV $< p_T^V <$ 250 GeV			
	$p_T^V > 250$ GeV			
Signal regions	$\Delta R(b_1, b_2)$ signal selection			
Control regions	High and low selections on $\Delta R(b_1, b_2)$			

Table 10.1: Event selection and categorisation in the 0-, 1-, and 2-lepton channels. The angular cuts between two objects are denoted by $\Delta\phi$, where ϕ is the azimuthal angle.

from data in the $e\mu$CR can be directly used to model the top background within the SR:

$$\frac{\mathcal{N}(\mathrm{SR}_{top}^{data})}{\mathcal{N}(\mathrm{CR}_{top}^{data})} = \frac{\mathcal{N}(\mathrm{SR}_{top}^{MC})}{\mathcal{N}(\mathrm{CR}_{top}^{MC})} \Rightarrow \mathcal{N}(\mathrm{SR}_{top}^{data}) = \frac{\mathcal{N}(\mathrm{SR}_{top}^{MC})}{\mathcal{N}(\mathrm{CR}_{top}^{MC})} \times \mathcal{N}(\mathrm{CR}_{top}^{data}). \tag{10.1}$$

This derivation is possible because the same kinematic selections are applied to both the SRs and $e\mu$CRs. In addition, due to the flavour-symmetric nature of the $t\bar{t}$ background and the very pure nature of the CR, there is very little difference between the two regions. The scale factor, $\frac{\mathcal{N}(\mathrm{SR}_{top}^{MC})}{\mathcal{N}(\mathrm{CR}_{top}^{MC})}$, is calculated to be 1.00 ± 0.01 and 1.01 ± 0.01 for the 2-jet and 3-jet categories [17], respectively, consistent with unity within the MC statistical uncertainty.

10.4.2 Multi-jet in the 1-lepton channel

Multi-jet (MJ) events can pass the event selection due to the reconstruction of fake isolated leptons. As MJ events are produced with a large cross section, it is likely that these events could contribute a sizeable amount to the background, even if the fake isolated lepton rate is very low. The MJ enriched control region (CR) is defined by inverting the lepton isolation requirement. Due to the limited number of events in the CR, only one b-tagged jet is required instead of two to increase the statistical precision of the background precision. The MJ template is derived from data in the CR after the subtraction of all simulated electroweak (EW) backgrounds.

A fit to the m_T^W distribution in data is performed to simultaneously extract the normalisation factors for the MJ and EW components. The m_T^W variable is chosen because it offers the best discrimination between the MJ and EW processes. After the normalisation factors are obtained, they are used in the CR to extract the shape template of a variable, which is used as the MJ background of the variable in the signal region.

10.5 Simplified Template Cross Sections regions

The main aim of the Simplified Template Cross Sections (STXS) framework [105] is to maximise the sensitivity of Higgs coupling measurements while minimising their theory

dependence. In the STXS framework, the cross sections (instead of signal strengths) are measured for each Higgs boson production mode in mutually exclusive regions of phase space, referred to as bins. The kinematic template for each bin is estimated from the particles produced in the hard scattering event in simulation, known as truth information.

For the VH production mode, the STXS regions are defined by splitting the VH production into the WH and ZH processes. Each production process is split into p_T^V bins. As the STXS regions are split using truth information, these regions are denoted in terms of $p_T^{V,t}$ in the summary of the regions in Table 10.2.

STXS region		Corresponding reconstructed analysis regions		
Process	$p_T^{V,t}$ 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, $\geq$3
ZH	>250 GeV	0	>250 GeV	2, 3
		2	>250 GeV	2, $\geq$3

Table 10.2: The simplified template cross section regions and their corresponding reconstructed analysis regions [17].

Chapter 11

Multivariate analysis

A multivariate analysis (MVA) technique was trained on simulated Monte Carlo (MC) samples to improve the separation between the signal and background processes, to improve the sensitivity of the analysis. Individual boosted decision trees (BDTs), as detailed in Section 5.1, were trained for each jet category within each lepton channel. Two versions of these BDTs with the same input variables and parameters were used. One is BDT_{VH}, which is used in the nominal analysis to extract the $VH, H \to b\bar{b}$ signal from all background processes. The other is BDT_{VZ}, which is designed to validate the analysis by separating the diboson $VZ, Z \to b\bar{b}$ process from other background processes and the VH signal. The BDTs were implemented in TMVA [68], which is a toolkit integrated in ROOT and contains a large variety of multivariate classification algorithms.

Variables describing the kinematics and topology of the events were used as inputs to the BDTs. The baseline selection of variables, which were all verified to be well-modelled, was optimised in Ref. [13], and is summarised in Table 11.1 for the different lepton channels. The variables were chosen based on their discriminating power between signal and background, with the most discriminating variables being the di-jet mass (m_{bb}), the angular separation between the two tagged b-jets ($\Delta R(b,b)$) and the transverse momentum of the reconstructed vector boson (p_T^V).

A 2-fold cross-validation was used to avoid bias in training and to make full use of the MC events for BDT training and in the analysis, given that the statistical uncertainty on the number of simulated events is one of the dominant uncertainties. The events were split into two subsets, depending on whether their event number was odd or even, to ensure that they were orthogonal to each other. The BDTs were trained on events with odd event numbers and tested on events with even event numbers, and vice versa. A good agreement of the ROC curves of the BDT outputs between training and test

data indicated no overfitting was present in the training.

Variable	Label	0-lepton	1-lepton	2-lepton
p_T^V	pTV	$\equiv E_T^{\text{miss}}$	×	×
E_T^{miss}	MET	×	×	×
$p_T^{b_1}$	pTB1	×	×	×
$p_T^{b_2}$	pTB2	×	×	×
m_{bb}	mBB	×	×	×
$\Delta R(b_1, b_2)$	dRBB	×	×	×
$\lvert\Delta\eta(b_1, b_2)\rvert$	dEtaBB	×		
$\Delta\phi(V, bb)$	dPhiVBB	×	×	×
$\lvert\Delta\eta(V, bb)\rvert$	dEtaVBB			×
m_{eff}	HT	×		
$\min(\phi(l, b))$	dPhiLBmin		×	
m_T^W	mTW		×	×
m_{ll}	mLL			×
m_{top}	Mtop		×	
$\lvert\Delta Y(V, bb)\rvert$	dYWH	×		
3-jet events only				
$p_T^{\text{jet}_3}$	PTJ3	×	×	×
m_{bbj}	mBBJ	×	×	×

Table 11.1: Input variables used for the multivariate analysis in the different lepton channels. The missing transverse energy is E_T^{miss}. The invariant mass of the dijet system is m_{bb}. The azimuthal angle, pseudorapidity and rapidity separations between the vector boson and the dijet system are $\Delta\phi(V, bb)$, $|\Delta\eta(V, bb)|$, and $|\Delta Y(V, bb)|$, respectively. The leading and sub-leading b-tagged jets are denoted by b_1 and b_2. The distance in η and ϕ between the 2 b-jets is $\Delta R(b_1, b_2)$. The angle between the lepton and the closest b-tagged jet in the transverse plane is denoted by $\min(\phi(l, b))$. The scalar sum of the transverse momenta of E_T^{miss} and all jets is m_{eff}. The transverse mass of the reconstructed W boson is m_T^W. The transverse momentum of an additional non-b-tagged jet is $p_T^{\text{jet}_3}$.

Distributions of the BDT outputs were transformed using a rebinning procedure [32] to maximise the sensitivity of the analysis, while minimising the number of bins to reduce statistical uncertainty on the simulated samples to a reasonable level. This procedure also allowed for better comparisons of the different BDT trainings. The rebinning procedure conducted was based on a transformation quantity:

$$Z = z_S \frac{n_S}{N_S} + z_B \frac{n_B}{N_B}, \tag{11.1}$$

where n_S and n_B are the number of signal and background events in a given bin interval, N_S and N_B are the total number of signal and background events, z_S and z_B are the

signal and background transformation parameters. The merging is conducted on an initial distribution with a large number of bins and starts from high to low BDT-output values, where Z is calculated for each bin that is added. The merging continues to form a new bin until $Z > 1$. Z is then reset to 0 and the procedure restarts with the remaining unmerged bins. Optimal values for the parameters were $z_S = 10$ and $z_B = 5$ for BDT_{VH}, and $z_S = 5$ and $z_B = 5$ for BDT_{VZ} due to lower MC statistics in diboson. These values also ensured that the expected statistical uncertainty on the simulated samples in each merged bin is smaller than 20%.

To evaluate the performance of a BDT training, the statistical sensitivity was calculated using the expected binned significance of the BDT output distribution:

$$S = \sqrt{\sum_{i=0}^{N} 2\left((n_S^i + n_B^i)\ln\left(1 + \frac{n_S^i}{n_B^i} - n_S^i\right)\right)}, \tag{11.2}$$

where N is the total number of bins in the distribution, and n_S^i and n_B^i are the number of signal and background events in bin i, respectively. The overall performance of each training was assessed by the quadrature sum of the transformed BDT statistical sensitivity for the 2-jet and 3-jet categories.

The studies in this chapter were performed in the 1-lepton channel, but where relevant were also applied to the other lepton channels.

11.1 $VH, H \to b\bar{b}$ high p_T^V region

The optimised configuration of the BDT_{VH} hyperparameters for the Run-1 analysis is summarised in Table 11.2. This was also used to train the baseline MVA for the Run-2 analysis [15] in the high p_T^V region ($p_T^V > 150$ GeV).

11.1.1 Hyperparameter optimisation and boosting algorithms

Due to an updated event selection and increased training statistics in the MC samples for the Run-2 analysis, a scan of several BDT hyperparameters and optimisation algorithms was performed to determine if these values were still optimal.

The boosting algorithm used for the Run-1 results and thus used to train the baseline

TMVA Option	Definition	Value
NTrees	Number of trees in forest	200
MaxDepth	Maximum depth of the trees	4
MinNodeSize	Minimum percentage of training events required in a node	5%
nCuts	Number of evenly spaced cuts per variable per node	100
BoostType	Boosting algorithm for trees in forest	AdaBoost
AdaBoostBeta	Learning rate for AdaBoost algorithm	0.15
SeparationType	Separation criterion for node splitting	GiniIndex
PruneMethod	Method for removing statistically insignificant branches	NoPruning

Table 11.2: Configuration of the BDT_{VH} hyperparameters used to train the optimised MVA for the Run-1 analysis and the baseline MVA for the Run-2 analysis.

MVA was adaptive boosting (AdaBoost). An alternative boosting algorithm available in TMVA is gradient boosting (GradBoost), as described in Section 5.1.1.2, and this was tested as an alternative. To ensure a fair comparison, the hyperparameters for both boosting algorithms were optimised. This motivated a 2D scan of the number of trees (NTrees) and the learning rates for AdaBoost and GradBoost.

The results are displayed in Fig. 11.1, where it can be observed that for both algorithms the performance generally increases with the number of trees and the learning rate until an optimal learning rate and number of trees is reached, which for both algorithms is around a learning rate of 0.5 to 0.55 and 600 to 800 trees, where the performance for both algorithms appeared to have reached a plateau. In addition, GradBoost generally performed better than AdaBoost for the different combinations of NTrees and learning rates tested, with the best GradBoost configuration (NTrees = 600, learning rate = 0.50) offering an improvement of 4.1% over the baseline configuration compared to 3.3% offered by the best AdaBoost configuration (NTrees = 800, learning rate = 0.55). Consequently, GradBoost was chosen as the optimal boosting algorithm and AdaBoost was not investigated further.

Since GradBoost with a learning rate of 0.50 showed a promising improvement in the sensitivity, further studies involving a 2D scan of the number of trees and the maximum depth of the trees (MaxDepth) was conducted using this configuration. The statistical sensitivity for each configuration is presented in Fig. 11.2 where it can be seen that the sensitivity increases with the MaxDepth value and plateaus at around 600 to 800 trees for the higher MaxDepth values of 4 and 5. As the improvement in the sensitivity was negligible when increasing NTrees from 600 to 800 and MaxDepth

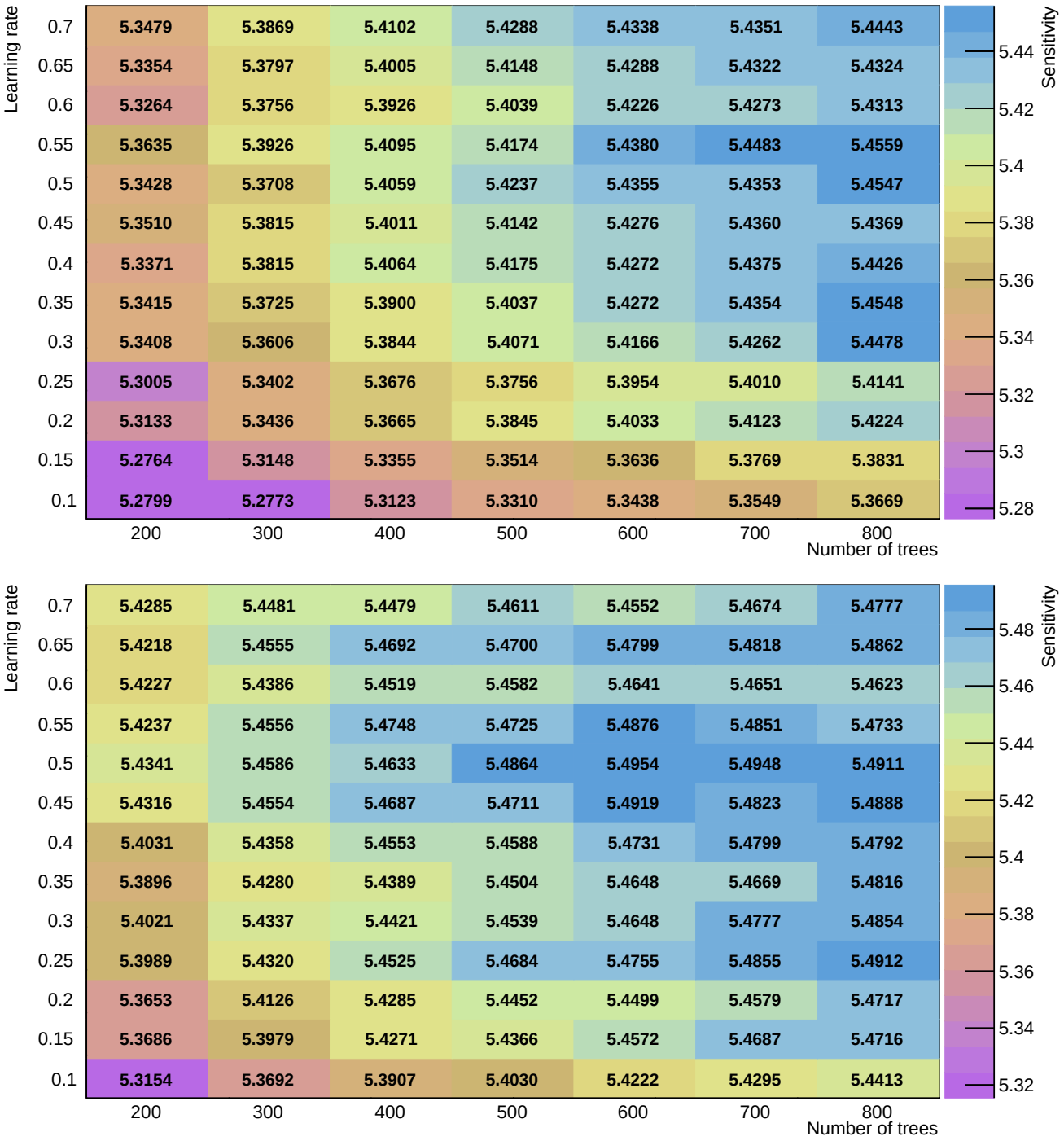


Figure 11.1: Combined BDT_{VH} statistical sensitivities for AdaBoost (top) and GradBoost (bottom) in the high p_T^V region using the parameters of Table 11.2 except for different combinations of learning rates and number of trees (and boosting algorithm).

from 4 to 5, and that bigger trees are more computationally expensive and more prone to overfitting, it was decided that NTrees = 600 and MaxDepth = 4 was the optimal number of trees and MaxDepth combination for the learning rate of 0.50. The ROC curves in Fig. 11.3 revealed no signs of overtraining in both jet categories for the optimal

GradBoost configuration.

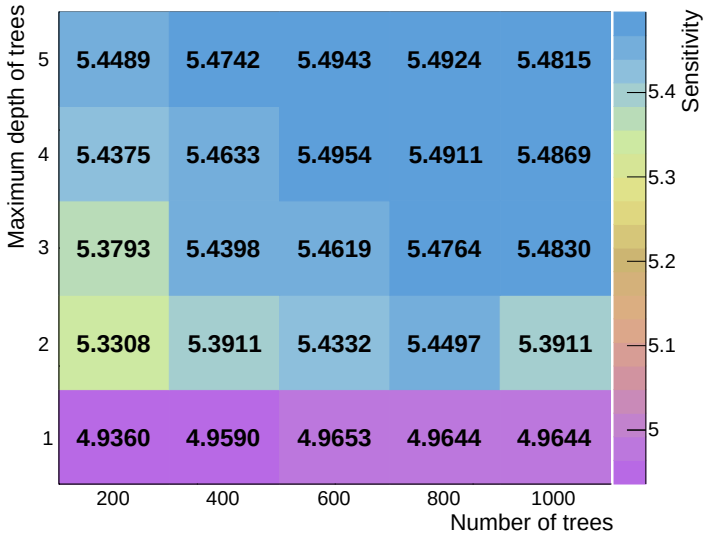


Figure 11.2: Combined BDT_{VH} statistical sensitivities for GradBoost with a learning rate of 0.50 with different combinations of number of trees and maximum tree depths in the high p_T^V region.

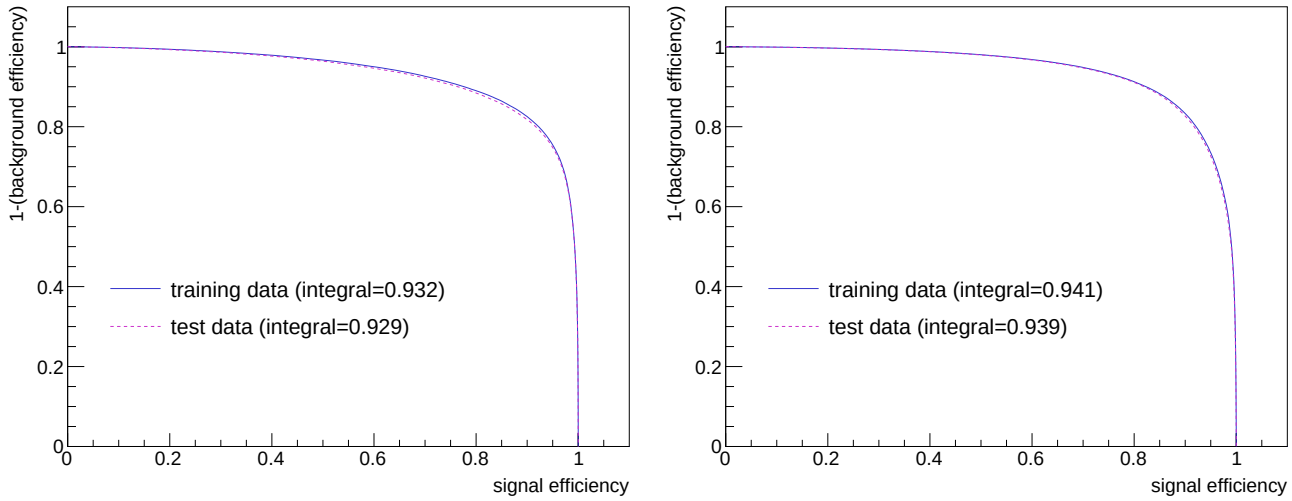


Figure 11.3: ROC curves as overtraining checks for the optimal BDT_{VH} GradBoost configuration for the 2-jet (left) and 3-jet (right) categories in the high p_T^V region.

To better understand the differences in the performance of the algorithms, the structure of the trees from the different trainings were analysed. Fig. 11.4 demonstrates how often each input variable is used at each depth for each algorithm and at learning rates of 0.10 and 0.50. It can be seen that GradBoost uses the input variables more equally, whereas AdaBoost mainly focuses on m_{bb}, $\Delta R(b,b)$ and the reconstructed top mass (m_{top}) to make decisions. In addition, increasing the learning rate generally

resulted in a small increase in the more equal use of the input variables. Thus, it can be concluded that the improvement in the performance was likely due to a better use of the available information.

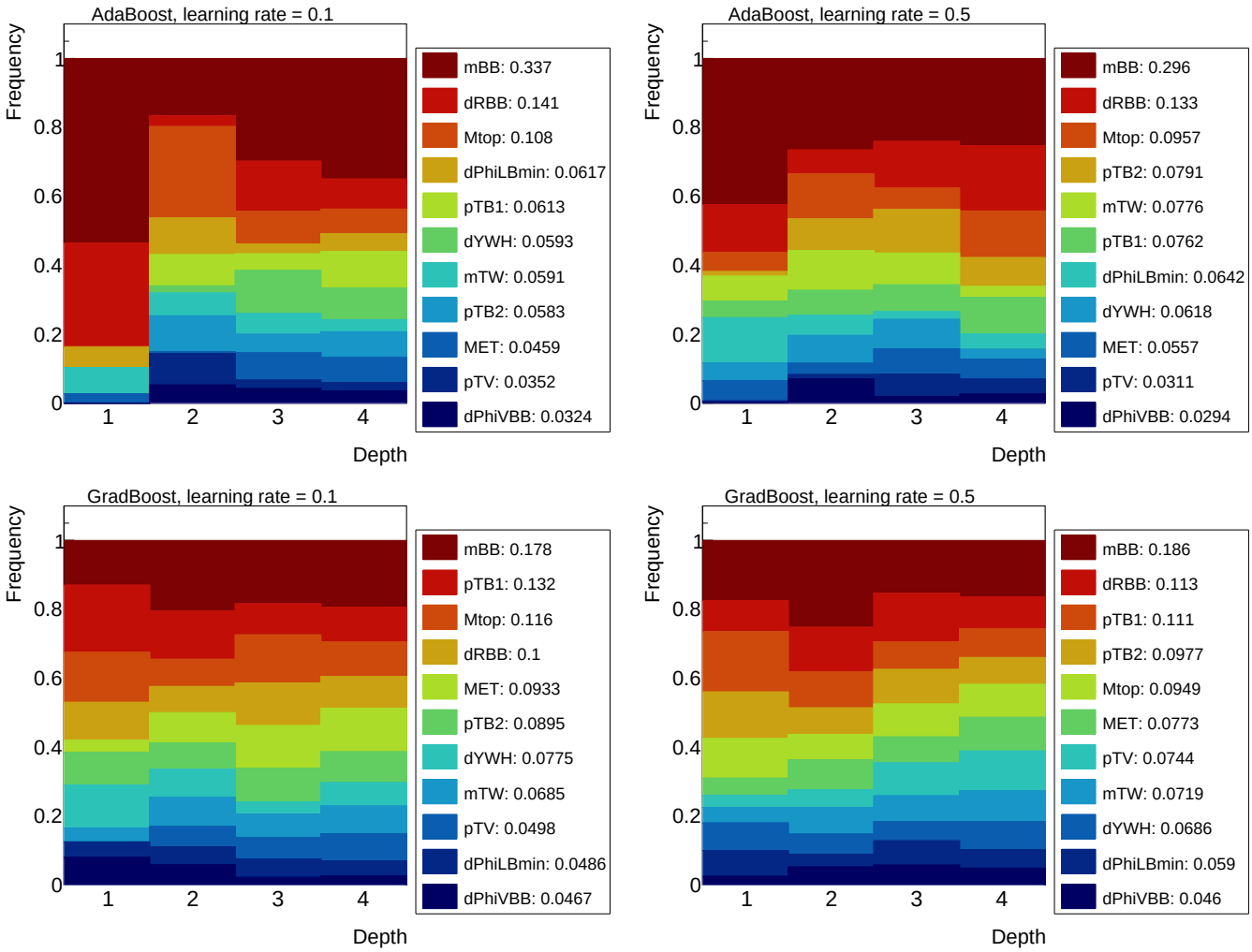


Figure 11.4: Frequencies of the usage of each input variable as a function of the tree depth for the 2 jet category for different boosting algorithms and learning rates. The values in the legend represent the overall proportional usage of each input variable in the training. GradBoost uses the input variables more equally than AdaBoost, which mainly uses m_{bb}, $\Delta R(b,b)$, and m_{top} to make the decisions. A more equal use of the input variables at higher tree depth is observed when increasing the learning rate from 0.10 to 0.50.

