
Identification of b-jets and investigation of the discovery potential of a Higgs boson in the $WH \rightarrow \ell \nu b \bar{b}$ channel with the ATLAS experiment

Nicola Giacinto Piacquadio

CERN-THESIS-2010-027
11/01/2010



FAKULTÄT FÜR MATHEMATIK UND PHYSIK
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG

Identification of b -jets and investigation of the
discovery potential of a Higgs boson in the
 $WH \rightarrow \ell \nu b \bar{b}$ channel with the ATLAS experiment

DISSERTATION
zur Erlangung des Doktorgrades
der Fakultät für Mathematik und Physik
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von
Nicola Giacinto Piacquadio

Leiter der Arbeit:
Prof. Dr. Karl Jakobs

November 2009

Dekan:	Prof. Dr. Kay Königsmann
Referent:	Prof. Dr. Karl Jakobs
Korreferent:	Prof. Dr. Markus Schumacher
Prüfer (Experiment):	Prof. Dr. Gregor Hertn
Prüfer (Theorie):	Prof. Dr. Stefan Dittmaier
Tag der Prüfung:	11. Januar 2010

Contents

Introduction	1
1 Theoretical framework	3
1.1 The Standard Model of fundamental interactions	3
1.1.1 Strong interactions	4
1.1.2 Weak interactions	4
1.2 Mass generation and the Higgs mechanism	6
1.2.1 Higgs mechanism in a quantum field theory	6
1.2.2 Higgs field and couplings	7
1.3 Bounds on the Higgs boson mass	8
1.3.1 Theoretical bounds	8
1.3.2 Indirect bounds from electroweak precision measurements	10
1.3.3 Bounds from direct searches	11
1.4 From theory to event generation	12
1.4.1 Definition of cross section	12
1.4.2 Luminosity	15
1.4.3 Higher order corrections	16
1.4.4 All-order resummation	18
1.4.5 Interfacing higher fixed-order calculations with a parton-shower	20
1.4.6 Hadronisation	21
1.4.7 Underlying event	21
1.4.8 Pile-up	22
2 Higgs phenomenology	23
2.1 Higgs decays	23
2.2 Production modes at the LHC	24
2.3 ATLAS Discovery potential	25
2.4 Higgs decays to b -quarks	26
2.5 Measurement of Higgs couplings	28
3 The ATLAS Experiment and its environment	31
3.1 The Large Hadron Collider	31
3.2 The ATLAS detector	33
3.2.1 Inner Detector	35
3.2.2 Calorimeter	37
3.2.3 Muon system	41
3.2.4 Trigger system	42
4 Event reconstruction	44
4.1 Reconstruction of charged particle tracks	44
4.2 Electron identification	46
4.3 Jet reconstruction	49

4.3.1	Input to jet finding	49
4.3.2	Jet finding algorithms	50
4.3.3	Jet calibration	51
4.4	Missing transverse energy	52
4.5	Muon reconstruction	52
5	Reconstruction of the primary interaction vertex	54
5.1	Introduction	54
5.2	Primary vertices in pile-up environment	55
5.3	Mathematical determination of the vertex position	56
5.3.1	Billoir method	58
5.3.2	Kalman filter	60
5.3.3	Linearised measurement equation: computation of Jacobians	64
5.3.4	Jacobians for a neutral track	67
5.3.5	Adaptive methods for vertex fitting	68
5.4	Vertex finding algorithms	71
5.4.1	Default vertex finder	71
5.4.2	Seed finders	73
5.4.3	Adaptive vertex finder	75
5.4.4	Adaptive Multi-Vertex Finder	76
5.4.5	Signal selection in the presence of pile-up	77
5.5	Performance	78
5.5.1	Track selection	79
5.5.2	Vertex reconstruction performance without pile-up	82
5.5.3	Performance with no beam spot constraint	87
5.5.4	Vertex reconstruction performance in the presence of pile-up	88
5.5.5	Selection of the signal primary vertex	92
5.5.6	Number of vertices as a measure of instantaneous luminosity	94
5.5.7	Improvement of track momenta resolution using vertex constraint	96
6	Identification of b-quark jets	98
6.1	Motivation	98
6.2	Selection and association of tracks to jets	100
6.3	Impact parameter based b -tagging algorithms	104
6.4	Secondary vertex reconstruction in b -jets	112
6.5	Inclusive secondary vertex reconstruction	112
6.6	The JetFitter vertex reconstruction algorithm	113
6.6.1	Reconstruction of the decay chain	113
6.6.2	Vertex fit: formalism and implementation	115
6.6.3	Vertex finding strategy	121
6.6.4	Optimisation of the algorithmic execution speed	122
6.7	Optimisation of the physics performance	123
6.7.1	Initial track selection	123
6.7.2	Detection of two-track vertices	124
6.7.3	Explicit reconstruction of neutral particles	129
6.7.4	Final selection of tracks and finding procedure	134
6.8	Geometrical performance of the secondary vertex finders	135
6.8.1	Vertex association efficiency and purity	136
6.8.2	Separation and resolution of b/c -decay vertices	139