The effect of increasing the number of trees on how the BDT uses the input variables was also investigated. Fig. 11.5 compares the tree structures from increasing the number of trees from the default value of 200 to 600 for the baseline configuration. While a more equal use of the input variables at higher tree depth is also observed with increasing the number of trees, the effect is not as prominent compared to increasing the learning

rate. A further examination into the tree structure from combining the change in the boosting algorithm and the increases in the learning rate and the number of trees was conducted. It can be seen from Fig. 11.6 that combining all the changes resulted in a more equal use of the input variables across all tree depths, suggesting that the increase in the sensitivity from the optimal GradBoost configuration could be linked to the BDT using input variables more equally in every tree depth. However, drawing conclusions from the frequency alone does not give the full perspective of how the BDT works, since it is possible that a variable can be used frequently but not impact the performance significantly.

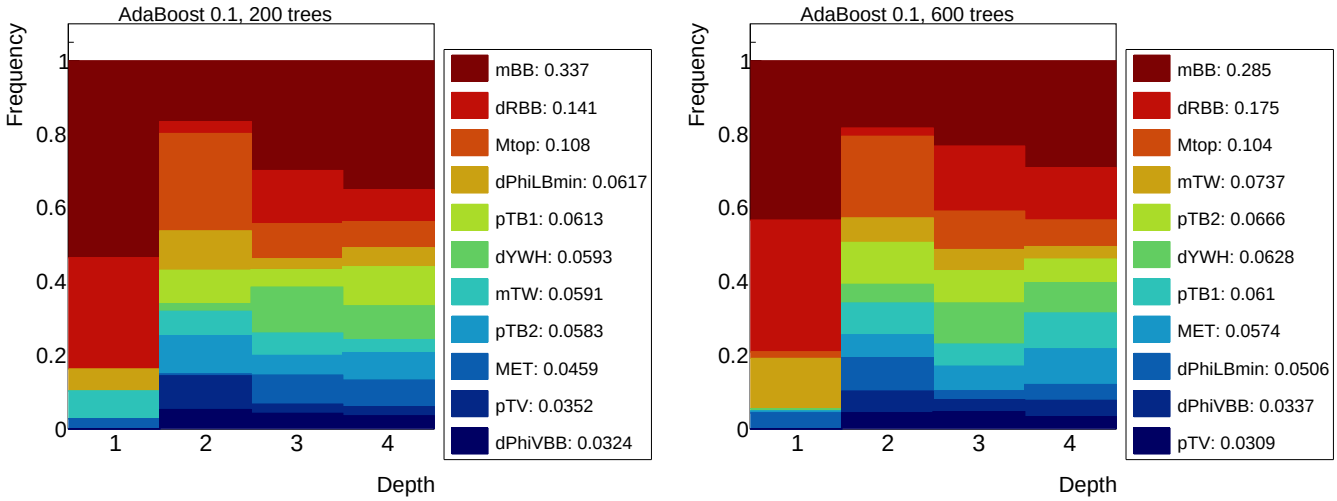


Figure 11.5: Frequencies of the usage of each input variable as a function of the tree depth for the 2 jet category for different number of trees. The values in the legend represent the overall proportional usage of each input variable in the training. A more equal use of the input variables at higher tree depth is observed with increasing the number of trees from 200 to 600.

Before the optimal GradBoost configuration was adapted into the analysis framework, a check was performed to ensure that GradBoost uses a similar phase space selection as AdaBoost to ascertain if this would result in significant differences in the events being probed. To study this, the region of the BDT which contained the top 20% of signal events was used to study the corresponding signal and background events for the most important input variables, m_{bb}, $\Delta R(b,b)$, p_T^V, and m_{top}. Comparison between the signal and background distributions of these input variables at high BDT for the two MVAs in Fig. 11.7 indicates that they are very similar, confirming that GradBoost uses a similar phase space selection as AdaBoost, and therefore can be used in the analysis without potential significant changes in the modelling and associated uncertainties.

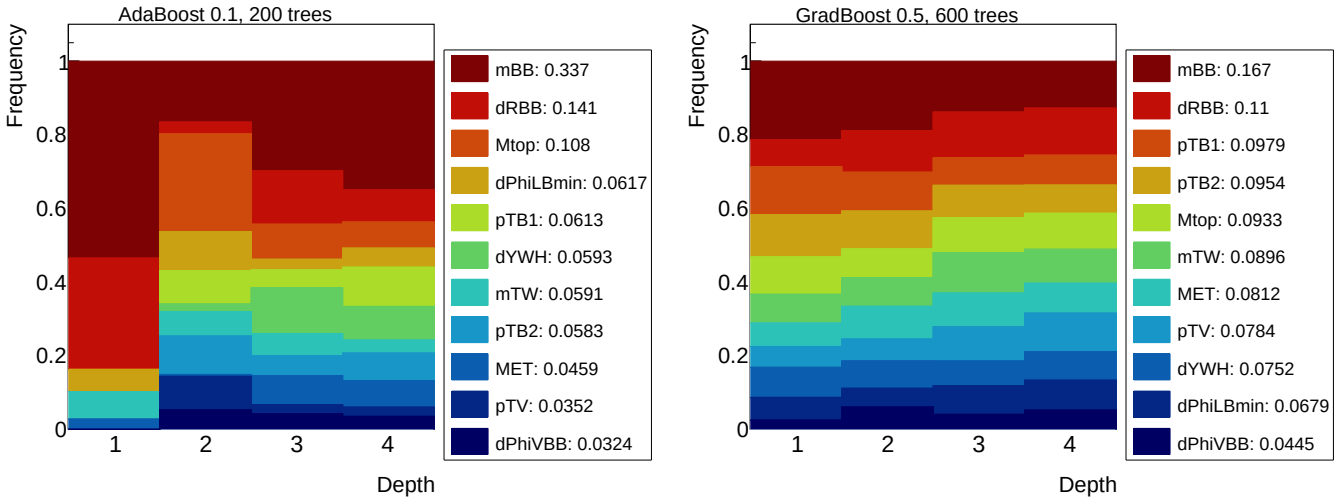


Figure 11.6: Frequencies of the usage of each input variable as a function of the tree depth for the 2 jet category for the baseline and for the optimal GradBoost configuration. The values in the legend represent the overall proportional usage of each input variable in the training. A more equal use of the input variables across all tree depths is observed with the optimal GradBoost configuration.

A summary of the optimised BDT_{VH} hyperparameters in the high p_T^V region of the 1-lepton channel for the Run-2 analysis is shown in Table 11.3. The optimised configuration yielded an improvement of 4.1% in the sensitivity over the baseline configuration.

TMVA Option	Value	
	Baseline MVA	Optimised MVA
NTrees	200	600
MaxDepth	4	4
MinNodeSize	5%	5%
nCuts	100	100
BoostType	AdaBoost	GradBoost
AdaBoostBeta/Shrinkage	0.10	0.50
SeparationType	GiniIndex	GiniIndex
PruneMethod	NoPruning	NoPruning
Performance		
Sensitivity	5.28	5.50
Improvement	-	4.1%

Table 11.3: Comparison of the optimised BDT_{VH} hyperparameters in the high p_T^V region of the 1-lepton channel for the Run-1 (Baseline MVA) and Run-2 (Optimised MVA) analysis. Shrinkage is the TMVA option for the learning rate of the GradBoost algorithm.

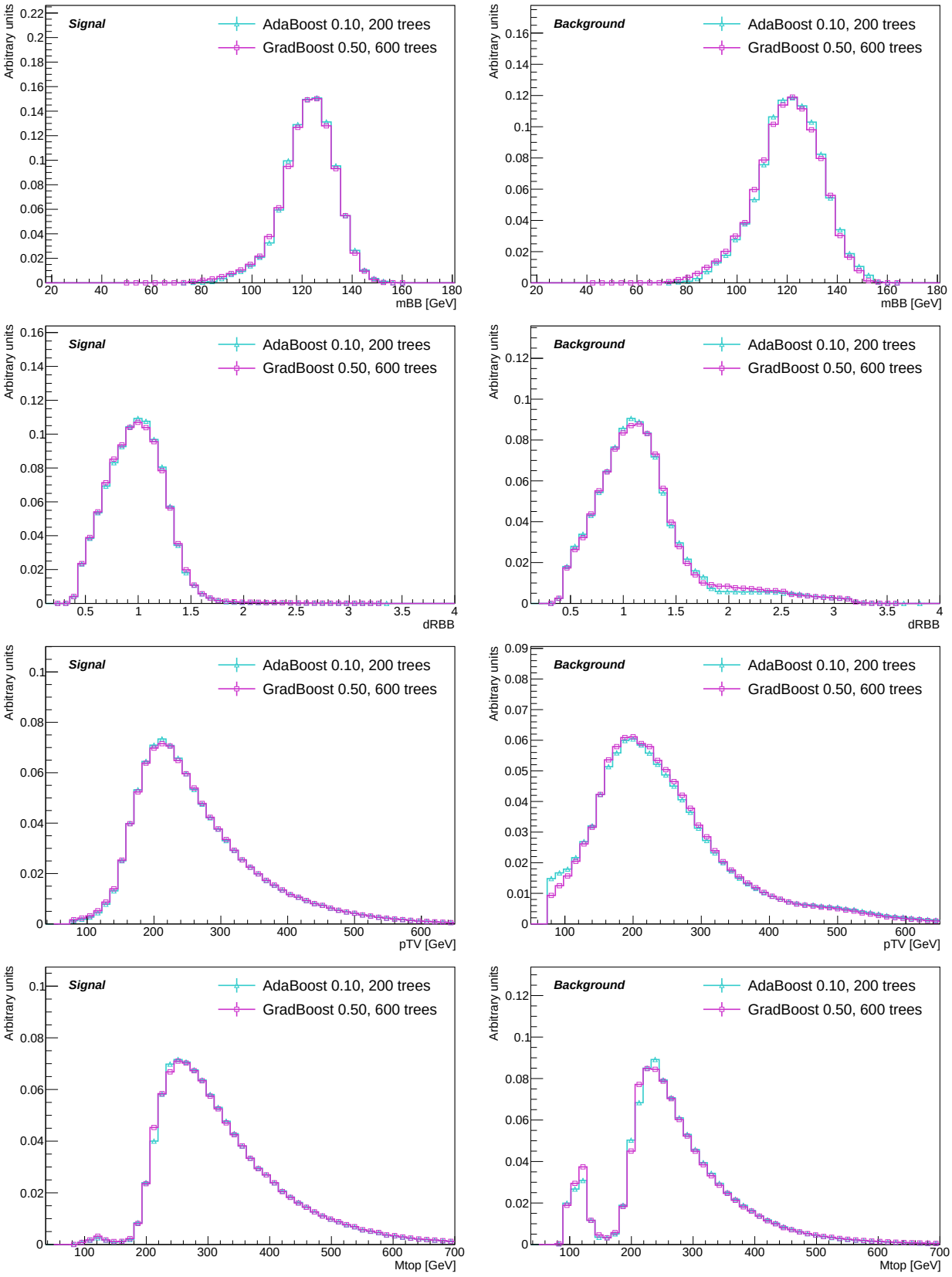


Figure 11.7: Signal and background distributions of key input variables at high BDT scores for the baseline AdaBoost and the optimised GradBoost configurations in the combined 2-jet and 3-jet categories. From top to bottom: m_{bb}, $\Delta R(b,b)$, p_T^V, and m_{top}. The distributions are very similar between the algorithms, suggesting that they use a similar phase space.

11.1.2 Range of input variables

Generally, each input variable distribution extends over a large range of values and can expand over several orders of magnitude, hence a default range that encompasses 99% of all signal events is used to find a threshold for each variable, above which all events are artificially set to the threshold value. This avoids wasting degrees of freedom in the BDT to categorise the small number of events in the tails of the distributions. Although it is not known whether this input variables range restriction can distort the correlations between the variables or lessen the performance of the BDT, therefore a training with no restriction was performed.

Removing the restriction revealed that some variables have ranges that can extend up to 1 million GeV. The correlation matrices in Fig. 11.8 demonstrate that the correlation between the input variables decreases for some variables in the no restriction case, which can be caused by a significant increase of statistical fluctuations originating from the tails of the distributions. The sensitivity from training with no restrictions in Table 11.4 signaled a deterioration in the performance of the BDT which was to be expected given the presence of some events at extreme values. As such, the default range of input variables was kept in the MVA training.

Input variables range	Sensitivity	Improvement
99% signal	5.28	-
No restriction	5.17	-2.1%

Table 11.4: Combined BDT_{VH} statistical sensitivities for trainings using different input variables range restriction.

11.1.3 Polarisation and other related variables

In the 2-lepton channel, it was discovered that the addition of variables relating to the polarisation of the Z-boson resulted in an improvement of up to 9.0% in the BDT_{VH} sensitivity [17, 106]. Consequently, the addition of variables relating to the W-boson polarisation was studied to see if a similar improvement can be achieved in the 1-lepton channel.

The 2-lepton polarisation variable used in the $VH, H \to b\bar{b}$ analysis was cosThetaLep,

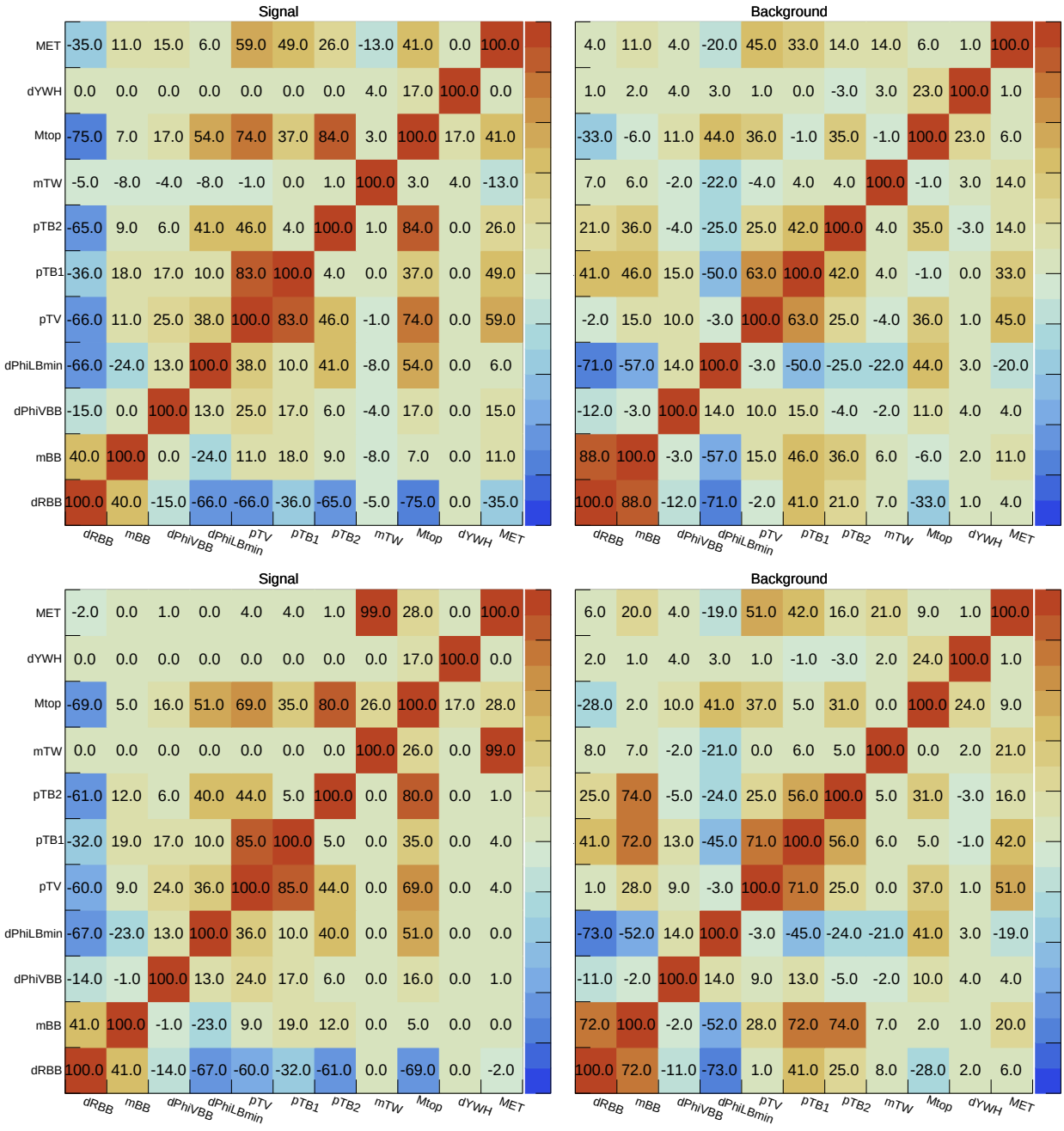


Figure 11.8: Matrices of the correlation coefficients between the input variables for the default input variable range restriction (top) and no restriction (bottom).

which is the cosine of the angle between the vector boson and negatively charged lepton in the rest frame. Since the transverse momenta of the leptons, $p_{T\ell 1}$ and $p_{T\ell 2}$, and the differences in the $p_{T\ell 1(2)}$ distributions are affected by the differences in the Z-polarisation, variables related to the Z-polarisation were added. The polarisation-related variables used in the 2-lepton channel were:

- pTL1: leading lepton transverse momentum,

- pTL1: sub-leading lepton transverse momentum,
- pTL1minusPTL2: difference between the lepton transverse momenta,
- pTL1minusPTL2divPTV: ratio of the pTL1minusPTL2 and the transverse momentum of the Z-boson i.e. $(p_{T\ell 1} - p_{T\ell 2})/p_T^V$.

Similar variables are available in 1-lepton by replacing a lepton with E_T^{miss} (MET), which represents the neutrino in $W \to \ell\nu$. As cosThetaLep in the 1-lepton channel was not computed in time for the analysis deadline, only variables related to the W-polarisation were studied:

- pTL: lepton transverse momentum,
- METminusPTL: difference between the lepton transverse momentum and MET,
- METminusPTLdivPTV: ratio of the METminusPTL and the transverse momentum of the W-boson i.e. $(E_T^{miss} - p_T^L)/p_T^V$.

The signal and background distributions for the three polarisation variables are shown in Fig. 11.9, where it can be observed that some degree of signal and background discrimination is provided by all three polarisation variables. The sensitivities from the trainings involving each of these variables are presented in Table 11.5. A training including all the polarisation variables was also conducted to determine if this could improve the performance of the BDT from exploiting correlations between them. Overall, the inclusion of the polarisation variables delivered negligible improvements of less than 1%. The studies were later repeated with GradBoost, which found the same results.

Polarisation variable added	Sensitivity	Improvement
None	5.28	-
pTL	5.30	0.38%
METminusPTL	5.28	0%
METminusPTLdivPTV	5.31	0.57%
All	5.32	0.76%

Table 11.5: Combined BDT_{VH} statistical sensitivities for the addition of different polarisation variables as input variables in the baseline MVA training.

The small improvements from the polarisation variables can be further understood from the high correlation between the polarisation variables and the existing input variables, as illustrated in Fig. 11.10. The polarisation variables are also highly correlated with each other, and hence explains the small improvement from adding all of them.

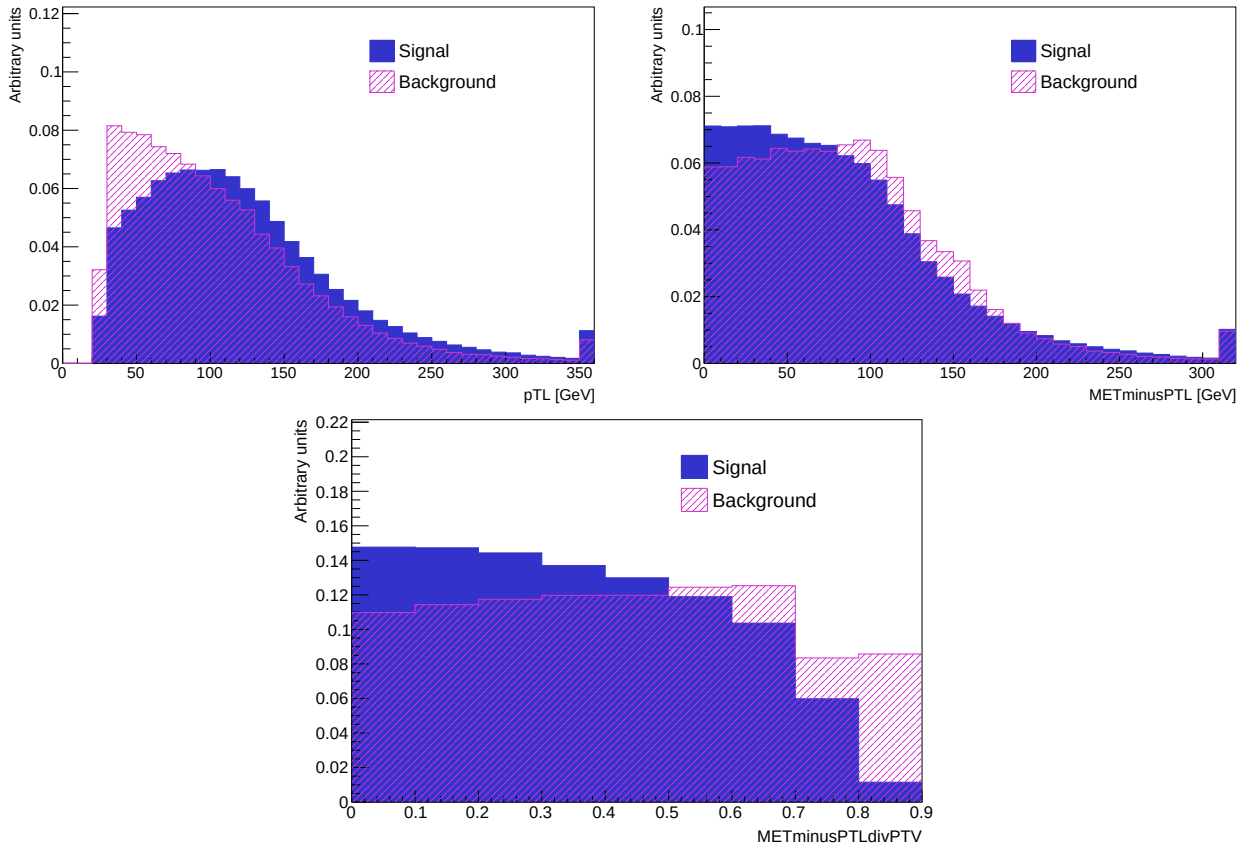


Figure 11.9: Signal and background distributions for variables associated to the W-boson polarisation in the 2-jet category. Clockwise from top-left: pTL, METminusPTL, METminusPTLdivPTV.

From these findings, it can be assumed that the polarisation information is already indirectly contained in the default input variables and therefore brought no improvement in sensitivity, thus the polarisation variables were not included in the 1-lepton channel.

11.1.4 Pseudo-continuous b-tagging variables

The two Higgs boson candidate jets are selected using a 70% b-tagging efficiency working point. Thus, the output distribution of the MV2c10 b-tagging algorithm for each candidate jet contains two bins corresponding to 70%-60% and 60%-0% b-tagging efficiencies, by definition of the pseudo-continuous b-tagging discriminant, as detailed in Chapter 6. Due to the large bc component in the background of the 1-lepton channel (85% of the $t\bar{t}$ background), these binned distributions are expected to provide further separation power between signal and background. This is because the 70%-60% bin is enriched in c- and light-jets, while the 60%-0% is enhanced in b-jets. The binned

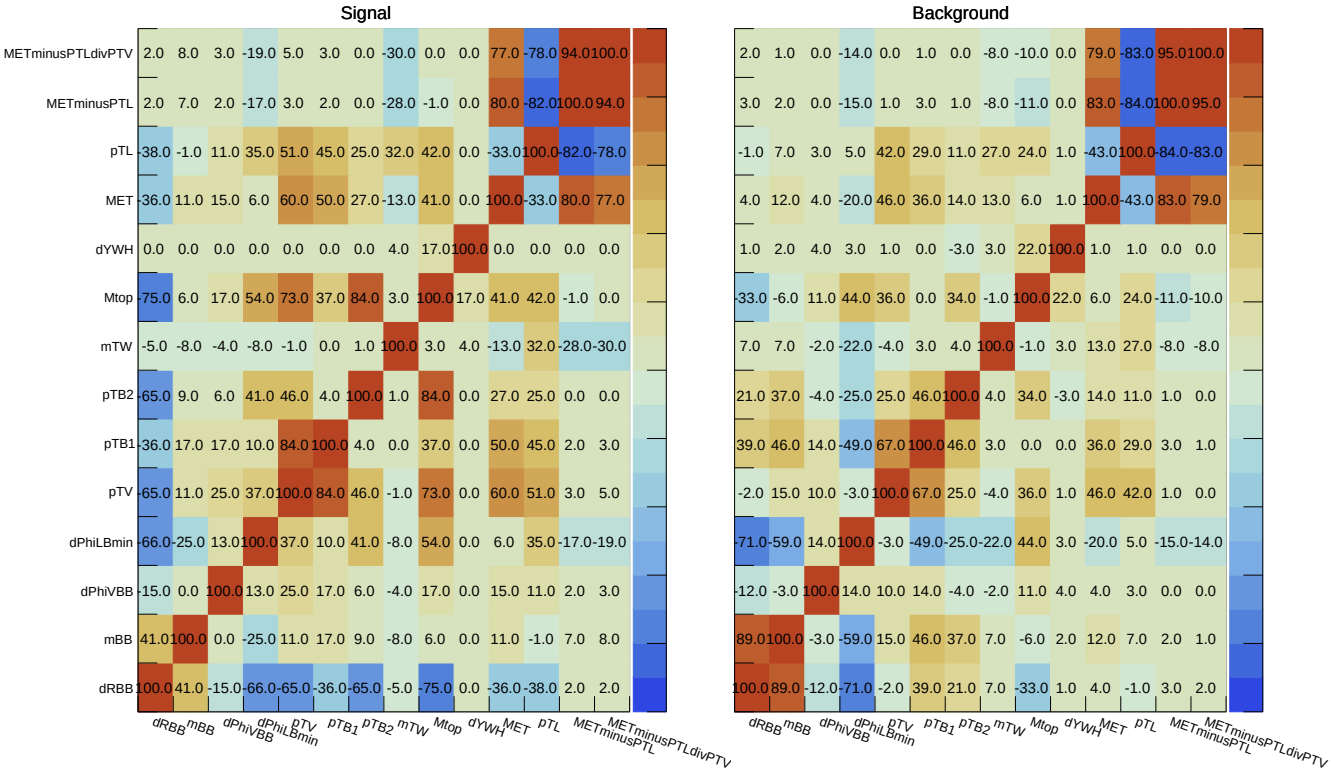


Figure 11.10: Matrices of the correlation coefficients between the input variables with the addition of polarisation variables for the 2 jet category. For both signal and background, the polarisation variables (pTL, METminusPTL, METminuspTLdivPTV) are strongly correlated with each other and the existing input variables, especially MET. This suggests that the existing input variables already contain sufficient information related to the polarisation.

MV2c10 output distributions for the leading and sub-leading Higgs candidate jets were incorporated into the MVA as two additional variables:

- bin_MV2c10B1: binned MV2c10 distribution of the leading b-tagged jet,
- bin_MV2c10B2: binned MV2c10 distribution of the sub-leading b-tagged jet.

The signal and background distributions for the additional variables are presented in Fig. 11.11

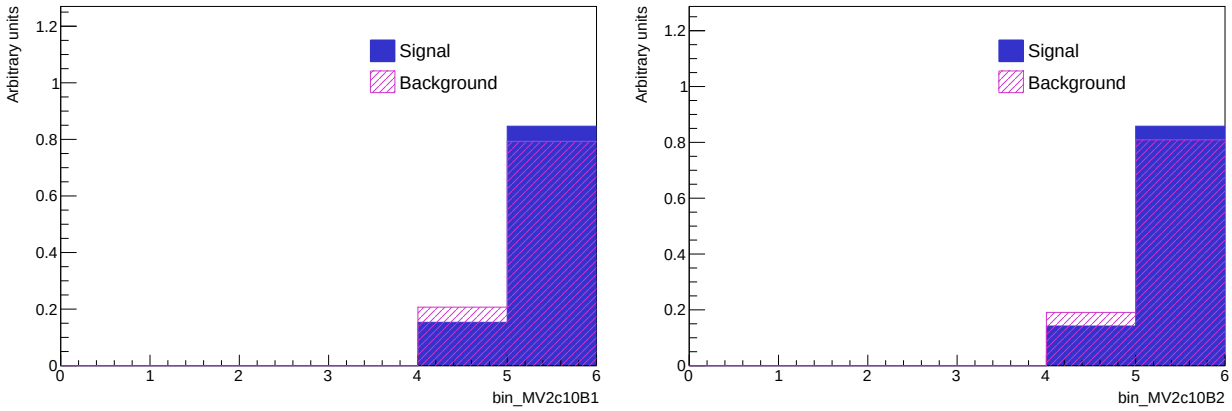


Figure 11.11: Signal and background distributions for the pseudo-continuous b-tagging variables, each with 6 pseudo-continuous b-tagging bins, in the 2-jet category.

The sensitivities of the trainings using the optimised Run-2 BDT_{VH} hyperparameters with and without pseudo-continuous b-tagging variables (PCBT) variables are displayed in Table 11.6, which demonstrates that the sensitivity gain from exploiting information on the flavour of jets was 9.3%.

Configuration	Sensitivity	Improvement
Without PCBT	5.50	-
With PCBT	6.01	9.3%

Table 11.6: Combined BDT_{VH} statistical sensitivities for MVA trainings using the optimised Run-2 BDT_{VH} hyperparameters with and without pseudo-continuous b-tagging (PCBT) variables.

11.1.5 Events in signal and control regions

During the Run-2 analysis, the introduction of new control regions (CR) redefined the signal region (SR), as detailed in Section 10.3.1. As a consequence, a training with events in SR only was tested to see if this can increase signal purity and therefore performance.

Events in both SR and CRs (SR+CR) were also used in a separate training to examine the effect of the new definitions on the BDT. These trainings were performed using the optimal GradBoost MVA configuration.

The sensitivities of the trainings are shown in Table 11.7, where it can be observed that there was negligible difference between the two configurations. In addition, using SR only events led to the removal of over 50% of background events, resulting in more significant overtraining, as shown in Fig. 11.12. As worse performance and overtraining was observed with SR only trainings, it was decided that SR+CR trainings would be used for subsequent studies.

Training	Sensitivity
Signal region only	5.53
Signal and control regions	5.54

Table 11.7: Combined BDT_{VH} statistical sensitivities for trainings with events in the signal region only and the combined signal and control regions using the optimal GradBoost configuration.

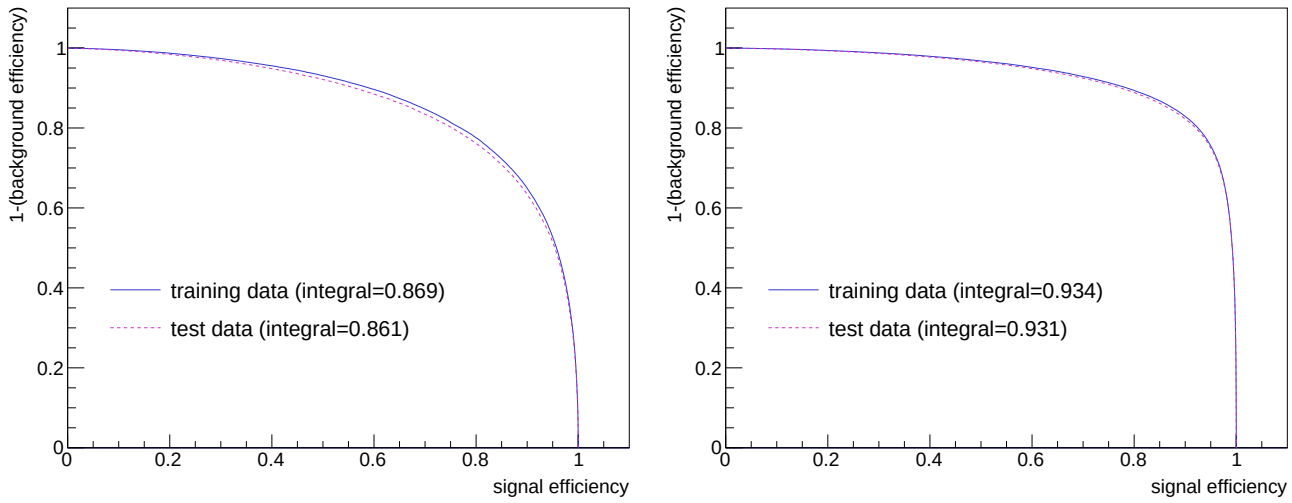


Figure 11.12: ROC curves as overtraining checks from BDT_{VH} trainings with SR only (left) and SR+CR (right) events for the 2-jet category.

11.1.6 Final configuration

Table 11.8 summarises the optimised BDT_{VH} hyperparameters used in the high p_T^V region for the Run-2 analysis. This optimised configuration with the addition of pseudo-continuous b-tagging variables was adopted as the default 1-lepton MVA in the high p_T^V

region for the Run-2 analysis. The new default 1-lepton MVA yielded an improvement of 12.3% over the baseline MVA, with similar improvements also adopted in other channels.

TMVA Option	Value	
	Baseline MVA	New default MVA
NTrees	200	600
MaxDepth	4	4
MinNodeSize	5%	5%
nCuts	100	100
BoostType	AdaBoost	GradBoost
AdaBoostBeta/Shrinkage	0.10	0.50
SeparationType	GiniIndex	GiniIndex
PruneMethod	NoPruning	NoPruning
Other configurations		
Restrict range	Yes	Yes
Signal and control regions	Both	Both
Variables	Nominal	Nominal+PCBT
Performance		
Sensitivity	5.28	6.01
Improvement	-	12.3%

Table 11.8: Comparison of BDT_{VH} hyperparameters used to train the baseline and new default 1-lepton MVA in the high p_T^V region for the Run-2 analysis. Shrinkage is the TMVA option for the learning rate of the GradBoost algorithm.

11.2 $VH, H \to b\bar{b}$ medium p_T^V region

As GradBoost performed better than AdaBoost in the high p_T^V region, only this algorithm was considered in the optimisation of the MVA for the medium p_T^V region (75 GeV $< p_T^V <$ 150 GeV). A 2D scan using the same combinations of the number of trees and the learning rates was conducted. Similar to the high p_T^V region, the performance plateaued at around 600 trees as shown in Fig. 11.13. As the improvement from 600 to 800 number of trees was not deemed significant compared to the variation observed when training multiple identical BDTs for the best performing learning rates of 0.45 and 0.5, it was decided that NTrees = 600 was the best choice in terms of performance and computational efficiency.

For NTrees = 600, the best performing learning rate was 0.5, hence this learning rate was also used for a 2D scan of the number of trees and MaxDepth. The performance

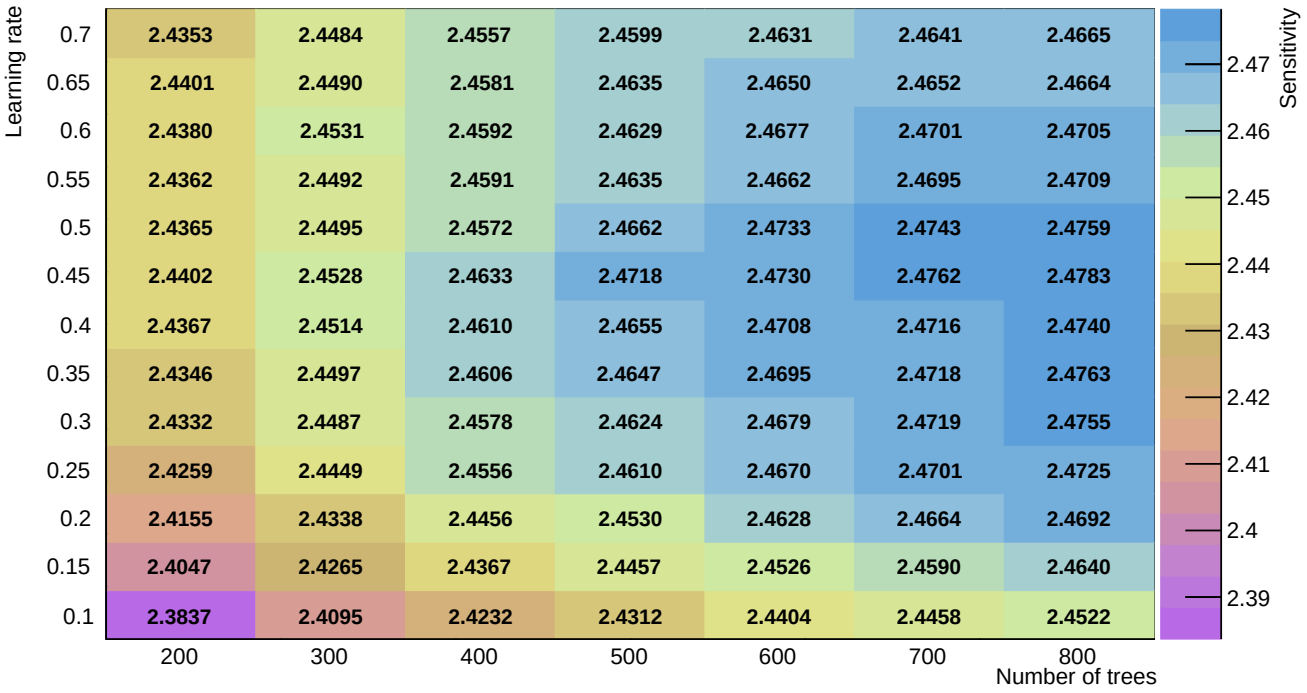


Figure 11.13: Combined BDT_{VH} statistical sensitivities for GradBoost trainings with different combinations of learning rates and number of trees in the medium p_T^V region.

for the different combinations is illustrated in Fig. 11.14, where it can be seen that the sensitivity plateaus at 600 to 1000 trees for MaxDepth = 4. As the improvement from 600 to 1000 trees is not significant, and as larger trees are more computationally expensive and more prone to overtraining, GradBoost with a learning rate of 0.5, 600 trees, and MaxDepth=4 was also chosen to be the optimised configuration for the medium p_T^V region. ROC curves in Fig 11.15 showed that there was no overtraining present in either jet categories for the optimal configuration. The optimised hyperparameter and boosting algorithm configuration provided an increase of 6.0% in the sensitivity with respect to that of the baseline MVA.

A summary of the optimised BDT_{VH} hyperparameters used in the medium p_T^V region for the Run-2 analysis is displayed in Table 11.9. Similarly to the high p_T^V region, a substantial improvement was seen from adding the pseudo-continuous b-tagging (PCBT) variables. Therefore, the new configuration with PCBT variables was used as the new default 1-lepton MVA in the medium p_T^V region for the Run-2 analysis, which offered an improvement of 10.7% compared to the baseline MVA.

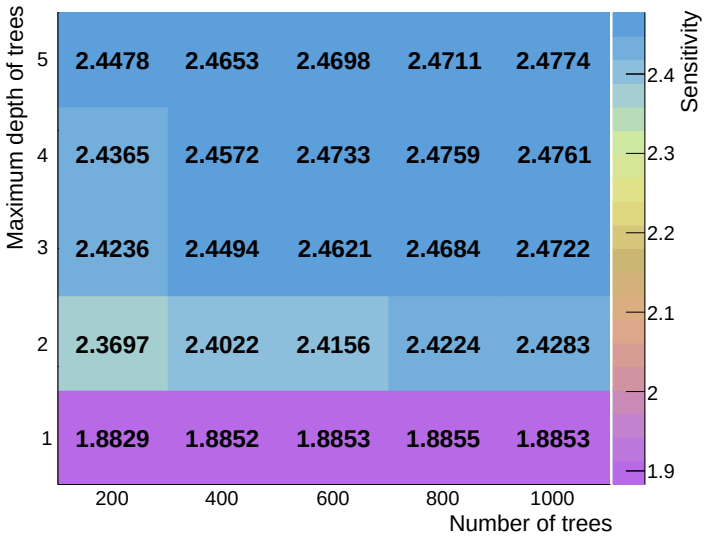


Figure 11.14: Combined BDT_{VH} statistical sensitivities for GradBoost with a learning rate of 0.50 with different combinations of number of trees and maximum tree depths in the medium p_T^V region.

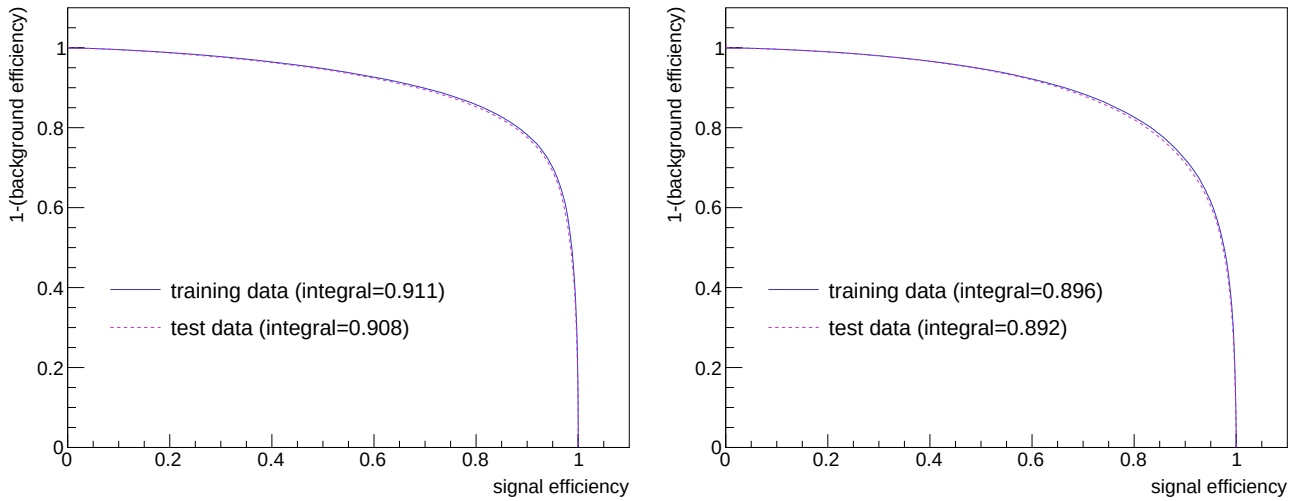


Figure 11.15: ROC curves as overtraining checks for the optimal GradBoost configuration for the 2-jet (left) and 3-jet (right) categories in the medium p_T^V region.

11.3 Combined $VH, H \to b\bar{b}$ performance

The sensitivities from the medium p_T^V and high p_T^V regions were summed in quadrature to ascertain the overall improvement in the combined $VH, H \to b\bar{b}$ MVA performance using the new default 1-lepton MVA for the Run-2 analysis. The comparison of the performance of the former and new default MVA is presented in Table 11.10.

TMVA Option	Value	
	Baseline MVA	New default MVA
NTrees	200	600
MaxDepth	4	4
MinNodeSize	5%	5%
nCuts	100	100
BoostType	AdaBoost	GradBoost
AdaBoostBeta/Shrinkage	0.10	0.50
SeparationType	GiniIndex	GiniIndex
PruneMethod	NoPruning	NoPruning
Other configurations		
Variables	Nominal	Nominal+PCBT
Performance		
Sensitivity	2.34	2.59
Improvement	-	10.7%

Table 11.9: Comparison of BDT_{VH} hyperparameters used to train the baseline and new default 1-lepton MVA in the medium p_T^V region for the Run-2 analysis. Shrinkage is the TMVA option for the learning rate of the GradBoost algorithm.

Configuration	Sensitivity	Improvement
Baseline (without PCBT)	5.78	-
Optimal GradBoost	6.54	11.7%

Table 11.10: Comparison of the combined BDT_{VH} statistical sensitivities for trainings in the medium p_T^V and high p_T^V regions between the default Run-1 MVA (Baseline) and the new default Run-2 MVA (Optimal GradBoost) in the 1-lepton channel.

11.4 $VZ, Z \to b\bar{b}$

The new default 1-lepton BDT_{VH} hyperparameters and additional PCBT variables were used to train the MVA diboson cross-check, BDT_{VZ}, but there was significant overtraining present due to lower training statistics for the VZ signal (1,513,174 events in 2-jet and 1,479,709 events in 3-jet for VH versus 928,875 events in 2-jet and 1,174,354 events in 3-jet for VZ).

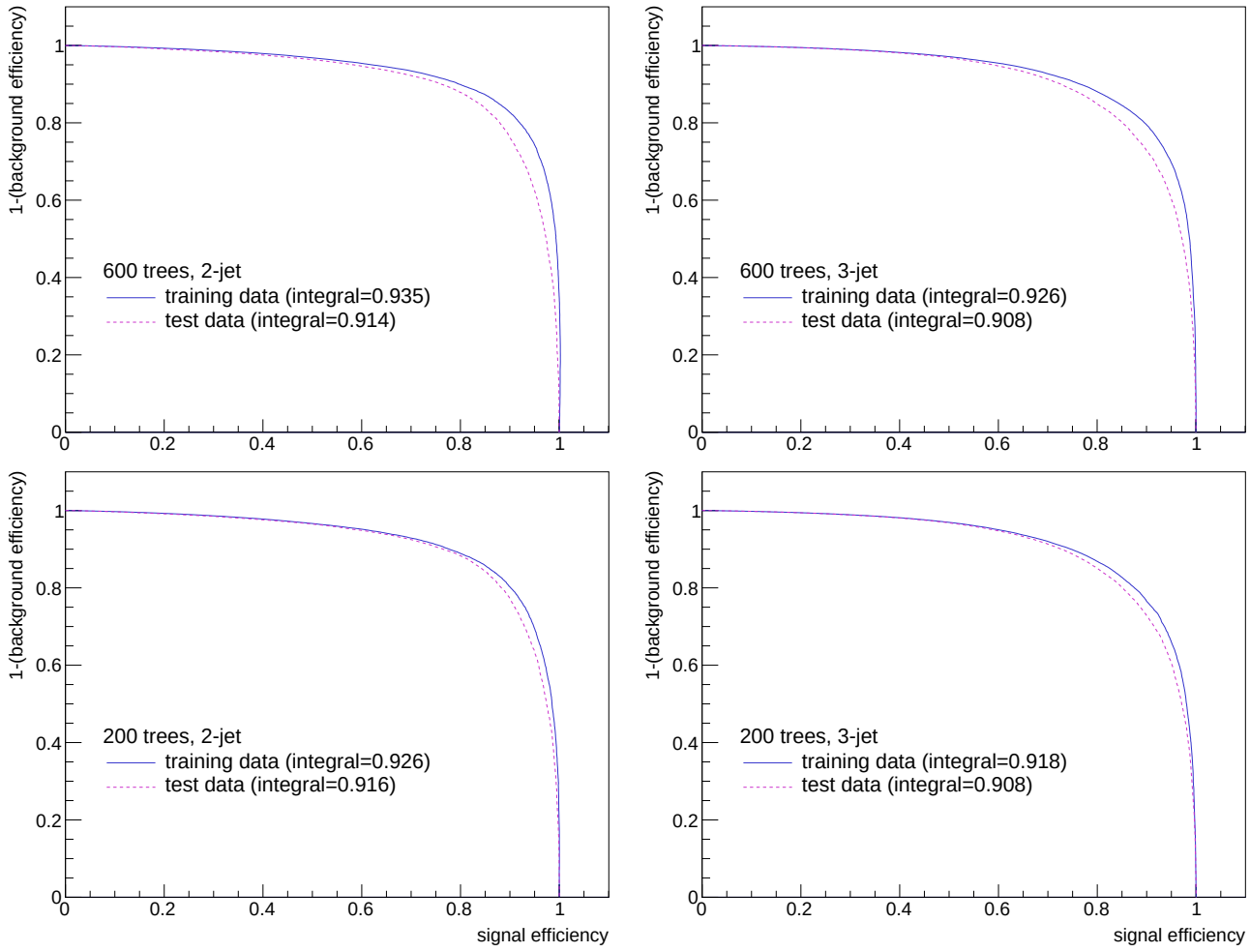


Figure 11.16: ROC curves as overtraining checks for GradBoost BDT_{VZ} trainings with a learning rate of 0.50 for the 2-jet (left) and 3-jet (right) categories. Decreasing the number of trees from 600 to 200 reduced overtraining significantly.