6.9	Secondary vertex b -tagging algorithms	142
6.9.1	Likelihood method based on a single inclusive vertex	143
6.9.2	Likelihood method based on JetFitter	143
6.9.3	Neural network discriminator based on JetFitter	147
6.9.4	Combination with impact parameter based algorithms	153
6.10	Evaluation of b -tagging performance	154
6.10.1	Definitions of efficiency and rejection	154
6.10.2	Physics events samples used	154
6.10.3	Performance in rejecting light-jets	155
6.10.4	Performance in rejecting charm-jets	161
7	The $WH \rightarrow \ell\nu b\bar{b}$ analysis at high transverse momenta	168
7.1	Introduction	168
7.2	Detector simulation	171
7.3	Event samples	171
7.4	Higgs Candidate Identification	177
7.4.1	Jet finding strategy	177
7.4.2	Invariant mass resolution	179
7.4.3	Identification of b -subjets	180
7.5	Analysis selection and sensitivity based on event counting	182
7.6	Trigger selection	186
7.7	Optimisation of b -tagging performance	190
7.8	Jet clustering: properties and optimisation	195
7.8.1	Comparison to traditional jet techniques	195
7.8.2	Optimisation of subjet clustering algorithm	199
7.9	Optimisation of jet energy scale and mass resolution	200
7.10	Comparison with fast simulation	204
7.11	Combination with the ZH channels	205
8	Fit based approach and systematic uncertainties	207
8.1	Impact of systematic uncertainties	207
8.2	Maximum likelihood formalism	208
8.2.1	Definition of signal and background PDFs	210
8.2.2	Determination of PDF shapes using Monte Carlo events	212
8.2.3	Correlation between fit variables	213
8.3	Definition of control samples	222
8.3.1	Control sample for mass and b -weight PDFs	222
8.3.2	Control sample for p_T of the additional jet PDF	227
8.3.3	Implementation in the likelihood fit	234
8.3.4	Control samples for the Wt background	236
8.3.5	Control samples for the W +jet background	236
8.4	Fit validation on mixed Monte Carlo sample	243
8.5	Fit validation and discovery potential using Monte Carlo pseudo-experiments	252
8.5.1	Fit validation	252
8.5.2	Discovery potential	253
8.6	Introduction of experimental systematic uncertainties	261
8.7	Modelling of systematic uncertainties	269
8.8	Impact of systematics on the maximum likelihood fit	279

Summary and conclusions	284
A Primary vertex reconstruction	287
A.1 Details on computation of track parameterisation	287
A.1.1 Computation of Jacobian derivatives	287
A.1.2 Approximation in the limit of large curvatures	288
A.2 Implementation in the ATLAS software ATHENA	289
A.2.1 Event Data Model	290
A.2.2 Vertexing algorithms and tools	292
A.3 Tuning of parameters in the adaptive vertex fitters	298
B Details about the JetFitter algorithm	300
B.1 Optimization of the algorithmic execution speed	300
B.2 Validation of neutral reconstruction	301
B.3 Implementation in the ATLAS reconstruction software	304
C WH analysis: further details	309
C.1 Parton level study of the $t\bar{t}Z$ background	309
C.2 $Wb\bar{b}$ background: comparison between HERWIG and AcerMC	309
C.3 Subjet technique compared to traditional jet clustering	311
C.3.1 Improved rejection of $t\bar{t}$ background	311
C.3.2 Impact of b -tagging on collinear $b\bar{b}$ pairs	311
C.4 Comparison between ATLFast-II and full simulation	312
C.4.1 Lepton reconstruction efficiency	312
C.4.2 Comparison of analysis selection on signal events	316
D WH analysis: fit based approach	324
D.1 Cross-check of correlations	324
Bibliography	335
Acknowledgements	339

Introduction

The Standard Model of particle physics describes three of the four known fundamental interactions between the elementary particles: the electromagnetic, weak and strong forces. It provides an extremely accurate description of the electroweak interactions up to the energy scales so far explored in high energy physics experiments.