To investigate if overtraining can be reduced, the number of trees was decreased from 600 to 200 and lower maximum tree depths were explored. While overtraining decreased with maximum tree depth, performance also decreased as indicated by the ROC integral in Fig. 11.17. However, the sensitivities in Table 11.11 revealed that the

performance for MaxDepth = 2, 3, 4 were comparable, which was also reflected by their ROC integral for test data. For MaxDepth = 2, there were no signs of overtraining in the 2-jet case but there were still some overtraining in the 3-jet case, which further indicated that it was the VZ signal training statistics that was causing overtraining as there are a lot more background events in 3-jet than 2-jet. Although MaxDepth = 1 had the least overtraining, it was decided that MaxDepth = 2 would be the optimal choice because it gave a comparable performance to MaxDepth = 4 without significant overtraining and ideally the BDT_{VH} and BDT_{VZ} hyperparameters should be as close as possible.

Maximum tree depth	Sensitivity
1	8.35
2	9.04
3	9.01
4	9.07

Table 11.11: Combined BDT_{VZ} statistical sensitivities for GradBoost trainings with a learning rate of 0.50, 200 trees and different maximum tree depths.

The optimised BDT_{VZ} hyperparameters for the Run-2 analysis are summarised in Table 11.12. The optimal configuration with the inclusion of pseudo-continuous b-tagging variables was used to train the new default VZ MVA in the 1-lepton channel for the Run-2 analysis.

11.5 Summary

The BDT_{VH} hyperparameters were optimised in the 1-lepton channel for the Run-2 analysis. The optimisation involved changing the boosting algorithm to GradBoost and increasing the learning rate to 0.50 and the number of trees to 600. The optimal configuration provided an improvement of 5.6% and 4.1% in the medium and high p_T^V regions, respectively, when compared to the baseline configuration, which was optimised for the Run-1 analysis.

The addition of pseudo-continuous b-tagging (PCBT) variables to the optimal configuration in the medium and high p_T^V regions yielded a larger improvement of 10.7% and 12.3%, respectively, over the baseline MVA. As a result, the optimal configuration

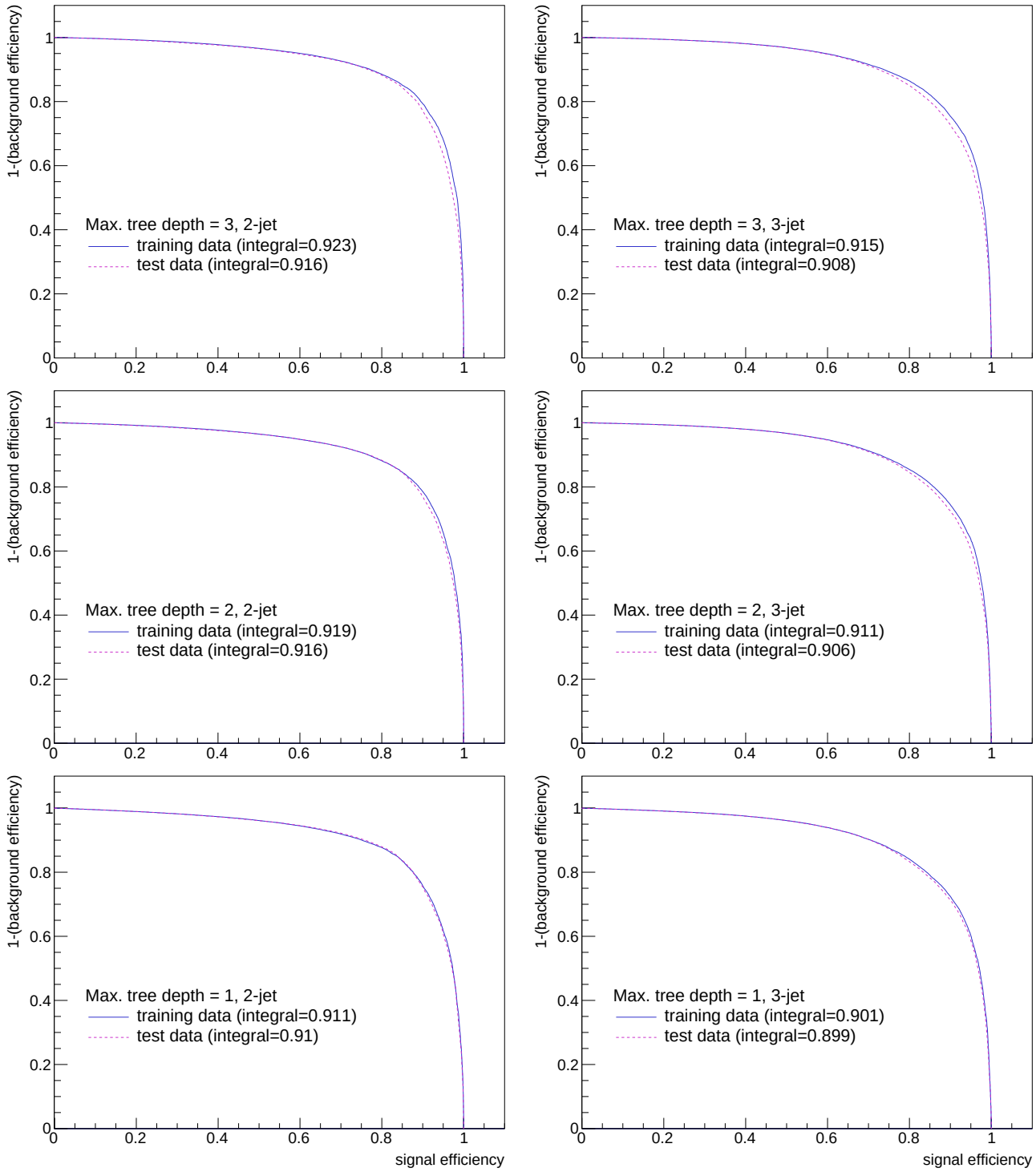


Figure 11.17: ROC curves as overtraining checks for GradBoost BDT_{VZ} trainings with a learning rate of 0.50, 200 trees and maximum tree depths of between 1 and 3 for the 2-jet (left) and 3-jet (right) categories. Overtraining is observed to decrease with maximum tree depth, but performance also decreases, as denoted by the ROC integrals. Furthermore, more overtraining is present in 3-jet than 2-jet.

with PCBT variables was chosen as the new default 1-lepton MVA for both p_T^V regions. Overall, the combined improvement in performance from both regions was 11.7%.

The new default 1-lepton MVA was used to train the diboson cross-check MVA,

TMVA Option	Value	
	Baseline MVA	New default MVA
NTrees	200	200
MaxDepth	4	2
MinNodeSize	5%	5%
nCuts	100	100
BoostType	AdaBoost	GradBoost
AdaBoostBeta/Shrinkage	0.10	0.50
SeparationType	GiniIndex	GiniIndex
PruneMethod	NoPruning	NoPruning
Other configurations		
Variables	Nominal	Nominal+PCBT
Performance		
Sensitivity	8.51	9.04
Improvement	-	6.2%

Table 11.12: Comparison of BDT_{VZ} hyperparameters used to train the baseline and new default 1-lepton VZ MVA for the Run-2 analysis. Shrinkage is the TMVA option for the learning rate of the GradBoost algorithm.

BDT_{VZ}. However, due to the presence of significant overtraining, the number of trees was reduced from 600 to 200 trees and MaxDepth value was decreased from 4 to 2 for the optimised BDT_{VZ} hyperparameters for the Run-2 analysis. The new default 1-lepton BDT_{VZ} resulted in an improvement of 6.2% in the BDT_{VZ} sensitivity.

Due to the improvements shown in the sensitivities, the new default 1-lepton BDT_{VH} and BDT_{VZ} were adopted as the default configurations in other lepton channels.

Chapter 12

Systematic uncertainties

Uncertainties affecting the $VH, H \rightarrow b\bar{b}$ analysis are separated into categories depending on the source of the uncertainty. The two main categories are experimental and modelling uncertainties. While experimental uncertainties are generally attributed to the simulation of detector related quantities and the reconstruction of objects, the modelling uncertainties are related to how well the simulated and data-driven backgrounds are modelled.

12.1 Experimental uncertainties

Experimental uncertainties are associated to the modelling of the detector-related quantities. The dominant uncertainties affecting the results are those related to the b-tagging efficiency, followed by the uncertainties in the jet energy scale and resolution. The main experimental uncertainties are summarised as follows:

- **Luminosity:** The uncertainty on the combined integrated luminosity for the 2015-2018 period is 1.7% [107], which is derived using the LUCID-2 detector [108] for the luminosity measurements.
- **Pile-up reweighting:** The pile-up in simulation is rescaled by a correction factor of 1.03 to describe the pile-up in data. The uncertainty is evaluated by varying the correction factor from 1.00 to 1.18 to obtain the respective up and down variations.
- **Leptons:** Uncertainties arise from the reconstruction, identification, isolation and trigger efficiencies of leptons. In addition, uncertainties relating to their energy scale and resolution are considered. The uncertainties are determined separately for electrons [55] and muons [56].
- $\mathbf{E_T^{miss}}$: The uncertainties on the jet and lepton scales and resolutions are propagated

to the E_T^{miss} calculation. Additional uncertainties on the reconstruction efficiency, scale, and resolution of the tracks used to calculate the soft term [63] also enter the E_T^{miss} calculation.

- **Jets:** Uncertainties associated to jets are divided into jet energy scale (JES) and jet energy resolution (JER) categories. The different sources of uncertainty in the calibration form a set of 30 independent NPs for the JES [60], while the uncertainty on JER is characterised by 8 NPs [61]. A dedicated uncertainty of 2%, corresponding to the modelling of the jet-vertex-tagger (JVT) discriminant, is assigned to the jets that were required to pass this selection.

- ***b*-tagging:** Uncertainties originate from the different sources related to the derivation of the data-to-MC scale factors, which are used to correct the mismodelling of b-tagging algorithms. As the scale factors are derived separately for the different jet flavours, a set of independent NPs, which incorporate the various uncertainties in each calibration, is assigned for each jet flavour, resulting in 29, 18 and 10 NPs for b-[75], c-[109] and light-flavour jets [110], respectively.

12.2 Modelling uncertainties

Modelling uncertainties are systematic uncertainties that originate from the approximations made in the descriptions of the MC simulations of signal and background processes. These uncertainties are determined by comparing a nominal model that uses a baseline set of approximations to an alternative model that uses alternative approximations. The alternative model is either created from an alternative generator or by varying the values of certain parameters of the generator used to simulate the nominal sample. For each process, there is often more than one systematic effect to consider, hence there is generally more than one alternative model. The modelling uncertainties are divided into normalisation, acceptance and shape uncertainties.

Normalisation uncertainties are calculated as an overall uncertainty on the total cross section of the process in the regions considered. Acceptance uncertainties account for the relative difference in the number of events predicted in each analysis region by different models. To control the correlations between the regions, the acceptance

uncertainties are calculated as double ratios between the yields in the different regions as

$$\frac{\text{Yield}(\text{Region}_A^{nom})}{\text{Yield}(\text{Region}_B^{nom})} \Big/ \frac{\text{Yield}(\text{Region}_A^{alt})}{\text{Yield}(\text{Region}_B^{alt})}, \tag{12.1}$$

where Region_A is the region with the highest purity and Region_B is the region to which the normalisation that is derived in the higher purity region is propagated, and the superscripts *nom* and *alt* denote the nominal and alternative MC models, respectively.

Shape uncertainties are determined for the most discriminant variables (m_{bb} and p_T^V) that enter the MVA as the variations in these variables are sufficiently large to account for the shape variations of the other input variables. In addition, m_{bb} and p_T^V are weakly correlated and this avoids the effect of double counting the variations. The standard procedure of deriving shape uncertainties is to obtain the distribution of a given variable for the nominal and alternative models. A ratio between the distributions of the nominal and alternative models is calculated, which corresponds to the uncertainty. As the nominal model is generally compared to more than one alternative model, the comparison resulting in the largest shape uncertainty is assigned as the overall uncertainty for the variable.

A BDT-reweighting (BDTr) approach [17] was introduced in the Run-2 analysis to improve the consideration of any differences that might occur in the other variables input to the MVA. In this approach, the phase space of the nominal model is reweighted using the p_T^V shape uncertainty to remove any difference, so this can be treated separately to m_{bb} and the other input variables. A BDT classifier is trained using all the MVA input variables except for p_T^V on the reweighted nominal sample and the alternative sample, separately, to produce the nominal and alternative output BDT distributions. The nominal BDT distribution is mapped onto the alternative BDT distribution to obtain a ratio, which is taken as a modelling uncertainty along with the p_T^V ratio.

The overall normalisation uncertainties for the main backgrounds (W+jets, Z+jets and $t\bar{t}$) are left unconstrained in the profile likelihood fit and are directly determined by data. Other background processes have prior normalisation uncertainties derived from cross section calculations as the data is less able to constrain the normalisation for these processes. The uncertainties in this section are obtained from Ref. [17].

12.2.1 Signal

The alternative models used to derive the signal modelling uncertainties are:

* Powheg [85] + MinLo [101] + Herwig 7 [111] for the PS/UE variations.
* Altering the renormalisation and factorisation scales of the nominal model for the scale variations.
* Varying the uncertainties related to the PDF4LHC set [112] for the PDF+α_S variations.

Normalisation uncertainties are implemented for the VH cross section and the $H \to b\bar{b}$ branching ratio. Acceptance uncertainties are considered for the scale, PS/UE, and PDF+α_S variations. The m_{bb} shape uncertainties are evaluated for the scale and PS/UE variations. Uncertainties relating to the production cross section and the $H \to b\bar{b}$ branching ratio follow the LHC Higgs Cross Section Working Group recommendations [113–115], while an uncertainty from the EW corrections at NLO is derived from the shape variation of the p_T^V distribution for the $qq \to VH$ process.

A summary of the signal modelling uncertainties is presented in Table 12.1.

Signal	
$H \to b\bar{b}$ branching ratio	1.7%
m_{bb} from scale variations	Shape only
m_{bb} from PDF+α_S variations	Shape only
m_{bb} from PS/UE variations	Shape only
p_T^V from NLO EW correction	Shape only
qq	
Cross section from scale variations	0.7%
Acceptance from scale variations	3-3.9% (WH), 6.7-12% (ZH)
Acceptance from PDF+α_S variations	1.8-2.2% (WH), 1.4-1.7% (ZH)
Acceptance from PS/UE variations	1-5%
gg	
Cross section from scale variations	25%
Acceptance from scale variations	37-100%
Acceptance from PDF+α_S variations	2.9-3.3%
Acceptance from PS/UE variations	5-20%

Table 12.1: Summary of the systematic uncertainties in the background modelling for the signal process.

12.2.2 V+ jets

For V+jets, the main background contributions are V+bb, V+bc, V+bl and V+cc, and are considered overall as the V+hf background. The alternative MC model is generated by MadGraph 5 [116] + Pythia 8 [102], where '+' denotes 'interfaced to', for the matrix element (ME), parton shower (PS), and underlying event (UE) variations. The NPs described in this section are implemented separately for W+ and Z+jets, with `V` = `W` or `Z` for the corresponding background process.

Due to the large uncertainty on the jet multiplicity distribution of the V+hf processes, the overall normalisation is separated for the 2-jet (`norm_Vbb_J2`) and 3-jet (`norm_Vbb_J3`) categories. The 1-lepton and 2-lepton channels have regions of high purity in W+hf and Z+hf, respectively, resulting in dedicated control regions (CRs) in each channel. These CRs are used to obtain constraints on the W+hf and Z+hf backgrounds, which are extrapolated to the 0-lepton channel. Therefore, acceptance uncertainties (`VbbNorm_L0`) are applied in the 0-lepton channel to account for these extrapolations.

The flavour composition uncertainties of the V+hf backgrounds are implemented as the ratio of the bc (`VbcVbbRatio`), bl (`VblVbbRatio`) and cc (`VccVbbRatio`) components relative to the dominant bb component. Due to the strong suppression of V+l and V+cl processes by the requirement of 2 b-tags, they compose less than 1% of the background in any region, hence it is sufficient to apply a normalisation uncertainty for each of the processes, named `VlNorm` and `VclNorm`, respectively.

Shape uncertainties are derived differently for W+ jets and Z+ jets:

- ***W*+jets:** The largest shape variation for p_T^V is due to the ME comparison between Sherpa 2.2.1 [103] and MadGraph 5 + Pythia 8, and is implemented as the NP, `WPtV_BDTr`. BDTr is included in the name of the NP to indicate that the p_T^V shape systematic is used along with the BDTr approach for all other input variables, as described in Section 12.2. The BDTr approach is used to compute the shape ratio between the BDT distributions of Sherpa 2.2.1 and MadGraph to obtain the BDTr-based shape uncertainty, `BDTr_W_SHtoMG5`.

- ***Z*+jets:** Shape uncertainties for m_{bb} and p_T^V are determined from comparisons

between data and the nominal model from simulation, SHERPA 2.2.1, excluding the events in the Higgs mass window (110 GeV $< m_{bb} <$ 140 GeV). The data-driven modelling comparisons involve subtracting non-Z+jets backgrounds from the data in the 2-lepton signal and control regions. The discrepancies in the distributions of the non-Z+jets background subtracted data and the nominal model for m_{bb} and p_T^V correspond to the shape uncertainties, `ZMbb` and `ZPtV`, respectively.

Table 12.2 summarises the uncertainties for V+jets background modelling.

W+jets	
W+hf normalisation	Unconstrained (2-jet, 3-jet)
W+ll normalisation	32%
W+cl normalisation	37%
W+bc-to-W+bb ratio	15% (0-lepton), 30% (1-lepton)
W+cc-to-W+bb ratio	10% (0-lepton), 30% (1-lepton)
W+ll-to-W+bb ratio	26% (0-lepton), 23% (1-lepton)
1-to-0 lepton extrapolation	5%
p_T^V	Shape only
BDTr	Shape only
Z+jets	
Z+hf normalisation	Unconstrained (2-jet, 3-jet)
Z+ll normalisation	18%
Z+cl normalisation	23%
Z+bc-to-Z+bb ratio	30-40%
Z+cc-to-Z+bb ratio	13-16%
Z+ll-to-Z+bb ratio	20-28%
2-to-0 lepton extrapolation	7%
p_T^V	Shape only
m_{bb}	Shape only

Table 12.2: Summary of the systematic uncertainties in the background modelling for W+jets and Z+jets production.

12.2.3 $t\bar{t}$

As the modelling of the $t\bar{t}$ background differs for the 0+1-lepton (Section 10.1) and 2-lepton (Section 10.4.1) channels, the uncertainties are considered separately in the 0+1-lepton and 2-lepton channels.

- **0+1-lepton:** The unconstrained normalisation is separated for the 2-jet (`norm_ttbar_J2`) and 3-jet (`norm_ttbar_J3`) categories. As the 1-lepton channel has a better constraining power on the $t\bar{t}$ normalisation, the constraint is

extrapolated from 1-lepton to 0-lepton. Therefore, an extrapolation uncertainty (`TTbarNorm_L0`) is applied in 0-lepton to represent the uncertainty in the prediction of the relative normalisations between the two regions.

Different variations arise from the comparison between the nominal POWHEG [85] + PYTHIA 8 [102] and the following alternative MC models:

* POWHEG + HERWIG 7 [111] for the PS variation.
* MADGRAPH 5_aMC@NLO [117] + PYTHIA 8 for the ME variation.
* Doubling or halving the renormalisation and factorisation scales of the nominal model for the scale variations.

Flavour composition uncertainties are separated into bb, bc, and Oth (bl, cc, cl, ll) components. Based upon the studies in Section 13.2, the flavour composition uncertainties are applied in the fit as the ratio of the bc and Oth components relative to the bb component separated into its ME and PS variations, resulting in the NPs: `TTbarbcMEACC, TTbarbcPSACC, TTbarOthMEACC`, and `TTbarOthPSACC`.

The largest shape variation for p_T^V originates from the ME variation, and is implemented as the NP, `TTbarPtV_BDTr`. BDTr is included in the name of the NP as it is used to reweight the nominal model prior to BDT training for the ME variation. Section 13.1 examined the NP, and due to the different behaviour observed between the jet categories, the NP was decorrelated between the 2-jet (`TTbarPtV_BDTr_J2`) and 3-jet (`TTbarPtV_BDTr_J3`) categories.

As the systematic effects of the ME and PS variations are dominant in the $t\bar{t}$ uncertainties, a BDT is trained for each variation to separate the nominal model from the alternative models. In each BDT, the nominal model is reweighted by the shape ratio from the comparison of the nominal model to its respective alternative model. The shape ratios of the BDT outputs of the nominal and the alternative ME and PS models enter the profile likelihood fit as `BDTr_ttbar_ME` and `BDTr_ttbar_PS`, respectively.

- **2-lepton:** As systematic uncertainties evaluated on the extrapolation from the $e\mu$CR to the SR are negligible, the limited statistics in the $e\mu$CR data and the MC statistical uncertainty on the scale factor are assigned as the main sources of

uncertainty in the data-driven estimation of the top background.

Table 12.3 presents a summary of the $t\bar{t}$ modelling uncertainties in 0- and 1-lepton.

	$t\bar{t}$
$t\bar{t}$ normalisation	Unconstrained (2-jet, 3-jet)
1-to-0 lepton extrapolation	8%
$t\bar{t}$ bc-to-bb ratio (ME variation)	7.6-8.2% (0-lepton), 1.3-3.8% (1-lepton)
$t\bar{t}$ bc-to-bb ratio (PS variation)	2.1-3.2% (0-lepton), 1.5-7.1% (1-lepton)
$t\bar{t}$ Oth-to-bb ratio (ME variation)	2.8-6.4% (0-lepton), 3.3-5.7% (1-lepton)
$t\bar{t}$ Oth-to-bb ratio (PS variation)	5.6-13.2% (0-lepton), 0.3-2.1% (1-lepton)
p_T^V	Shape only
BDTr ME variation	Shape only
BDTr PS variation	Shape only

Table 12.3: Summary of the systematic uncertainties in the background modelling for the $t\bar{t}$ production.

12.2.4 Single-top

The single top-quark (single-top) production arises from three different channels, and normalisation uncertainties are adopted separately for the t- (`stoptNorm`), s- (`stopsNorm`), and Wt- (`stopWtNorm`) channels. As there are negligible contributions to the background in the 0- and 2-lepton channels, the uncertainties are determined in 1-lepton, and applied to 0-/2-lepton. The single-top contribution in the 1-lepton is dominated by the t- and Wt-channels, and due to its very minor contribution, acceptance and shape uncertainties are not considered for the s-channel.

The variations from the comparisons of the nominal Powheg [85] + Pythia 8 [102] using the DR scheme in the Wt-channel (Section 10.1) to the alternative MC models are as follows:

* Powheg + Herwig++ [118] for the PS variation.
* MadGraph 5_aMC@NLO [117] + Herwig++ for the ME variation.
* Doubling or halving the renormalisation and factorisation scales of the nominal model for the scale variations.
* Powheg + Pythia 8 using the diagram subtraction (DS) [104] scheme, which involves a subtraction term in the ME to cancel events that overlap with single-top events in the Wt-channel.

The topology significantly differs between the decays of $Wt \to b\bar{b}$ and $Wt \to Oth$, where Oth represents events with one or fewer b-jets present, and hence the sub-processes are considered separately. Acceptance uncertainties for the Wt- (`stopWtbbACC` and `stopWtOthACC`) and t- (`stoptACC`) channels are calculated between the 2-jet and 3-jet categories.