The Large Hadron Collider (LHC), which is presently starting to operate, will provide proton-proton collisions with an unprecedented centre-of-mass energy of $\sqrt{s} = 14$ TeV and with instantaneous luminosities of up to $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, and is therefore ideally suited to explore the TeV energy domain. Two multipurpose experiments, ATLAS and CMS, were built to analyse the collisions.

The high instantaneous luminosities achievable at the LHC will result in a significant contamination of the signal processes by additional soft proton-proton collisions, usually known as pile-up interactions. In the course of this thesis several algorithms were developed for the ATLAS experiment to reconstruct the position of the primary interaction vertex with improved precision, which rely on adaptive methods to reduce both the influence of pile-up interactions and of secondary interactions. The Kalman filter and mathematical formalism these methods rely on is described in detail and their performance is analysed through the use of Monte Carlo simulations.

There is one particle predicted by the Standard Model whose existence has not yet been proven: the Higgs boson, which plays the crucial role of giving other particles a mass without breaking the gauge symmetry the model is built upon. All physical properties of this particle are predicted by the theory, except its mass.

The most stringent limit on the Higgs boson mass is provided by the direct searches performed at the LEP2 collider, which exclude a Standard Model Higgs boson with a mass below $114.4 \text{ GeV}/c^2$ at 95% confidence level. Indirect constraints from electroweak precision observables, where the Higgs boson enters through virtual corrections, predict a Higgs boson mass of $87_{-26}^{+35} \text{ GeV}/c^2$. Under the assumption that the Standard Model is valid, the Higgs boson mass is expected to be found close to the LEP2 limit. In the low mass region, the Higgs boson decays preferentially into a pair of b -quarks, which can be observed in the detector as b -quark jets.

The experimental identification of b -jets and the rejection of the copious backgrounds from u , d , s -quarks and gluon jets is made possible by the relatively long lifetime of b -hadrons ($\beta\gamma c\tau = O(\text{mm})$), whose decay particles are thus produced at a secondary vertex which is sufficiently displaced with respect to the primary interaction vertex. The most promising b -jet identification algorithms are based either on detecting individual charged particle tracks which are displaced with respect to the primary interaction vertex or on the explicit reconstruction of the secondary vertex position. In this thesis a sophisticated secondary vertex reconstruction algorithm is presented, which exploits the topological structure of weak b - and c -hadron decays inside a b -jet and is implemented mathematically as an extension of the Kalman filter formalism for vertex reconstruction. The application to a b -jet identification algorithm is discussed in detail and its performance is compared with the standard algorithms available in ATLAS.

Even under the hypothesis that an excellent b -jet identification performance can be achieved,

the search for Higgs bosons decaying to a b -quark pair will still suffer from copious irreducible backgrounds with b -jets produced by strong interactions. This is the main reason why no Higgs boson search based on the $b\bar{b}$ decay mode at the LHC is actually considered to yield a significant discovery potential. Searches with the Higgs boson decaying into a pair of photons or taus are typically preferred in the $m_H = 115\text{-}130 \text{ GeV}/c^2$ mass region.

Recently, a new approach for the search for Higgs bosons decaying to a pair of b -quarks was proposed [1], which relies on the already well known associated production of a Higgs boson with a W or Z boson, but where only the phase space region where the Higgs and the vector bosons are produced at large transverse momenta is considered. This allows a more efficient rejection of the backgrounds, but, at the same time, poses several new problems to the object identification algorithms.

A first detector level study of the $WH \rightarrow \ell\nu b\bar{b}$ Higgs boson search channel, based on a realistic simulation of the ATLAS detector, is presented in this thesis. Special emphasis is put on the implementation of the jet reconstruction algorithm and on the optimisation of the b -jet identification performance in this specific scenario. The discovery potential, based on simple event counting, is analysed in terms of statistical significance.