In the Wt-channel, the largest variation in the p_T^V shape (`StopWtPTV`) originated from the ME comparison for both sub-processes, and the largest m_{bb} shape variation (`StopWtMBB`) was from the ME comparison for the $Wt \to b\bar{b}$ process. In the t-channel, the scale variations and PS comparison provided the largest shape variations in m_{bb} (`StoptMBB`) and p_T^V (`StoptPTV`), respectively.

A summary of the single-top modelling uncertainties is shown in Table 12.4.

	Single-top
Normalisation	4.6% (s-), 4.4% (t-), 6.2% (Wt-)
Acceptance in 2-jet	17% (t-), 55% ($Wt \to bb$), 24% ($Wt \to Oth$)
Acceptance in 3-jet	20% (t-), 51% ($Wt \to bb$), 21% ($Wt \to Oth$)
m_{bb}	Shape only (t-, $Wt \to bb$, $Wt \to Oth$)
p_T^V	Shape only (t-, $Wt \to bb$, $Wt \to Oth$)

Table 12.4: Summary of the systematic uncertainties in the background modelling for the t-, s-, and Wt-channels of the single-top process.

12.2.5 Diboson

Three different contributions (WW, WZ, ZZ) are considered for the diboson background, giving rise to the normalisation uncertainties, `WWNorm`, `WZNorm`, and `ZZNorm`, respectively. Only the normalisation uncertainty is evaluated for WW due to its minor contribution ($<$1%) to the background.

Alternative comparisons to the nominal Sherpa 2.2.1 [103] include

* Powheg + Herwig++ [118] for the PS/UE variations.
* Powheg + Pythia 8 [102] for the ME variation.
* Changing the renormalisation and factorisation scales of the nominal model for the scale variations.

The PS/UE variations give rise to the NP, `VZ_UEPS_Acc`, with an acceptance uncertainty (`VZ_UEPS_J3`) added to account for the relative yields between the jet

categories. Additional uncertainties are included in the fit to account for the residual effects of the PS/UE variations in the 0-lepton channel relative to the 1-lepton (`WZUEPSResid_L0`) and 2-lepton (`ZZUEPSResid_L0`) channels.

Scale variations in inclusive jet multiplicity bins are used to compute an overall acceptance uncertainty in the 2-jet category (`VZ_QCDscale_J2`), an uncertainty for the migration of events between the 2-jet and 3-jet categories (`VZ_QCDscale_J3`), and an uncertainty due to the veto of events with 4 jets in the 0- and 1-lepton channels (`VZ_QCDscale_JVeto`).

Shape uncertainties for m_{bb} and p_T^V are considered for the ME (`VZMbbME` and `VZPTVME`, respectively) and PS/UE variations (`VZMbbPSUE` and `VVPTVPSUE`, respectively).

Table 12.5 summarises the diboson modelling uncertainties.

WW	
Normalisation	25%
WZ	
Normalisation	26%
Acceptance from PS/UE variations in 2-jet	4%
Acceptance from PS/UE variations for 2- to 3-jet	11%
1-to-0 lepton residual effect	11%
Acceptance from scale variations	13-21%
m_{bb}, p_T^V from scale variations	Shape only
m_{bb}, p_T^V from PS/UE variations	Shape only
m_{bb}, p_T^V from ME variations	Shape only
ZZ	
Normalisation	20%
Acceptance from PS/UE variations in 2-jet	6%
Acceptance from PS/UE variations for 2- to 3-jet	7% (0-lepton), 3% (2-lepton)
2-to-0 lepton residual effect	6%
Acceptance from scale variations	10-18%
m_{bb}, p_T^V from scale variations	Shape only
m_{bb}, p_T^V from PS/UE variations	Shape only
m_{bb}, p_T^V from ME variations	Shape only

Table 12.5: Summary of the systematic uncertainties in the background modelling for the diboson process. The shape only uncertainties are correlated between WZ and ZZ uncertainties.

12.2.6 Multi-jet

The multi-jet background is negligible in the 0- and 2-lepton channels, thus the uncertainties are only considered in the 1-lepton channel, where the background is obtained

using a data-driven estimation, as described in Section 10.4.2.

Normalisation uncertainties are uncorrelated between the electron and muon sub-channels in the 2-jet (`MJNorm_2J_El` and `MJNorm_2J_Mu`) and 3-jet (`MJNorm_3J_El` and `MJNorm_3J_Mu`) categories. The uncertainties are estimated from variations in the m_T^W template distribution, arising from the use of an alternative variable that also provides good discrimination between multi-jet and electroweak processes in the template fit, and from including the $E_T^{miss} < 30$ GeV region in the electron sub-channel fit, which also increases the multi-jet contribution.

Shape uncertainties are determined from changing the definition of the multi-jet control region by tightening the isolation requirements, resulting in the `MJReduced` nuisance parameter. An additional shape uncertainty (`MJSFsCR`) originates from the discrepancy in the nominal multi-jet template evaluated with and without applying the fitted normalisation of the top and W+jet backgrounds in the multi-jet CR.

Chapter 13

Statistical analysis

The $VH, H \to b\bar{b}$ fitting procedure employs a profile likelihood formalism to extract the significance and signal strength, which is defined as

$$\mu = \frac{\sigma \times B}{\sigma_{SM} \times B_{SM}}, \tag{13.1}$$

where σ is the cross section of the production process and B is the branching ratio of the decay channel, the subscript SM denotes the SM prediction, and the numerator refers to the experimental measurements.

The profile likelihood formalism is implemented to extract the parameter of interest, μ, by maximising a likelihood function for binned distributions, which is defined using the Poisson probabilities of the bins:

$$\mathcal{L}(\mu) = \prod_{i=1}^{N_{bins}} \frac{(\mu s_i + b_i) n_i}{n_i!} e^{-(\mu s_i + b_i)}, \tag{13.2}$$

where N_{bins} is the total number of bins, n_i is the number of observed data events in bin i, and s_i and b_i are the expected number of signal and background events in bin i, respectively. The effects of the systematic uncertainties on the signal and background predictions are incorporated into the likelihood function as a vector of nuisance parameters (NPs), $\boldsymbol{\theta}$, which are constrained by Gaussian or log-normal probability distribution functions, with each NP corresponding to a systematic uncertainty, such that Equation (2.1) becomes

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N_{bins}} \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})) n_i}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \times \mathcal{L}_{aux}(\boldsymbol{\theta}), \tag{13.3}$$

where $\mathcal{L}_{aux}(\boldsymbol{\theta})$ is a penalty term.

For a number of systematic uncertainties, N_{syst}, the penalty term is defined as

$$\mathcal{L}_{aux}(\boldsymbol{\theta}) = \prod_{j=1}^{N_{syst}} \frac{1}{\sqrt{2\pi}\sigma_j} e^{\frac{-(\bar{\theta_j}-\theta_j)^2}{2\sigma_j^2}}, \tag{13.4}$$

where, for systematic uncertainty j, $\bar{\theta_j}$ is the central value of the Gaussian, θ_j is the best fit value, and σ_j is the prior uncertainty of θ_j. The central values and prior uncertainties are estimated from data or external measurements. This term penalises NPs that have large deviations from their central values. Through maximising the likelihood function, the NPs can be shifted or "pulled" away from their prior estimate to improve the agreement between MC simulation and data and their uncertainties "constrained". The pulls and constraints of the NPs can be visualised in a 'pulls comparison' plot. Nuisance parameters that are completely determined from the fit, known as free floating normalisations, have no prior uncertainties and hence no penalty is set for them. In addition, the statistical uncertainties of the MC simulations are considered in the likelihood function by introducing Poissonian priors in each bin of the fitted distribution by applying the Beeston-Barlow technique [119].

The test statistic, q_μ, is defined using a profile likelihood ratio as

$$q_\mu = -2\ln\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{13.5}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are parameters that maximise the likelihood fit, known as the unconditional fit, and $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ are the nuisance parameter values that maximise the likelihood fit for a given μ value, known as the conditional fit.

Discrepancies between the hypothesis ($\mu s+b$) and the observed data can be quantified using the test statistic

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu, \mu) dq_\mu, \tag{13.6}$$

where $q_{\mu,obs}$ is the observed test statistic and $f(q_\mu, \mu)$ is the probability density function of q_μ.

For testing the background-only hypothesis, where $\mu = 0$, the value of p_0 is used. If p_0 is small, this indicates that the probability of the background-only hypothesis is small, and hence the excess is significant. By converting p_0 into standard deviations of

a Gaussian distribution using the normal inverse cumulative distribution function [120], the observed significance becomes

$$Z_0 = \mathbf{\Phi}^{-1}(1 - p_0) = \sqrt{q_0}. \tag{13.7}$$

The measured signal strength is given by $\hat{\mu} \pm \sigma_\mu$, where σ_μ is obtained from shifting q_μ by one unit of its variance. Individual NPs can be ranked depending on their impact on the posterior uncertainty on μ. The ranking is performed by fixing each NP to its best fit value, $\hat{\theta}$, and varying a single NP by one unit of its fitted uncertainty. Then the estimated impact of the systematic uncertainty is quantified by the difference between $\hat{\mu}$ and the varied μ value, $\Delta\mu$. The higher the $\Delta\mu$ value for a given NP, the higher the impact of the NP, and hence its position in the ranking. The impact of a group of systematic uncertainties can also be evaluated and presented in a breakdown table, where each entry is calculated from the difference in the quadrature of the error on μ from a fit with all NPs and the error on μ from a fit excluding the NPs of interest.

Expected 'prefit' results involve an Asimov dataset, which is a representative dataset where the best fit values of μ and $\boldsymbol{\theta}$ are set to their generated values (i.e. $\mu = 1$) and the expected event yields from simulation are used. The expected 'postfit' results are derived by applying the fit procedure to an Asimov dataset. Distributions obtained from the expected 'postfit' results are used to ascertain the agreement between data and MC.

The nominal fit uses the BDT_{VH} as the distribution input to the fit in the signal regions. This is referred to as the VH MVA analysis. To cross-check and validate the VH MVA analysis, two other versions of the analysis are constructed. As m_{bb} is the most important variable in the fit, it is used to check the VH MVA results. The VZ process contains a $b\bar{b}$ resonance, hence it is used to validate the analysis process. The BDT_{VZ} discriminant and the m_{bb} distribution are used as the input distribution in the signal region for the VZ MVA fit and the di-jet mass fit, respectively, giving rise to the corresponding diboson validation analysis and di-jet mass cross-check analysis. As the fitting procedure produces a large amount of information, the results of the studies in this chapter are separated into pulls comparison, rankings, breakdowns, and significances.

As $t\bar{t}$ is one of the dominant backgrounds and the uncertainties associated with it are

some of the leading systematics, it is crucial that the uncertainties are correctly treated with the optimal correlation scheme in the fit, since the pull and constraint of a correlated NP are propagated across all regions. Hence, it is important that the propagation of the pull and constraint of a NP is justified, otherwise the NP should be decorrelated into its components corresponding to different regions, which are represented by suffixes in the NP names. The impact of the different correlation schemes on $t\bar{t}$ uncertainties are presented in this chapter.

13.1 Understanding the $t\bar{t}$ p_T^V shape uncertainty

To study the $t\bar{t}$ p_T^V shape uncertainty in the 0+1-lepton channels, which is derived from shape variations between different simulated MC samples, the effect of decorrelating this uncertainty between the jet categories and the signal and control regions was investigated in the 0-lepton, 1-lepton and combined (0-, 1- and 2-lepton) VH MVA fits, but as similar conclusions were drawn from the results from all fits for each type of decorrelation, only the 1-lepton MVA fit results are presented in this section. The uncertainty, correlated across all regions, is denoted by the nuisance parameter, `TTbarPtV_BDTr`.

13.1.1 Decorrelation between jet categories

The uncertainty was decorrelated into its 2-jet and 3-jet components, creating two nuisance parameters, `TTbarPtV_BDTr_J2` and `TTbarPtV_BDTr_J3`.

- **Pulls comparison:** It can be observed from Fig. 13.1 that the correlated NP (`TTbarPtV_BDTr`) is significantly constrained and pulled upwards. The 3-jet component is also pulled upwards and constrained whereas the 2-jet component is not affected, showing that the pull and constraint of `TTbarPtV_BDTr` originates in the 3-jet region only, which is expected as the 3-jet region is dominated by $t\bar{t}$.
- **Rankings:** Table 13.1 reveals that the 3-jet component ranked higher than the 2-jet component in the decorrelated case, but both ranks are very similar suggesting that they have a similar impact on the signal strength. In the correlated case, the p_T^V shape uncertainty ranks lower than the separated components, suggesting that the impact of the uncertainty on the signal strength is reduced from correlating.

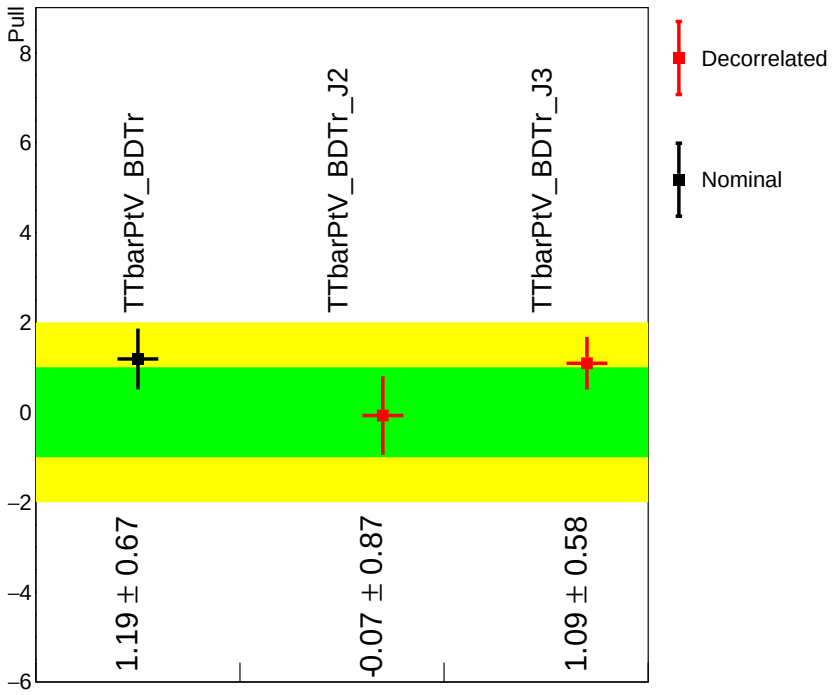


Figure 13.1: Fitted nuisance parameters for the $t\bar{t}$ p_T^V shape uncertainty from correlating (black) and decorrelating (red) the uncertainty between jet categories. The number of σ the central value of each nuisance parameter has shifted by in the fit is shown by the pull (y-axis) and the relative constraint on the original prior is shown by the size of the uncertainty on the data point. The exact values of the pull and constraint for each nuisance parameter are indicated along the x-axis. The green and yellow bands show that the NP is within $\pm 1\sigma$ and $\pm 2\sigma$ of its central value, respectively.

Nuisance parameter	Ranking
Correlated	
`TTbarPtV_BDTr`	36
Decorrelated	
`TTbarPtV_BDTr_J2`	24
`TTbarPtV_BDTr_J3`	22

Table 13.1: Rankings of the NPs from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between jet categories based on their impact on the signal strength.

- **Breakdowns:** The total impact of the uncertainties on the signal strength is increased slightly by the decorrelation and the $t\bar{t}$ contribution increases from 0.049 to 0.061, as presented in Table 13.2. The W+jets contribution also increases slightly, which is likely due to a new correlation between the 3-jet components of the W+jets p_T^V shape uncertainty and the $t\bar{t}$ p_T^V shape uncertainty which results from decorrelating.
- **Significances:** Comparison of the expected significances in Table 13.3 revealed

Set of uncertainties	Uncertainty on μ	
	Correlated	Decorrelated
Total	± 0.286	± 0.288
Statistical	± 0.192	± 0.192
Systematic	± 0.212	± 0.215
Floating normalisations	± 0.093	± 0.098
Theoretical and modelling uncertainties		
VH Signal	± 0.054	± 0.054
Multi-jet	± 0.025	± 0.021
Single-top	± 0.057	± 0.056
$t\bar{t}$	± 0.049	± 0.061
W+jets	± 0.089	± 0.096
Z+jets	± 0.003	± 0.002
Diboson	± 0.042	± 0.044
MC statistical	± 0.058	± 0.058

Table 13.2: Breakdown of the sources of uncertainties from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between jet categories.

that the sensitivity in both cases are very similar, although there is a slight decrease of around 1% in the significances from decorrelating.

	Expected significances			
	Prefit	Difference	Postfit	Difference
Correlated	3.776	-	4.022	-
Decorrelated	3.743	-0.87%	4.005	-0.42%

Table 13.3: Expected significances from a fit to an Asimov dataset (prefit) and a conditional fit to data with $\mu = 1$ (postfit) from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between jet categories.

The similarity of the different metrics demonstrates that the results were not very sensitive to the correlation scheme for the $t\bar{t}$ p_T^V shape uncertainty. As there was no strong motivation to propagate the constraint from the 3-jet component to 2-jet component and the contribution of the uncertainties from $t\bar{t}$ was reduced when correlating, it was more conservative and desirable to have the jet components decorrelated, hence the decision was made to decorrelate the uncertainty.

13.1.2 Decorrelation between signal and control regions

The nuisance parameters, `TTbarPtV_BDTr_DSR`, `TTbarPtV_BDTr_DCRHigh` and `TTbarPtV_BDTr_DCRLow` were introduced from decorrelating the uncertainty into its

signal region (SR) and control regions (CRHigh and CRLow) components. The jet components of the uncertainty were kept correlated for this study.

- **Pulls comparison:** Fig. 13.2 illustrates that the SR and CRHigh components pull in the same direction as `TTbarPtV_BDTr`, indicating that they are compatible with each other. The SR and CRHigh components are constrained but this is expected as these regions have large $t\bar{t}$ background components. The CRLow component is not pulled or constrained, suggesting that this region is insensitive to `TTbarPtV_BDTr`.

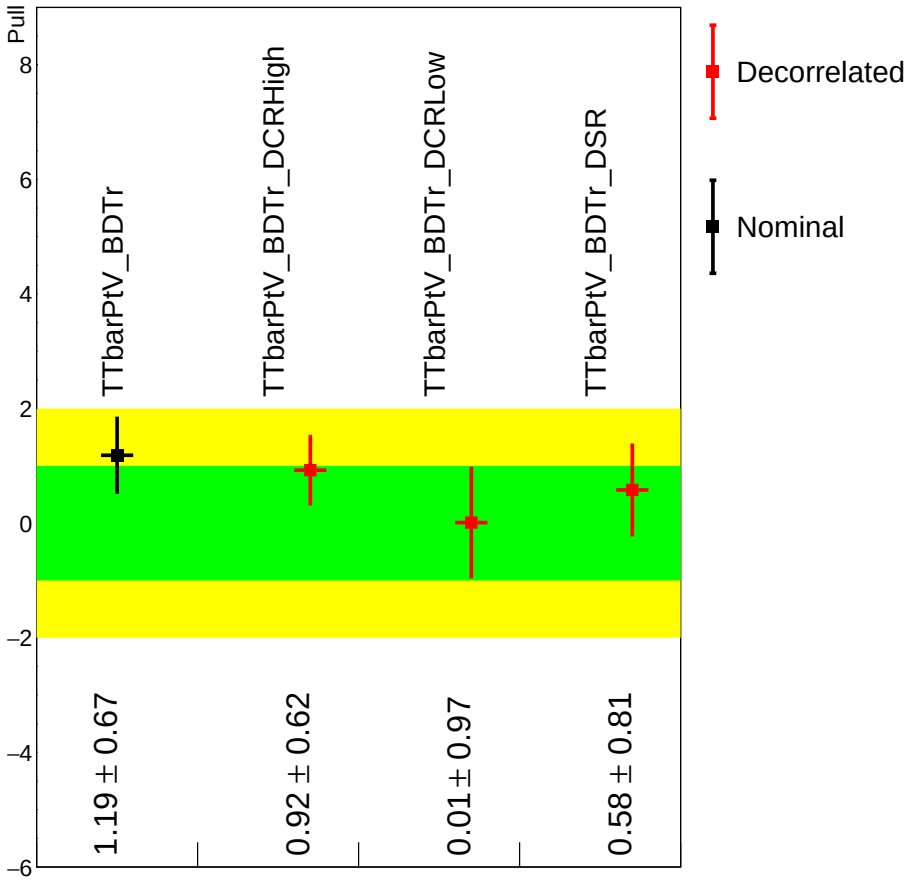


Figure 13.2: Fitted nuisance parameters for the $t\bar{t}$ p_T^V shape uncertainty from correlating (black) and decorrelating (red) the uncertainty between signal (SR) and control (CRHigh and CRLow) regions. The number of σ the central value of the each nuisance parameter has shifted by in the fit is shown by the pull (y-axis) and the relative constraint on the original prior is shown by the size of the uncertainty on the data point. The exact values of the pull and constraint for each nuisance parameter are indicated along the x-axis. The green and yellow bands show that the NP is within $\pm 1\sigma$ and $\pm 2\sigma$ of its central value, respectively.

- **Rankings:** Table 13.4 shows that the SR and CRLow components in the decorrelated case are ranked low, indicating that they have an insignificant impact on

the signal strength. The CRHigh component in the decorrelated case is ranked one higher than the correlated p_T^V shape uncertainty, suggesting that most of the impact of the uncertainty on the signal strength comes from the CRHigh component.

Nuisance parameter	Ranking
Correlated	
`TTbarPtV_BDTr`	36
Decorrelated	
`TTbarPtV_BDTr_DCRHigh`	35
`TTbarPtV_BDTr_DSR`	91
`TTbarPtV_BDTr_DCRLow`	124

Table 13.4: Rankings of the NPs from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between signal and control regions based on their impact on the signal strength.

- **Breakdown:** A breakdown of the sources of uncertainties in Table 13.5 showed that the contributions between the correlated and decorrelated cases were very similar.

Set of uncertainties	Uncertainty on μ	
	Correlated	Decorrelated
Total	± 0.286	± 0.286
Statistical	± 0.192	± 0.192
Systematic	± 0.212	± 0.212
Floating normalisations	± 0.093	± 0.093
Theoretical and modelling uncertainties		
VH Signal	± 0.054	± 0.054
Multi-jet	± 0.025	± 0.025
Single-top	± 0.057	± 0.058
$t\bar{t}$	± 0.049	± 0.049
W+jets	± 0.089	± 0.089
Z+jets	± 0.003	± 0.003
Diboson	± 0.042	± 0.041
MC statistical	± 0.058	± 0.058

Table 13.5: Breakdown of the sources of uncertainties from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between signal and control regions.

- **Significances:** The significances between the correlated and decorrelated cases were consistent, as in Table 13.6.

Since there was good compatibility between the NPs of the two cases and there were trivial changes in the significances, rankings, and breakdown, the simpler approach

	Expected significances			
	Prefit	Difference	Postfit	Difference
Correlated	3.776	-	4.022	-
Decorrelated	3.775	-0.3%	4.022	0%

Table 13.6: Expected significances from a fit to an Asimov dataset (prefit) and a conditional fit to data with $\mu = 1$ (postfit) from correlating and decorrelating the $t\bar{t}$ p_T^V shape uncertainty between signal and control regions.

was taken to keep the $t\bar{t}$ p_T^V shape uncertainty correlated between its SR and CRs components.

13.2 $t\bar{t}$ flavour composition uncertainty

The $t\bar{t}$ background process can be broken down into 6 flavour categories: bb, bc, bl, cc, cl, and ll. As $t \to Wb$ has a branching fraction of approximately 1 [92] and a 70% b-tagging efficiency is employed in the analysis, only the bb and bc components are non-negligible. Hence the $t\bar{t}$ background is classified using a reduced set of flavour categories: bb, bc, Oth, where Oth represents the bl, cc, cl, and ll components. The flavour composition of the nominal $t\bar{t}$ sample is presented in Table 13.7.

Flavour	2-jet		3-jet	
	Number of events	Percentage	Number of events	Percentage
bb	96092	93.3%	374874	91.5%
bc	5798	5.6%	29869	7.3%
Oth	1096	1.1%	5003	1.2%

Table 13.7: Events of the nominal $t\bar{t}$ sample weighted using truth-tagging and decomposed into the $t\bar{t}$ flavour categories.

Flavour composition uncertainties enter the profile likelihood fit as acceptance ratios of the non-bb components of the background processes relative to the dominant bb component. The flavour composition priors from the matrix element (ME) and parton shower (PS) variations are presented in Table 13.8 for each jet category in the 0+1-lepton channels. Initially, the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties were applied in the combined (0-, 1- and 2-lepton) MVA fit as `TTbarbcACC` and `TTbarOthACC` in the 'Total(ME, PS)' systematic scheme, where the NPs correspond to the quadratic sum of the ME and PS variations. Fig. 13.3 shows that neither NP was constrained. While

Systematic uncertainty		Priors			
		0-lepton		1-lepton	
		2-jet	3-jet	2-jet	3-jet
TTbarbcACC	Matrix element	+8.2%	+7.6%	+1.3%	+3.8%
	Parton shower	-2.1%	-3.2%	+7.1%	+1.5%
	Total	±8.5%	±8.2%	±7.2%	±4.1%
TTbarOthACC	Matrix element	-6.36%	-2.8%	+3.3%	+5.7%
	Parton shower	-13.2%	-5.6%	+0.3%	+2.1%
	Total	±14.7%	±6.3%	±3.3%	±6.1%

Table 13.8: Priors of the $t\bar{t}$ acceptance ratios of bc and Oth relative to bb from the matrix element and parton shower variations in each jet category for the 0- and 1-lepton channels. The total is given by the quadratic sum of the variations.

TTbarOthACC exhibited no pull, TTbarbcACC was pulled down.

To further probe the impact of this uncertainty, given the importance of the bc background and the difference in behaviour between the ME, PS and total uncertainties in Table 13.8, a systematic scheme known as 'ME+PS', which involved the uncertainties separated into its ME and PS variations, was studied. Decorrelation of the NPs in both schemes between the 0-lepton (0L) and 1-lepton (1L) channels was also investigated. The systematic and correlation schemes and their associated nuisance parameters are summarised in Table 13.9.

13.2.1 Combined MVA fit

The correlated and decorrelated cases of the systematic schemes were first applied in the combined (0-, 1- and 2-lepton) MVA fit.

- **Pulls comparison:** Fig. 13.3 shows that all components of the $t\bar{t}$ Oth/bb acceptance uncertainty have no significant pulls or constraints. In the decorrelated Total(ME, PS) scheme, the 0L component of (TTbarbcACC) is pulled down whereas the 1L component is not pulled, which suggests that the fit prefers less $t\bar{t}$ bc flavour in 0L compared to the prediction in MC. It can be seen that this 0L pull is dominated by ME because the 0L component of TTbarbcMEACC is pulled down but the 0L component of TTbarbcPSACC is not pulled. There is also tension between the NPs relating to the $t\bar{t}$ bc/bb acceptance uncertainty in the decorrelated ME+PS scheme as the 0L and 1L components of both TTbarbcMEACC and TTbarbcPSACC

Systematic scheme		Nuisance parameters
Total(ME, PS)	Correlated	TTbarbcACC TTbarOthACC
	Decorrelated	TTbarbcACC_L0 TTbarbcACC_L1 TTbarOthACC_L0 TTbarOthACC_L1
ME+PS	Correlated	TTbarbcMEACC TTbarbcPSACC TTbarOthMEACC TTbarOthPSACC
	Decorrelated	TTbarbcMEACC_L0 TTbarbcMEACC_L1 TTbarbcPSACC_L0 TTbarbcPSACC_L1 TTbarOthMEACC_L0 TTbarOthMEACC_L1 TTbarOthPSACC_L0 TTbarOthPSACC_L1

Table 13.9: Nuisance parameters of the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties. The 0L and 1L components are denoted by L0 and L1, respectively.

pull in opposite directions. Though some of the NPs related to the $t\bar{t}$ bc/bb acceptance uncertainty are pulled significantly, all NPs from the correlated and decorrelated cases of both schemes were not constrained, suggesting that the fit is not sensitive to the NPs and was using the pulled components to fix some discrepancies between data and MC.

- **Rankings:** In the correlated and decorrelated cases of both schemes, the NPs corresponding to the $t\bar{t}$ Oth/bb acceptance uncertainty are all ranked very low, as displayed in Table 13.10, indicating that this uncertainty has trivial impact on the results. In the decorrelated cases, the 1L component of `TTbarbcACC` and `TTbarbcMEACC` are all ranked significantly lower than the 0L component, implying that $t\bar{t}$ bc in 0L has a much higher impact on the signal strength. The correlated $t\bar{t}$ bc/bb acceptance uncertainty ranks higher than the separated components, suggesting that the impact of the uncertainty on the signal strength is increased from correlating.

- **Breakdowns:** Breakdowns of the sources of uncertainties in Table 13.11 demon-

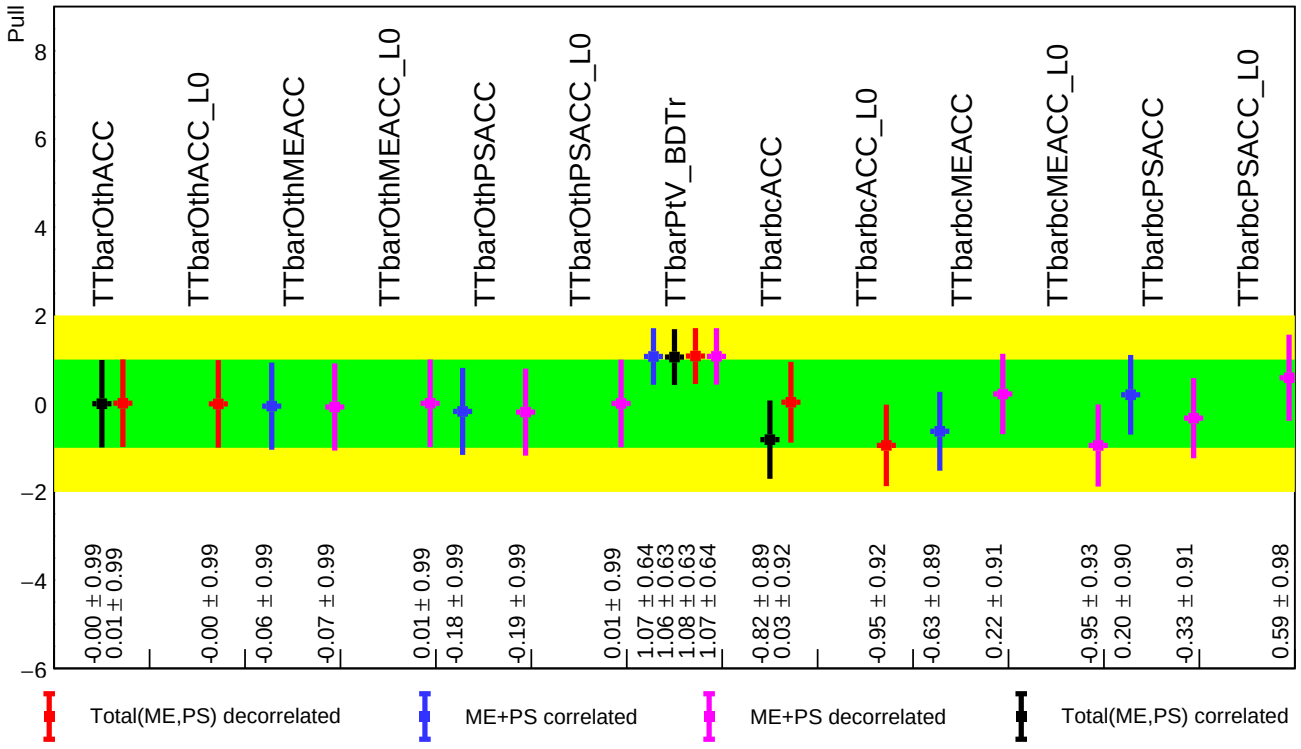


Figure 13.3: Fitted nuisance parameters from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb (`TTbarbcACC`) and $t\bar{t}$ Oth/bb (`TTbarOthACC`) acceptance uncertainties in the combined MVA fit. The separation of the uncertainties into their matrix element and parton shower variations are denoted by the addition of ME and PS, respectively, in the nuisance parameter name. The correlated cases are shown in black and blue, while the decorrelated cases correspond to red and pink. The 0-lepton component is denoted by L0, whereas the 1-lepton component is not explicitly stated. The number of σ the central value of the each nuisance parameter has shifted by in the fit is shown by the pull (y-axis) and the relative constraint on the original prior is shown by the size of the uncertainty on the data point. The exact values of the pull and constraint for each nuisance parameter are indicated along the x-axis. The green and yellow bands show that the NP is within $\pm1\sigma$ and $\pm2\sigma$ of its central value, respectively.

Nuisance parameter		Ranking
Total(ME, PS) scheme		
Correlated	TTbarbcACC	16
	TTbarOthACC	290
Decorrelated	TTbarbcACC_L0	20
	TTbarbcACC_L1	60
	TTbarOthACC_L0	274
	TTbarOthACC_L1	416
ME+PS scheme		
Correlated	TTbarbcMEACC	25
	TTbarbcPSACC	163
	TTbarOthMEACC	210
	TTbarOthPSACC	326
Decorrelated	TTbarbcMEACC_L0	20
	TTbarbcMEACC_L1	273
	TTbarbcPSACC_L0	77
	TTbarbcPSACC_L1	41
	TTbarOthMEACC_L0	422
	TTbarOthMEACC_L1	243
	TTbarOthPSACC_L0	354
	TTbarOthPSACC_L1	238

Table 13.10: Rankings of the NPs from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and $t\bar{t}$ Oth/bb acceptance uncertainties based on their impact on the signal strength in the combined MVA fit. The 0L and 1L components are denoted by L0 and L1, respectively.

strated that the contributions from different groups of uncertainties are very similar between the correlated and decorrelated cases of both schemes.

- **Significances:** Comparison of the expected significances between both schemes and from a nominal fit without the uncertainties in Table 13.12 illustrates that the significances are very similar.

Due to the tension demonstrated by the NPs related to the $t\bar{t}$ bc/bb acceptance uncertainty in the decorrelated ME+PS scheme and the pulls exhibited by the 0L components, further studies were carried out to explore the behaviour of this uncertainty in the combined di-jet mass fit. As the combined MVA fit revealed that the $t\bar{t}$ Oth/bb acceptance uncertainty has little impact on the results, which can be understood by the size of the $t\bar{t}$ Oth contribution, only NPs related to the $t\bar{t}$ bc/bb acceptance uncertainty will be discussed in the combined di-jet mass fit.

Set of uncertainties	Uncertainty on μ			
	Total(ME, PS)		ME+PS	
	Correlated	Decorrelated	Correlated	Decorrelated
Total	±0.177	±0.177	±0.178	±0.178
Statistical	±0.115	±0.115	±0.115	±0.115
Systematic	±0.135	±0.134	±0.136	±0.136
Floating normalisations	±0.038	±0.038	±0.037	±0.036
Theoretical and modelling uncertainties				
VH Signal	±0.055	±0.055	±0.059	±0.059
Multi-jet	±0.008	±0.008	±0.008	±0.008
Single-top	±0.019	±0.020	±0.022	±0.021
$t\bar{t}$	±0.024	±0.023	±0.022	±0.023
W+jets	±0.035	±0.036	±0.037	±0.037
Z+jets	±0.016	±0.016	±0.016	±0.016
Diboson	±0.040	±0.040	±0.040	±0.040
MC statistical	±0.028	±0.028	±0.028	±0.028

Table 13.11: Breakdown of the sources of uncertainties from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and $t\bar{t}$ Oth/bb acceptance uncertainties in the combined MVA fit.

Systematic scheme		Expected significances			
		Prefit	Difference	Postfit	Difference
Nominal		6.932	-	6.749	-
Total(ME, PS)	Correlated	6.910	-0.31%	6.751	-0.03%
	Decorrelated	6.915	-0.25%	6.769	-0.30%
ME+PS	Correlated	6.918	-0.20%	6.752	0.04%
	Decorrelated	6.905	-0.38%	6.760	0.16%

Table 13.12: Expected significances from a fit to an Asimov dataset (prefit) and a conditional fit to data with $\mu = 1$ (postfit) from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties in the combined MVA fit. The significances are compared to those of a nominal fit without the acceptance uncertainties.

13.2.2 Combined di-jet mass fit

As di-jet mass is the most discriminating variable that contributes to the output VH MVA discriminant, it is likely that uncertainties with the most impact on the signal strength in the di-jet mass fit will have some effect on the signal strength in the combined MVA fit. Hence, the correlated and decorrelated cases of both schemes were studied in the combined (0-, 1-, and 2-lepton) di-jet mass fit.

- **Rankings:** The correlated $t\bar{t}$ bc/bb acceptance uncertainty ranked second, and the 0L and ME components of the uncertainty are all ranked very highly, with the 0L

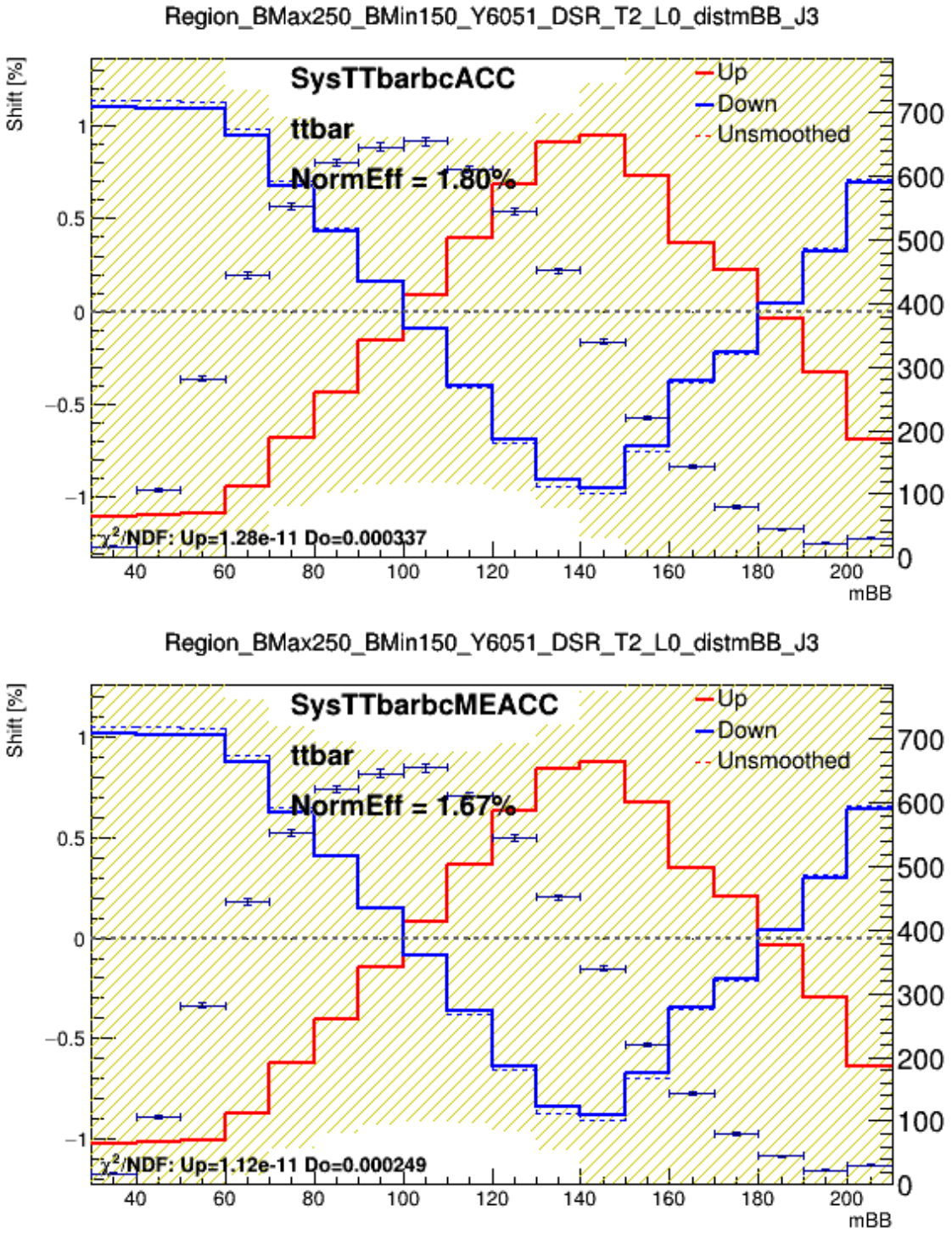


Figure 13.4: Shapes of the $t\bar{t}$ bc/bb acceptance uncertainty in the Total(ME,PS) scheme (top) and the ME component in the ME+PS scheme (bottom) for the 3-jet category in the 0-lepton channel of the combined di-jet mass fit for 150 GeV $< p_T^V <$ 250 GeV. The yellow shaded area is the uncertainty. The up and down variations of the uncertainty are the red and blue solid lines, respectively, and are quantified by the y-axis on the left. The di-jet mass signal is denoted by the blue markers and is quantified by the y-axis on the right. The impact of the uncertainty on the normalisation is given by "NormEff".

components ranking second, suggesting that most of the impact of the uncertainty comes from these components. This is evident from the shape variations of `TTbarbcACC` and `TTbarbcMEACC` in Fig. 13.4, which are identical and very signal-like, indicating that uncertainty has a substantial impact on the signal strength in this fit, in agreement with the rankings. The 1L component of the ME variation of the uncertainty is also ranked much lower than the 0L component, which likely caused the correlated `TTbarbcMEACC` to rank lower than its 0L component and reduced the impact of the ME variation on the signal strength slightly by correlating.

Nuisance parameter		Ranking
Total(ME, PS) scheme		
Correlated	`TTbarbcACC`	2
Decorrelated	`TTbarbcACC_L0`	2
	`TTbarbcACC_L1`	35
ME+PS scheme		
Correlated	`TTbarbcMEACC`	5
	`TTbarbcPSACC`	99
Decorrelated	`TTbarbcMEACC_L0`	2
	`TTbarbcMEACC_L1`	308
	`TTbarbcPSACC_L0`	50
	`TTbarbcPSACC_L1`	25

Table 13.13: Rankings of the NPs from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb acceptance uncertainty based on their impact on the signal strength in the combined di-jet mass fit. The 0L and 1L components are denoted by L0 and L1, respectively.

- **Breakdowns:** Breakdowns of the sources of uncertainties in Table 13.14 showed that the contributions from different groups of uncertainties between the correlated and decorrelated cases of both schemes are very similar. However, the correlated Total(ME, PS) scheme has the largest $t\bar{t}$ contribution while the correlated ME+PS scheme has the lowest $t\bar{t}$ contribution, indicating that the impact of $t\bar{t}$ on the error of the fit is reduced slightly by separating the $t\bar{t}$ bc/bb acceptance uncertainty into its ME and PS variations.
- **Significances:** Table 13.15 illustrates that the differences in the significances are negligible between the correlated and decorrelated cases of both schemes.

Set of uncertainties	Uncertainty on μ			
	Total(ME, PS)		ME+PS	
	Correlated	Decorrelated	Correlated	Decorrelated
Total	±0.221	±0.221	±0.220	±0.221
Statistical	±0.157	±0.157	±0.157	±0.157
Systematic	±0.152	±0.156	±0.154	±0.155
Floating normalisations	±0.027	±0.026	±0.028	±0.025
Theoretical and modelling uncertainties				
VH Signal	±0.057	±0.056	±0.058	±0.058
Multi-jet	±0.008	±0.010	±0.011	±0.010
Single-top	±0.034	±0.038	±0.040	±0.039
$t\bar{t}$	±0.040	±0.036	±0.031	±0.036
$Wjets$	±0.025	±0.026	±0.027	±0.025
$Zjets$	±0.030	±0.029	±0.027	±0.028
Diboson	±0.062	±0.062	±0.062	±0.061
MC statistical	±0.041	±0.040	±0.040	±0.040

Table 13.14: Breakdown of the sources of uncertainties from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties in the di-jet mass fit.

Systematic scheme		Expected significances			
		Prefit	Difference	Postfit	Difference
Nominal		5.164	-	5.091	-
Total(ME, PS)	Correlated	5.082	-1.59%	5.119	0.55%
	Decorrelated	5.098	-1.28%	5.096	0.10%
ME+PS	Correlated	5.119	-0.87%	5.074	-0.39%
	Decorrelated	5.096	-1.31%	5.087	-0.14%

Table 13.15: Expected significances from a fit to an Asimov dataset (prefit) and a conditional fit to data with $\mu = 1$ (postfit) from the correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties in the combined di-jet mass fit. The significances are compared to those of a nominal fit without the acceptance uncertainties.

13.2.3 Summary

The correlated and decorrelated 0L/1L cases of the systematic schemes of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties (Table 13.9) all yielded similar breakdowns and significances, demonstrating that the results are not sensitive to the correlation cases and systematic schemes of these uncertainties.

However, from the priors and the pulls comparison of the $t\bar{t}$ bc/bb and Oth/bb acceptance uncertainties, it is clear that the 0L component of the $t\bar{t}$ bc/bb acceptance

uncertainty is dominated by its ME variation, while its PS variation is slightly more dominant in the 1L component. In addition, the rankings of all fits revealed that most of the impact of the $t\bar{t}$ bc/bb acceptance uncertainty comes from its ME variation in 0L. It is evident from pulls comparison, priors, and rankings, that the ME and PS variations should be separated.

In the VH MVA fit, the differences between the correlated and uncorrelated cases of the ME+PS scheme are not obviously clear, hence the di-jet mass fit was used to decide on a correlation case as it is the most discriminating variable in the VH MVA discriminant. As the shape variations `TTbarbcACC` and `TTbarbcMEACC` in Fig. 13.4 are very signal-like in the di-jet mass fit, it is likely to cause a larger uncertainty in the fit, which is reflected in the rankings. As the behaviour of the correlated/decorrelated NPs in the fit and the overall impact of the different correlation schemes on the expected significances are similar, the NPs of the ME+PS scheme were kept correlated to reduce the number of NPs in the fit, and therefore improve the stability.

Chapter 14

Results

The results from the profile likelihood fit of the $VH, H \to b\bar{b}$ analysis conducted using the full Run-2 dataset is presented in this chapter.

14.1 Signal strength measurements

For the fit with one parameter of interest, related to the VH signal strength, Table 14.1 presents the postfit normalisation factors for the unconstrained backgrounds in the profile likelihood fit, while the postfit signal and background yields for the 0-, 1- and 2-lepton channels are shown in Tables 14.2, 14.3 and 14.4, respectively. The postfit BDT_{VH} output distributions illustrated in Fig. 14.1 for the high p_T^V signal regions in the 2-jet category, which have the highest sensitivity, demonstrates a good agreement between data and simulated predictions.

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 GeV $< p_T^V <$ 150 GeV	1.28 ± 0.08
Z+hf 3-jet, 75 GeV $< p_T^V <$ 150 GeV	1.17 ± 0.05
Z+hf 2-jet, $p_T^V >$ 150 GeV	1.16 ± 0.07
Z+hf 3-jet, $p_T^V >$ 150 GeV	1.09 ± 0.04

Table 14.1: The postfit normalisation factors applied to the unconstrained backgrounds, $t\bar{t}$, W+hf, and Z+hf [17]. The errors are composed of the combined statistical and systematic uncertainties.

The measured VH signal strength for a Higgs boson with a mass of 125 GeV from a

Signal region	150 GeV $< p_T^V <$ 250 GeV		$p_T^V >$ 250 GeV	
Sample	2-jet	3-jet	2-jet	3-jet
Z+jets	2846 ± 80	3830 ± 160	338 ± 13	533 ± 23
W+jets	634 ± 63	1500 ± 110	83 ± 9	220 ± 19
Single-top	237 ± 35	770 ± 130	9 ± 2	36 ± 8
$t\bar{t}$	1157 ± 76	5470 ± 220	39 ± 5	151 ± 16
Diboson	360 ± 55	333 ± 79	86 ± 13	70 ± 17
Total bkg.	5234 ± 63	11910 ± 100	554 ± 15	1011 ± 21
Signal ($\mu = 1.02$)	147 ± 24	130 ± 22	40 ± 6	33 ± 6
Data	5397	11875	578	1046

Table 14.2: The postfit signal and background yields for each signal region in the 0-lepton channel [17]. The yields after the full event selection are normalised to the results of the profile likelihood fit. The errors include all systematic uncertainties.