A significant degradation of the discovery potential is expected due to the sizable systematic uncertainties affecting the expected amount of background feeding into the signal selection. In order to reduce the impact of these uncertainties, a maximum likelihood based estimator is defined to extract the number of signal and background events directly from the data, based on the expected distributions of a few discriminating variables. Special care is taken to evaluate the effect of experimental systematic uncertainties and the discovery potential is evaluated for Higgs boson masses in the range between 115 and 130 GeV/c^2 .

The thesis is organised as follows: a short overview of the theory of elementary particles and interactions, with emphasis on the mechanism of symmetry breaking, is given in Chapter 1. An introduction to Higgs boson searches at the LHC, with some details about Higgs boson searches with decays into a pair of b -quarks, is contained in Chapter 2. The LHC collider and the ATLAS detector are briefly described in Chapter 3, while an overview of how physics events are reconstructed and how the final state particles are identified is given in Chapter 4. The reconstruction of the primary interaction vertex, the implementation of new vertex finding techniques and their performance are presented in Chapter 5. In Chapter 6, the b -jet identification in ATLAS is discussed, with emphasis on the design and application of a novel technique to reconstruct b - to c -hadron vertex decay chains in b -jets. The event counting based $WH \rightarrow \ell\nu b\bar{b}$ analysis is presented in Chapter 7, while the maximum likelihood based analysis is presented in Chapter 8.

1 Theoretical framework

In this chapter a short overview of the present theory of elementary particles and fundamental interactions is given, with some emphasis on the Higgs mechanism of symmetry breaking, which is the object of the second part of this thesis. As a general rule, in this thesis natural units are used, defining $\hbar = 1$ and $c = 1$ (except in the introduction and in the summary).

1.1 The Standard Model of fundamental interactions

The Standard Model (SM) [2–4] is a quantum field theory which describes the electromagnetic, weak and strong interactions between all elementary particles whose existence has been proven up to today.

This model contains elementary particles with spin 1/2 as building blocks of matter:

- six leptons (and their antiparticles), organised in three families

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} \quad (1.1)$$

- six quarks (and their antiparticles), organised in three families

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \quad (1.2)$$

In addition it contains the particles which are responsible for the fundamental interactions, given by the following vector bosons:

- a gauge boson for the electromagnetic interactions, the photon γ ;
- three gauge bosons for the weak interaction, W^\pm and Z^0 ;
- eight gauge bosons for the strong interaction, called gluons;

and one scalar boson:

- the Higgs boson, H (not yet experimentally observed).

The Standard Model is based on a symmetry group of the kind $SU(3)_C \otimes SU(2)_I \otimes U(1)_Y$, where $SU(3)_C$ describes the *colour* symmetry of strong interactions, $SU(2)_I$ describes the *weak isospin* I for the unified electroweak interactions and $U(1)_Y$ the invariance under *hypercharge* Y transformations. The theory is based on a base Lagrangian density where the masses of all particles are set to zero, which is invariant under transformations of the group symmetry the model is built upon. The Lagrangian of the Standard Model can be separated into a sum of a term \mathcal{L}_{QCD} which is responsible for the strong interactions and an electroweak term \mathcal{L}_{EW} .

1.1.1 Strong interactions

Quantum chromodynamics (QCD) is the theory of strong interactions [5–7], which describes the interactions of quarks and gluons. This theory is based on the assumption that hadrons consist of structureless, confined quarks which obey exact $SU(3)$ colour symmetry. Its Lagrangian density is given by

$$\mathcal{L}_{QCD} = i \sum_f \bar{q}_f \gamma^\mu D_\mu q_f - \frac{1}{4} G_{\mu\nu}^i G_{\mu\nu}^i \quad (1.3)$$

where q_f is a colour triplet of quarks with flavour f and the covariant derivative D_μ is defined as

$$D_\mu = \partial_\mu - ig_3 \frac{\lambda_i}{2} G_\mu^i \quad (1.4)$$

where g_3 is the $SU(3)$ coupling constant and the tensor fields $G_{\mu\nu}$ are defined as

$$G_{i,\mu\nu} = \partial_\mu G_{i,\nu} - \partial_\nu G_{i,\mu} + gf_{ijk} G_{j,\mu} G_{k,\nu}, \quad (1.5)$$

where $G_{i,\nu}$ are the eight gluon fields, λ_i the generators of $SU(3)_C$ (Gell-Mann matrices) and f_{ijk} the structure constants of $SU(3)_C$.