Signal region	150 GeV $< p_T^V <$ 250 GeV		$p_T^V >$ 250 GeV	
Sample	2-jet	3-jet	2-jet	3-jet
Z+jets	102 ± 5	207 ± 11	10 ± 1	25 ± 2
W+jets	1850 ± 160	4080 ± 270	353 ± 33	935 ± 69
Single-top	990 ± 160	3570 ± 600	70 ± 15	313 ± 66
$t\bar{t}$	4600 ± 210	21030 ± 620	188 ± 19	970 ± 58
Diboson	229 ± 57	264 ± 83	57 ± 15	68 ± 22
Multi-jet	130 ± 47	24 ± 14	16 ± 10	7 ± 8
Total bkg.	7899 ± 82	29170 ± 170	694 ± 23	2318 ± 42
Signal ($\mu = 1.02$)	148 ± 24	125 ± 21	52 ± 9	44 ± 7
Data	8044	29316	727	2378

Table 14.3: The postfit signal and background yields for each signal region in the 1-lepton channel [17]. The yields after the full event selection are normalised to the results of the profile likelihood fit. The uncertainties include all systematic uncertainties.

Signal region	75 GeV $< p_T^V <$ 150 GeV		150 GeV $< p_T^V <$ 250 GeV		$p_T^V >$ 250 GeV	
Sample	2-jet	$\geq$ 3-jet	2-jet	$\geq$ 3-jet	2-jet	$\geq$ 3-jet
Z+jets	5900 ± 100	11630 ± 160	716 ± 19	2499 ± 52	84 ± 3	537 ± 16
W+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 ± 12	17.6 ± 3
Data	9463	20927	881	3148	123	614

Table 14.4: The postfit signal and background yields for each signal region in the 2-lepton channel [17]. The yields after the full event selection are normalised to the results of the profile likelihood fit. The uncertainties include all systematic uncertainties.

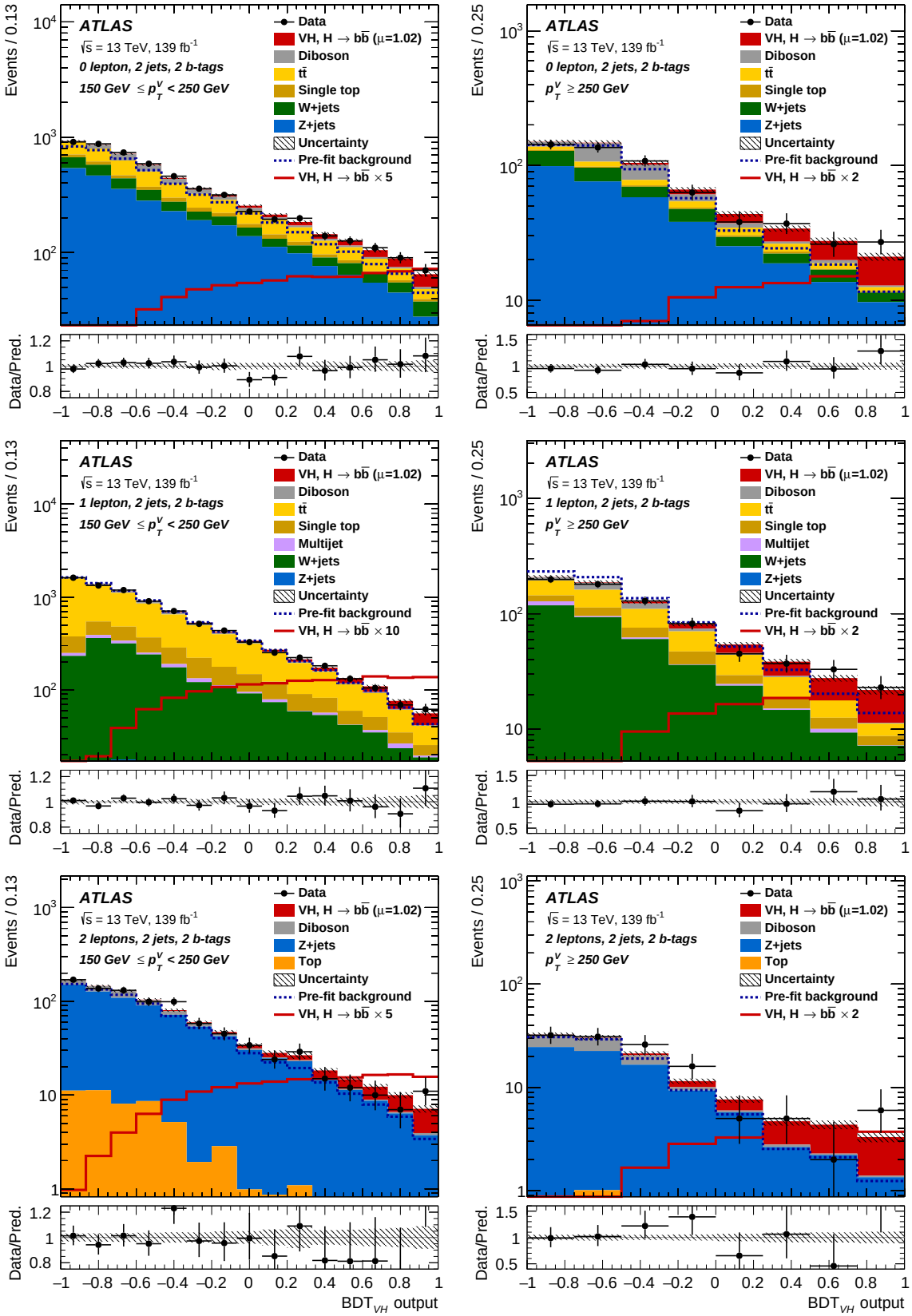


Figure 14.1: The postfit BDT_{VH} output distributions of 2-b-tag events in the 2-jet category for the $150 < p_T^V < 250$ GeV (left) and $p_T^V > 250$ GeV (right) regions in the 0- (top), 1- (middle) and 2- (bottom) lepton channels [17]. The filled histograms represent the fitted signal and background distributions. The $VH \to b\bar{b}$ signal is also shown as an unstacked histogram that is scaled by the factor indicated in the legend.

combined (0-, 1-, and 2-lepton) fit is

$$\mu_{VH}^{bb} = 1.02^{+0.18}_{-0.17} = 1.02^{+0.12}_{-0.11}(\text{stat.})^{+0.14}_{-0.13}(\text{syst.}). \tag{14.1}$$

The measured VH signal strength indicates that the experimental results are consistent with the SM prediction. The background-only hypothesis is rejected with an observed signal significance of 6.7 standard deviations (σ), compared to an expected significance of 6.7σ.

A combined fit including separate signal strength measurements for the WH and ZH production processes is also performed. The measured signal strength for the WH production process is

$$\mu_{WH}^{bb} = 0.95^{+0.27}_{-0.25} = 0.95 \pm 0.18(\text{stat.})^{+0.19}_{-0.18}(\text{syst.}), \tag{14.2}$$

while the fitted signal strength for the ZH production process is

$$\mu_{ZH}^{bb} = 1.08^{+0.25}_{-0.23} = 1.08 \pm 0.17(\text{stat.})^{+0.18}_{-0.15}(\text{syst.}). \tag{14.3}$$

The observed significances of 4.0σ and 5.3σ for the WH and ZH processes, respectively, are compared to their corresponding expected significances of 4.1σ and 5.3σ. The signal strengths are displayed in Fig. 14.2, and are compatible with the SM predictions.

Breakdowns of the different contributions to the uncertainties on the measured signal strengths of VH, WH, and ZH are listed in Table 14.5. For the WH and ZH processes, the contributions from the total statistical and systematic uncertainties are similar. The dominant contributions to the uncertainties on the measured signal strengths originate from the systematic uncertainties on b-tagging, jets, E_T^{miss} and modelling.

14.1.1 Di-jet mass cross-check analysis

As di-jet mass is the most important input variable in the MVA classifier, it is used to cross-check the VH fit, which is nominally obtained using the VH MVA distribution. Fig 14.3 shows the postfit m_{bb} output distributions for the 1-lepton channel, and indicates that the simulated predictions are consistent with data.

Source of uncertainty		σ_μ		
		VH	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	0.064
	c-jets	0.035	0.068	0.010
	light-flavour jets	0.009	0.004	0.014
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

Table 14.5: Breakdown of the contributions to the uncertainty on the measured signal strengths for VH, WH and ZH [17].

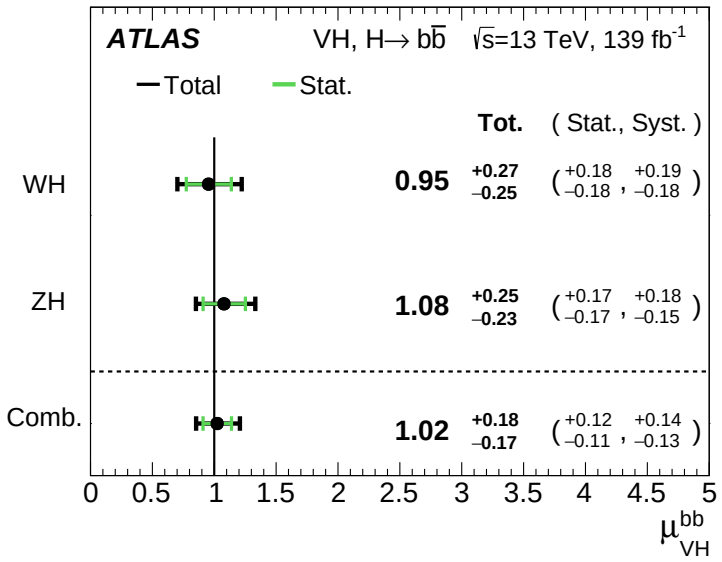


Figure 14.2: Measured signal strengths of the WH and ZH production processes, and their combination [17]. The signal strengths of the WH and ZH are independently unconstrained in a fit across all lepton channels to simultaneously extract their values.

The fitted signal strength from a combined fit to the di-jet mass distribution is

$$\mu_{VH}^{bb} = 1.17^{+0.25}_{-0.23} = 1.17 \pm 0.16(\text{stat.})^{+0.19}_{-0.16}(\text{syst.}). \tag{14.4}$$

The observed significance is 5.5 σ, compared to an expected significance of 4.9 σ. The compatibility between the nominal fit and the di-jet mass fit was checked using the bootstrap method [122], which showed that the fits were in agreement at the level of 1.1 σ.

The di-jet mass distribution, after subtraction of all backgrounds except for the WZ and ZZ processes, is shown in Fig. 14.4. The distribution is obtained by summing the di-jet mass contributions from each p_T^V region in each jet category across all lepton channels, with each contribution weighted by the ratio of the fitted Higgs boson signal to the background yield in that region.

14.1.2 Diboson cross-check analysis

Since the $VZ, Z \to b\bar{b}$ process contains a bb resonance, it is used as a standard reference to validate the VH fit. The postfit BDT_{VZ} output distributions for the 1-lepton channel are shown in Fig. 14.5, which illustrate a good agreement between data and simulated

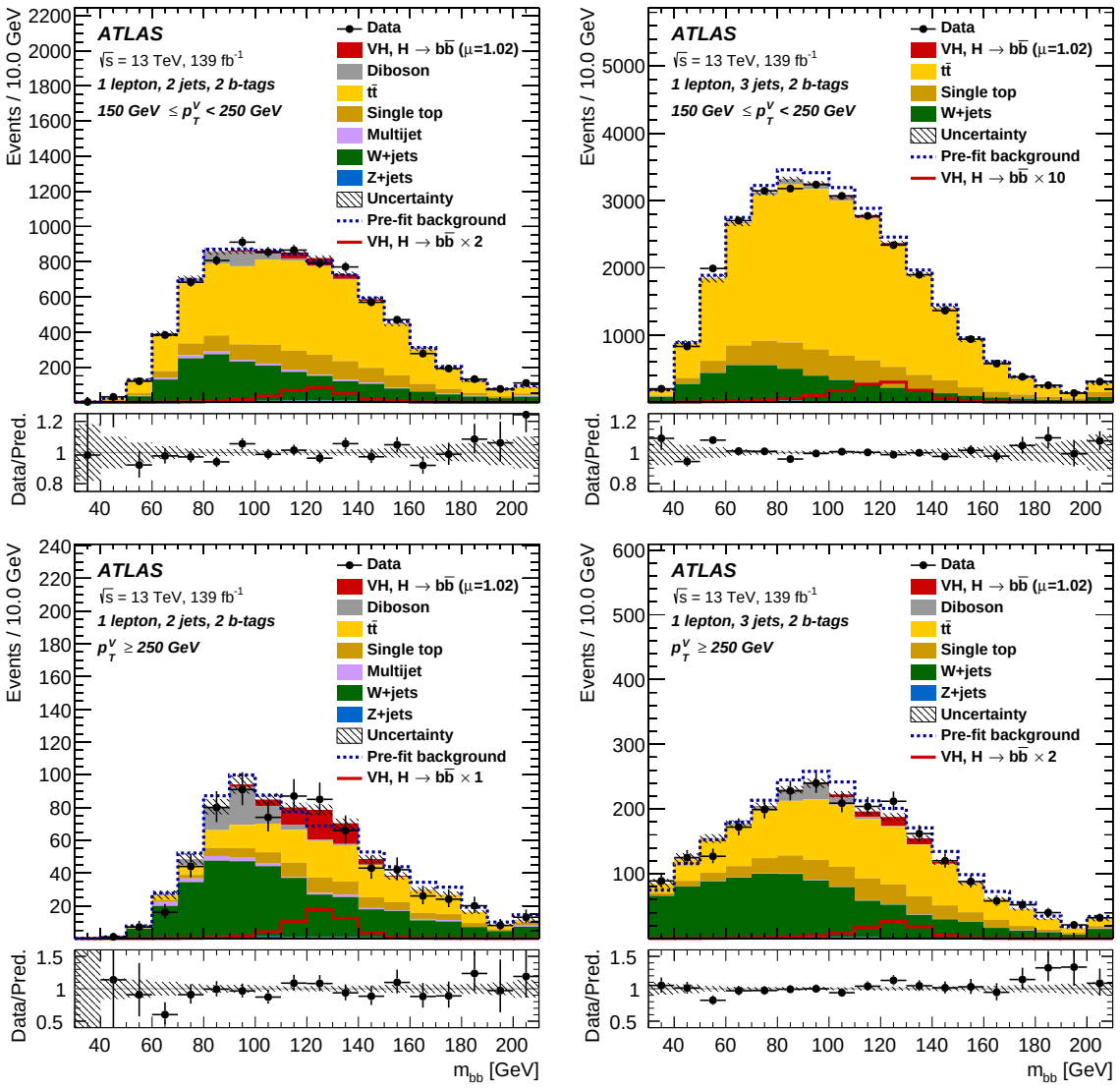


Figure 14.3: The postfit m_{bb} output distributions of 2-b-tag events in the 2-jet (left) and 3-jet (right) categories for the $150 < p_T^V < 250$ GeV (top) and $p_T^V > 250$ GeV (bottom) regions in the 1-lepton channel [121]. The filled histograms represent the fitted signal and background distributions. The $VH \to b\bar{b}$ signal is also shown as an unstacked histogram that is scaled by the factor indicated in the legend.

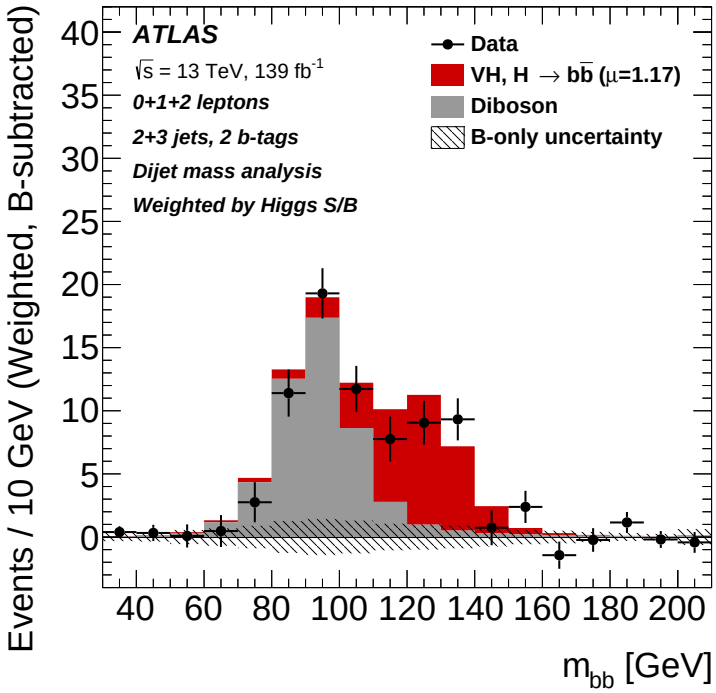


Figure 14.4: The di-jet mass distribution after subtraction of all backgrounds except for the WZ and ZZ diboson processes, weighted by the ratio of the fitted Higgs boson signal (S) to background (B) yields [17]. The expected $VH, H \to bb$ signal is scaled by the measured signal strength ($\mu = 1.17$). The shaded area indicates the total uncertainty on the fitted background.

predictions.

The measured signal strength from a combined fit to BDT_{VZ} is

$$\mu_{VZ}^{bb} = 0.93^{+0.16}_{-0.13} = 0.93^{+0.07}_{-0.06}(\text{stat.})^{+0.14}_{-0.12}(\text{syst.}), \tag{14.5}$$

which demonstrates a good agreement between experiment and SM predictions.

Similarly to the nominal analysis, a combined fit in which signal strengths are measured independently for WZ and ZZ was conducted. The fitted signal strengths are illustrated in Fig. 14.6.

14.2 Simplified template cross section measurements

For the STXS (Section 10.5) measurements, the likelihood fit is performed simultaneously in all STXS regions, as each region has its own signal template. The fitting procedure is similar to those performed to provide the results in the previous section except the

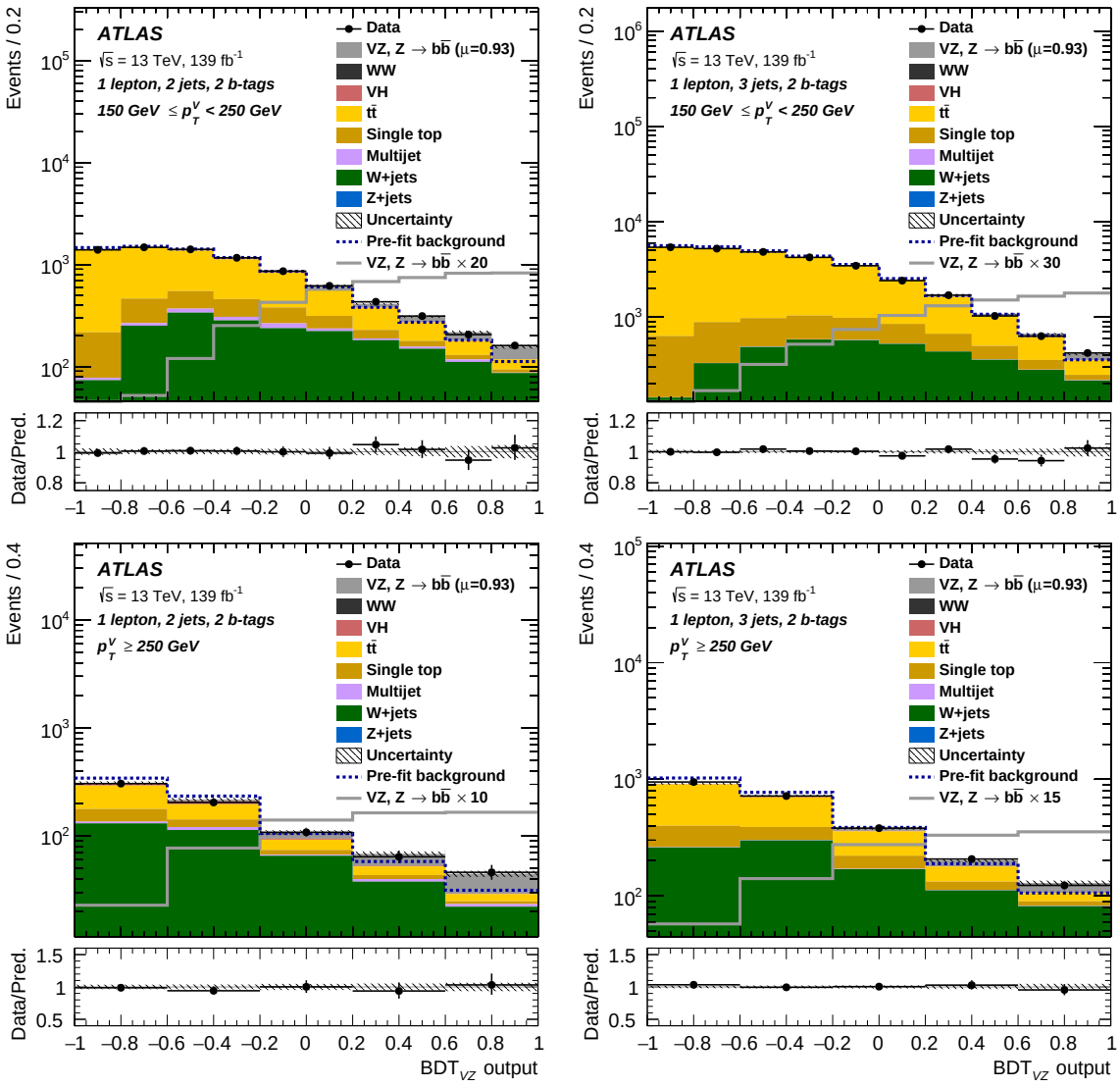


Figure 14.5: The postfit BDT_{VZ} output distributions of 2-b-tag events in the 2-jet (left) and 3-jet (right) categories for the $150 < p_T^V < 250$ GeV (top) and $p_T^V > 250$ GeV (bottom) regions in the 1-lepton channel [121]. The filled histograms represent the fitted signal and background distributions. The $VZ \to b\bar{b}$ signal is also shown as an unstacked histogram that is scaled by the factor indicated in the legend.

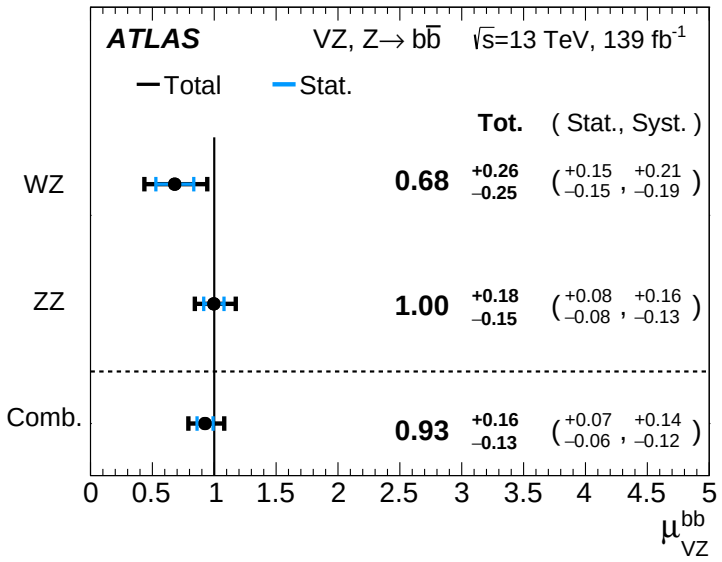


Figure 14.6: Measured signal strengths of the WZ and ZZ production processes, and their combination [17]. The signal strengths of the WZ and ZZ are independently unconstrained in a fit across all lepton channels to simultaneously extract their values.

nuisance parameters relating to the prediction of the production cross-section for each of the STXS regions are not included in the fit. In addition, compared to a fit used to derive the VH signal strength where one parameter of interest, i.e. μ, is extracted, multiple parameters of interest corresponding to the signal strength in each STXS p_T^V region are extracted in the likelihood fit for STXS measurements.

The measurement of the product of the VH cross section and the $H \to b\bar{b}$ and $V \to$ leptons branching ratios, $\sigma \times B$, in the STXS regions is compared to the SM predictions in Table 14.6 and Fig. 14.7, which both indicate that the experimental values are in agreement with theoretical predictions. It can be seen from Table 14.6 that the dominant contribution to the total uncertainty in all regions stems from the data statistical uncertainty. In the lower p_T^V regions, there is also a noticeable contribution to the total uncertainty from systematic uncertainties.

STXS region		SM prediction			Result			Stat. unc.	Syst. unc. [fb]							
Process	$p_{\mathrm{T}}^{V,\,\mathrm{t}}$ interval	[fb]			[fb]			[fb]	Th. sig.		Th. bkg.		Exp.			
$W(\ell\nu)H$	150–250 GeV	24.0	±	1.1	19.0	±	12.1	±	7.7	±	0.9	±	5.5	±	6.0	
$W(\ell\nu)H$	$>$ 250 GeV	7.1	±	0.3	7.2	±	2.2	±	1.9	±	0.4	±	0.8	±	0.7	
$Z(\ell\ell/\nu\nu)H$	75–150 GeV	50.6	±	4.1	42.5	±	35.9	±	25.3	±	5.6	±	17.2	±	19.7	
$Z(\ell\ell/\nu\nu)H$	150–250 GeV	18.8	±	2.4	20.5	±	6.2	±	5.0	±	2.3	±	2.4	±	2.3	
$Z(\ell\ell/\nu\nu)H$	$>$250 GeV	4.9	±	0.5	5.4	±	1.7	±	1.5	±	0.5	±	0.5	±	0.3	

Table 14.6: The product of the VH cross section and the $H \to b\bar{b}$ and $V \to$ leptons branching ratios, $\sigma \times B$, from experimental results and Standard Model predictions in the STXS regions [17]. The contributions to the total uncertainty in the measurements are from statistical (Stat. unc.) and systematic uncertainties (Syst. unc.) from signal (Th. sig.) and background (Th. bkg.) predictions and experimental performance (Exp.).

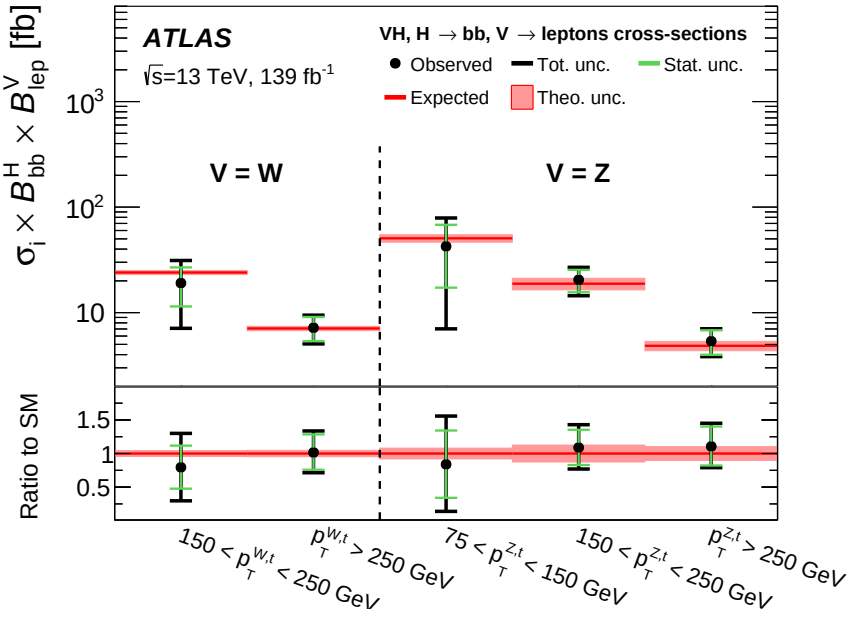


Figure 14.7: The measurement of the product of the VH cross section and the $H \to b\bar{b}$ and $V \to$ leptons branching ratios, $\sigma \times B$, in the STXS regions [17].

Chapter 15

Conclusion

A measurement of the $VH, H \to b\bar{b}$ final state was undertaken using the full Run-2 dataset collected at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, which corresponds to a total integrated luminosity of 139 fb^{-1}.

A key component of the $VH, H \to b\bar{b}$ analysis involves the identification of jets originating from b-quarks (b-tagging). As the reconstruction of tracks directly impact b-tagging, the quality of the track collections is critical to the performance of b-tagging. To ascertain which aspects of the tracking algorithms impacts the flavour-tagging efficiency the most, so future studies can focus on improving these aspects, different track collections were constructed and their impact upon b-tagging were studied. This identified that for the currents algorithms, the largest impact upon b-tagging could be realised by improving the pattern recognition of the tracks not from b-hadron decays, to reduce mis-measurements and thereby reduce the fake-rate. This had a bigger impact than improving the efficiency with which the tracking algorithm reconstructed tracks from b-hadron decays. However, there is the possibility that if the lower-level algorithms were retuned better use could be made of any additional tracks that are reconstructed. Optimisation studies were also conducted on a higher-level b-tagging algorithm, which revealed similar performances to the default algorithm.

The b-jet tagging efficiency, which is one of the dominant uncertainties in the VH analysis, is calibrated to data using di-leptonic top-quark events. Several techniques were explored to automate the calibration software to provide a more robust procedure that saves both time and manpower. This is especially important within large experiments, where contributions of work often rely on inherited code and frequent updates to most aspects of a calibration are undertaken. These can pose common issues in reproducing results, which the automation aims to solve by increasing the ability to reproduce and validate results for all updates and changes to the calibration methods. It was

also demonstrated that this procedure could reproduce the latest calibration results when triggered by updated data or code changes, demonstrating the feasibility of this approach.

A multivariate analysis (MVA) technique involving boosted decision trees in the $VH, H \to b\bar{b}$ analysis was used to classify signal and background events. Studies including optimisation of the hyperparameters and different input variables were conducted in the 1-lepton channel, which yielded significant improvements in the sensitivity of above 10%. These improvements were also adopted in the other lepton channels, resulting in an improvement of 7% and 6% in the sensitivity of the 0- and 2-lepton channels, respectively. The effects of the uncertainties related to one of the dominant backgrounds in the analysis were also tested to ensure the statistical fit is robust.

The observed significance of the $VH, H \to b\bar{b}$ signal is 6.7σ, which corresponds to a measured signal strength of $\mu_{VH}^{bb} = 1.02^{+0.18}_{-0.17} = 1.02^{+0.12}_{-0.11}(\text{stat.})^{+0.14}_{-0.13}(\text{syst.})$. The observed significances and measured signal strengths of the WH and ZH production processes are 4.0σ with $\mu_{WH}^{bb} = 0.95^{+0.27}_{-0.25} = 0.95 \pm 0.18(\text{stat.})^{+0.19}_{-0.18}(\text{syst.})$ and 5.3σ with $\mu_{ZH}^{bb} = 1.08^{+0.25}_{-0.23} = 1.08 \pm 0.17(\text{stat.})^{+0.18}_{-0.15}(\text{syst.})$, respectively. These results all indicate that the experimental result is consistent with the Standard Model prediction. This result is the first single analysis observation of VH production and $H \to b\bar{b}$ decay. In addition, ZH production was observed for the first time and strong evidence for WH production was found. Furthermore, the measurement of the product of the VH cross section and the $H \to b\bar{b}$ and $V \to$ leptons branching ratios in the simplified cross section template regions between experiment and prediction across all regions was conducted, allowing the behaviour of the Higgs boson to be more probed as a function of p_T, which has since been used to place constraints upon Effective Field Theory [123] models [17].

Bibliography

1. The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **716,** 1–29. (2012). arXiv: 1207.7214.

2. The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **716,** 30–61. (2012). arXiv: 1207.7235.

3. Evans, L. & Bryant, P. LHC Machine. *Journal of Instrumentation* **3,** S08001–S08001. (2008).

4. The ATLAS Collaboration. Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **92,** 012006. (2015). arXiv: 1412.2641.

5. The ATLAS Collaboration. Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Physical Review D - Particles, Fields, Gravitation and Cosmology.* (2015). arXiv: 1408.5191.

6. The CMS Collaboration. Measurement of the properties of a Higgs boson in the four-lepton final state. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **89,** 092007. (2014). arXiv: 1312.5353.

7. The ATLAS Collaboration. Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Physical Review D - Particles, Fields, Gravitation and Cosmology* **90,** 112015. (2014). arXiv: 1408.7084.

8. The CMS Collaboration. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *European Physical Journal C* **74,** 1–49. (2014). arXiv: 1407.0558.

9. The ATLAS Collaboration; The 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. *Journal of High Energy Physics.* (2016). arXiv: 1606.02266.

10. The ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **784,** 173–191. (2018). arXiv: 1806.00425.

11. The CMS Collaboration. Observation of $t\bar{t}H$ Production. *Physical Review Letters* **120,** 231801. (2018).

12. CDF Collaboration & DØ Collaboration. Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron. *Physical Review Letters* **109.** (2012). arXiv: 1207.6436.

13. The ATLAS Collaboration. Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector. *Journal of High Energy Physics* **2017,** 1–69. (2017). arXiv: 1708.03299v2 [hep-ex].

14. The CMS Collaboration. Evidence for the Higgs boson decay to a bottom quark–antiquark pair. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **780,** 501–532. (2018). arXiv: 1709.07497.

15. The ATLAS Collaboration. Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **786,** 59–86. (2018). arXiv: 1808.08238 [hep-ex].

16. The CMS Collaboration. Observation of Higgs boson decay to bottom quarks. *Physical Review Letters* **121.** (2018). arXiv: 1808.08242.

17. The ATLAS Collaboration. Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. *The European Physical Journal C* **81,** 1–41. (2021).

18. Halzen, F. & Martin, A. D. *Quark & Leptons: An Introductory Course In Modern Particle Physics* (John Wiley & Sons, 1984).

19. Thomson, M. *Modern Particle Physics* (Cambridge University Press, 2013).

20. Galbraith, David & Burgard, C. *UX: Standard Model of the Standard Model* 2012. https://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/ (2021).

21. Glashow, S. L. Partial-symmetries of weak interactions. *Nuclear Physics* **22,** 579–588. (1961). (2021).

22. Salam, A. & Ward, J. C. Electromagnetic and weak interactions. *Physics Letters* **13,** 168–171. (1964). (2021).

23. Weinberg, S. A Model of Leptons. *Physical Review Letters* **19,** 1264–1266. (1967). (2021).

24. Englert, F. & Brout, R. Broken Symmetry and the Mass of Gauge Vector Mesons. *Physical Review Letters* **13,** 321–323. (1964). (2021).

25. Higgs, P. W. Broken symmetries, massless particles and gauge fields. *Physics Letters* **12,** 132–133. (1964). (2021).

26. Ellis, R. K., Stirling, W. J. & Webber, B. R. *QCD and Collider Physics* (Cambridge University Press, 2003).

27. Butterworth, J. M., Dissertori, G. & Salam, G. P. Hard Processes in Proton-Proton Collisions at the Large Hadron Collider. *Annual Review of Nuclear and Particle Science* **62,** 387–405. (2012). (2021).

28. Murray, W. & Sharma, V. Properties of the Higgs Boson Discovered at the Large Hadron Collider. *Annual Review of Nuclear and Particle Science* **65,** 515–554. (2015).

29. ETH Zurich - Grab Group. *Higgs boson* 2021. https://grab-group.ethz.ch/research/higgs-boson.html (2021).

30. The ATLAS Collaboration. *LHC Higgs Working Group* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG#Latest_plots (2021).

31. The ATLAS Collaboration. Search for Higgs bosons produced via vector-boson fusion and decaying into bottom quark pairs in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector. *Physical Review D* **98,** 052003. (2018). arXiv: 1807.08639.

32. The ATLAS Collaboration. Search for the $H \to b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. *Journal of High Energy Physics* **2015.** (2015). arXiv: 1409.6212v2 [hep-ex].

33. The ATLAS Collaboration. Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *Physical Review D* **97,** 072016. (2018).

34. Palmer, C., Emanuele, R., Tramontano, F. & Pandini, C. *The 15th Workshop of the LHC Higgs Cross Section Working Group* 2018. https://indico.cern.ch/event/740110/contributions/3192764/ (2021).

35. Sjöstrand, T. Monte Carlo Generators. *2006 European School of High-Energy Physics, ESHEP 2006,* 51–73. (2006). arXiv: hep-ph/0611247.

36. Buckley, A. *et al.* General-purpose event generators for LHC physics. *Physics Reports* **504,** 145–233. (2011). (2021).

37. SLAC. *Simulations* 2016. https://theory.slac.stanford.edu/our-research/simulations (2021).

38. Marcastel, F. *CERN's Accelerator Complex* 2013. https://cds.cern.ch/record/1621583 (2020).

39. The ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation* **3,** S08002. (2008).

40. The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation* **3,** S08003. (2008).

41. The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation* **3,** S08004. (2008).

42. The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation* **3,** S08005. (2008).

43. The ATLAS Collaboration. *LuminosityPublicResultsRun2* https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (2022).

44. Pequenao, J. *Computer generated image of the whole ATLAS detector* 2008. https://cds.cern.ch/record/1095924 (2020).

45. The ATLAS Collaboration. Study of the material of the ATLAS inner detector for Run 2 of the LHC. *Journal of Instrumentation* **12,** P12009. (2017). arXiv: 1707.02826.

46. Capeans, M. *et al. Insertable B-Layer* Tech. rep. (CERN, Geneva, 2010). https://cds.cern.ch/record/1291633.

47. Pequenao, J. *Computer Generated image of the ATLAS calorimeter* 2008. https://cds.cern.ch/record/1095927 (2020).

48. Pequenao, J. *Computer generated image of the ATLAS Muons subsystem* 2008. https://cds.cern.ch/record/1095929 (2020).

49. Nedden, M. Z. The LHC Run 2 ATLAS trigger system: Design, performance and plans. *Journal of Instrumentation* **12,** C03024. (2017).

50. Adams, D. *et al. Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014* Tech. rep. (CERN, Geneva, 2015), 12. https://cds.cern.ch/record/1700874.

51. The ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *European Physical Journal C.* (2017).

52. The ATLAS Collaboration. *The Expected Performance of the ATLAS Inner Detector* Tech. rep. (CERN, Geneva, 2009). http://cds.cern.ch/record/1118445.

53. The ATLAS Collaboration. *Early Inner Detector Tracking Performance in the 2015 data at* $\sqrt{s}$ *= 13 TeV* Tech. rep. (CERN, Geneva, 2015). https://cds.cern.ch/record/2110140.

54. The ATLAS Collaboration. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *European Physical Journal C.* (2017).

55. The 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. *European Physical Journal C* **79.** (2019). arXiv: 1902.04655.

56. The ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *European Physical Journal C* **76.** (2016).

57. The ATLAS Collaboration. *RecommendedIsolationWPs* 2020. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/RecommendedIsolationWPs (2020).

58. The ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *European Physical Journal C.* (2017). arXiv: 1603.02934.

59. Cacciari, M., Salam, G. P. & Soyez, G. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics.* (2008). arXiv: 0802.1189.

60. The ATLAS Collaboration. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physical Review D* **96.** (2017). arXiv: 1703.09665.

61. The ATLAS Collaboration. Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector. *European Physical Journal C* **73.** (2013). arXiv: 1210.6210.

62. The ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *European Physical Journal C* **76.** (2015).

63. The ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV. *European Physical Journal C* **78.** (2018). arXiv: 1802.08168.

64. Radovic, A. *et al.* Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560,** 41–48. (2018).

65. Coadou, Y. Boosted Decision Trees and Applications. *EPJ Web of Conferences* **55** (eds Delemontex, T. & Lucotte, A.) 02004. (2013).

66. Statinfer. *Building a Decision Tree in Python* 2017. https://statinfer.com/204-3-7-building-a-decision-tree-in-python/ (2021).

67. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, 2009).

68. Hoecker, A. *et al.* TMVA - Toolkit for Multivariate Data Analysis. (2007). arXiv: physics/0703039.

69. Alves, S. *Deep Learning in PyTorch with CIFAR-10 dataset* 2020. https://medium.com/@sergioalves94/deep-learning-in-pytorch-with-cifar-10-dataset-858b504a6b54 (2021).

70. Kingma, D. P. & Ba, J. L. *Adam: A method for stochastic optimization* in (International Conference on Learning Representations, ICLR, 2015). arXiv: 1412.6980.

71. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323,** 533–536. (1986).

72. Sharma, S. *Activation Functions in Neural Networks* 2017. https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6 (2021).

73. Zyla, P. A. *et al.* Review of Particle Physics. *Progress of Theoretical and Experimental Physics* **2020.** (2020). (2021).

74. Bartosik, N. *b-tagging* http://bartosik.pp.ua/hep_sketches/btagging (2016).

75. The ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s}$=13 TeV. *European Physical Journal C.* (2019). arXiv: 1907.05120.

76. Müller, A. C. & Guido, S. *Introduction to Machine Learning with Python* ("O'Reilly Media, Inc.", 2016).

77. The ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run* Tech. rep. (CERN, Geneva, 2017). https://cds.cern.ch/record/2273281.

78. The ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector* Tech. rep. (CERN, Geneva, 2017). https://cds.cern.ch/record/2270366.

79. Piacquadio, G. & Weiser, C. A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS. *Journal of Physics: Conference Series* **119,** 032032. (2008).

80. The ATLAS Collaboration. *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC* Tech. rep. (CERN, Geneva, 2018). https://cds.cern.ch/record/2645405.

81. The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C* **70,** 823–874. (2010). arXiv: 1005.4568.

82. Agostinelli, S. *et al.* GEANT4 - A simulation toolkit. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506,** 250–303. (2003).

83. Lange, D. J. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **462,** 152–155. (2001).

84. Ball, R. D. *et al.* Parton distributions with LHC data. *Nuclear Physics B* **867,** 244–289. (2013).

85. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX. *Journal of High Energy Physics* **2010,** 43. (2010). arXiv: 1002.2581.

86. Ball, R. D. *et al.* Parton distributions for the LHC run II. *Journal of High Energy Physics* **2015,** 1–148. (2015). arXiv: 1410.8849.

87. Onwujekwegn, G. & Yoon, V. Y. Analyzing the Impacts of Activation Functions on the Performance of Convolutional Neural Network Models. *AMCIS 2020 Proceedings.* (2020). https://aisel.aisnet.org/amcis2020/ai_semantic_for_intelligent_info_systems/ai_semantic_for_intelligent_info_systems/4.

88. Smith, L. N. Cyclical Learning Rates for Training Neural Networks. *IEEE Winter Conference on Applications of Computer Vision, WACV 2017,* 464–472. (2017). arXiv: 1506.01186.

89. Boettiger, C. An introduction to Docker for reproducible research, with examples from the R environment. *ACM SIGOPS Operating Systems Review, Special Issue on Repeatability and Sharing of Experimental Artifacts* **49,** 71–79. (2015). arXiv: 1410.0846.

90. Cortes Rios, J. C. *et al.* A Methodology for Using GitLab for Software Engineering Learning Analytics. *2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE),* 3–6. (2019).

91. Cito, J. & Gall, H. C. Using Docker containers to improve reproducibility in software engineering research. *Proceedings of the 38th International Conference on Software Engineering Companion,* 906–907. (2016).

92. Déliot, F., Hadley, N., Parke, S. & Schwarz, T. Properties of the top quark. *Annual Review of Nuclear and Particle Science.* (2014). arXiv: 1803.00656.

93. Combe, T., Martin, A. & Di Pietro, R. To Docker or Not to Docker: A Security Perspective. *IEEE Cloud Computing* **3,** 54–62. (2016).

94. Kurtzer, G. M., Sochat, V. & Bauer, M. W. Singularity: Scientific containers for mobility of compute. *PLOS ONE* **12** (ed Gursoy, A.) e0177459. (2017).

95. Soltesz, S., Pötzl, H., Fiuczynski, M. E., Bavier, A. & Peterson, L. Container-based operating system virtualization: A scalable, high-performance alternative to hypervisors. *EuroSys '07: Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007,* 275–287. (2007).

96. Pahl, C. & Lee, B. Containers and Clusters for Edge Cloud Architectures – A Technology Review. *2015 3rd International Conference on Future Internet of Things and Cloud,* 379–386. (2015).

97. Ebert, C., Gallardo, G., Hernantes, J. & Serrano, N. DevOps. *IEEE Software* **33,** 94–100. (2016).

98. Ben-Kiki, O. & Evans, C. *YAML Ain't Markup Language (YAML™) Version 1.2 3rd Edition, Patched at 2009-10-01* Tech. rep. (2001). http://yaml.org/spec/1.2/spec.html.

99. CERN. *Worldwide LHC Computing Grid* http://wlcg-public.web.cern.ch/ (2018).

100. Ferrando, J., Geisen, J., Pollard, C. S., Saimpert, M. & Schmoeckel, J. C. *Measurement of the b-jet identification efficiency with $t\bar{t}$ events using an improved likelihood method* Tech. rep. (CERN, Geneva, 2018). https://cds.cern.ch/record/2631547.

101. Hamilton, K., Nason, P. & Zanderighi, G. MINLO: Multi-scale improved NLO. *Journal of High Energy Physics* **2012,** 155. (2012). arXiv: 1206.3572.

102. Sjöstrand, T., Mrenna, S. & Skands, P. A brief introduction to PYTHIA 8.1. *Computer Physics Communications* **178,** 852–867. (2008). arXiv: 0710.3820.

103. Gleisberg, T. *et al.* Event generation with SHERPA 1.1. *Journal of High Energy Physics* **2009,** 007. (2009). arXiv: 0811.4622.

104. White, C. D., Frixione, S., Laenen, E. & Maltoni, F. Isolating Wt production at the LHC. *Journal of High Energy Physics* **2009,** 074. (2009). arXiv: 0908.0631.

105. Berger, N. *et al. Simplified template cross sections - Stage 1.1* Tech. rep. (CERN, Geneva, 2019). arXiv: 1906.02754.

106. Gonçalves, D. & Nakamura, J. Role of the Z polarization in the $H \to b\bar{b}$ measurement. *Physical Review D* **98.** (2018). arXiv: 1805.06385.

107. The ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC* Tech. rep. (CERN, Geneva, 2019). https://cdsweb.cern.ch/record/2677054.

108. Avoni, G. *et al.* The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *Journal of Instrumentation* **13,** P07017. (2018).

109. The ATLAS Collaboration. Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector. *The European Physical Journal C 2022 82:1* **82,** 1–27. (2022).

110. The ATLAS Collaboration. *Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s}$ = 13 TeV* Tech. rep. (CERN, Geneva, 2018). https://cds.cern.ch/record/2314418.

111. Bellm, J. *et al.* Herwig 7.0/Herwig++ 3.0 release note. *European Physical Journal C* **76,** 196. (2016). arXiv: 1512.01178.

112. Butterworth, J. *et al.* PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics* **43,** 023001. (2016). arXiv: 1510.03865.

113. LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. (2011). arXiv: 1101.0593 [hep-ex].

114. LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. (2012). arXiv: 1201.3084 [hep-ex].

115. LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. (2013). arXiv: 1307.1347 [hep-ex].

116. Alwall, J., Herquet, M., Maltoni, F., Mattelaerc, O. & Stelzerd, T. MadGraph 5: Going beyond. *Journal of High Energy Physics* **2011,** 128. (2011). arXiv: 1106.0522.

117. Alwall, J. *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics* **2014,** 79. (2014). arXiv: 1405.0301.

118. Bähr, M. *et al.* Herwig++ physics and manual. *European Physical Journal C* **58,** 639–707. (2008). arXiv: 0803.0883.

119. Barlow, R. & Beeston, C. Fitting using finite Monte Carlo samples. *Computer Physics Communications* **77,** 219–228. (1993).

120. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. Asymptotic formulae for likelihood-based tests of new physics. *European Physical Journal C* **71.** (2011). arXiv: 1007.1727.

121. The ATLAS Collaboration. *Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector* https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/HIGG-2018-51/ (2022).

122. The 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. *Journal of High Energy Physics 2015 2015:1* **2015,** 1–89. (2015).

123. Contino, R., Ghezzi, M., Grojean, C., Mühlleitner, M. & Spira, M. Effective Lagrangian for a light Higgs-like scalar. *Journal of High Energy Physics 2013 2013:7* **2013,** 1–43. (2013).