1.1.2 Weak interactions

All experimental observations where weak decays take place are consistent with the assumption that the current of the weak interactions is always of the form $(V - A)$, i.e. a specific combination of axial and vector currents. It seems therefore natural to define the fields of the interacting particles in terms of the respective right- and left-handed helicity projections, which, for a generic field $\phi(x)$, are defined as:

$$\begin{aligned} \phi_L(x) &= P_L \phi(x) = \frac{1}{2} (1 - \gamma_5) \phi(x) \\ \phi_R(x) &= P_R \phi(x) = \frac{1}{2} (1 + \gamma_5) \phi(x). \end{aligned}$$

The different transformation properties of the two components $\phi(x)$ and $\gamma_5 \phi(x)$ lead to an asymmetry in the theory, which translates into a violation of the symmetries of parity \mathcal{P} and charge conjugation \mathcal{C} . This makes it natural to define *left* and *right* fields of quarks and leptons into the following multiplets:

$$Q_L^I(x) = \begin{pmatrix} U_L^I(x) \\ D_L^I(x) \end{pmatrix} = (3, 2)_{+1/6} \quad u_R^I(x) = (3, 1)_{+2/3} \quad d_R^I(x) = (3, 1)_{-1/3} \quad (1.6)$$

for the quarks, and

$$\Psi_L(x) = \begin{pmatrix} \nu_L(x) \\ l_L(x) \end{pmatrix} = (1, 2)_{-1/2} \quad l_R(x) = (1, 1)_{-1} \quad \nu_R(x) = (1, 1)_0 \quad (1.7)$$

for the leptons, where $(3, 2)_{+1/6}$ reflects a triplet of colour $SU(3)_C$, a doublet of weak isospin $SU(2)_I$ with hypercharge $Y = Q - I_3 = +1/6$ and analogously for the other representations. The electroweak Lagrangian can therefore be written as

$$\mathcal{L}_{EW} = i[\bar{Q}_L^I(x) \gamma^\mu D_\mu Q_L^I(x) + \bar{u}_R(x) \gamma^\mu D_\mu u_R(x) + \bar{d}_R(x) \gamma^\mu D_\mu d_R(x) +$$

$$\begin{aligned} & \bar{\Psi}_L(x)\gamma^\mu D_\mu \Psi_L(x) + \bar{\nu}_R(x)\gamma^\mu D_\mu \nu_R(x) + \bar{l}_R(x)\gamma^\mu D_\mu l_R(x) \\ & - \frac{1}{4}B_{\mu\nu}(x)B^{\mu\nu}(x) - \frac{1}{4}W_{\mu\nu}^i(x)W_i^{\mu\nu}(x) \end{aligned} \quad (1.8)$$

where the covariant derivative is given by

$$D_\mu = \partial_\mu + i\frac{g}{2}\tau_j W_j^\mu(x) - ig'YB^\mu(x) \quad (1.9)$$

where g and g' are the coupling constants associated to the gauge fields $W_j(x)$ ($j = 1, 2, 3$) and $B(x)$ related respectively to the symmetry group of isospin $SU(2)_I$ and of hypercharge $U(1)_Y$, while τ_j ($j = 1, 2, 3$) are the Pauli matrices. The last two terms in the Lagrangian contain the kinetic term of the vector bosons: as in the case of the gluons, the tensor field $F_{\mu\nu}^i$ can be written as a function of the vector field A_i as

$$F_{i,\mu\nu} = \partial_\mu A_{i,\nu} - \partial_\nu A_{i,\mu} + gf_{ijk}A_{j,\mu}A_{k,\nu} \quad (1.10)$$

where g is the coupling constant associated to the vector field and f_{ijk} are the structure constants of the Lie algebra related to the gauge symmetry group. Since $SU(2)_I$ is a non-abelian group, triple and quartic self-coupling terms in the vector bosons will necessarily appear. The Lagrangian can be separated in two terms, a term which describes the free theory and a second one which accounts for the interaction, according to

$$\mathcal{L}_{EW} = \mathcal{L}_0 + \mathcal{L}_{INT}. \quad (1.11)$$

Through a redefinition, the gauge fields W_j and B can be expressed in terms of the observable vector bosons W^\pm , Z and A :

$$\begin{aligned} W_1^\mu(x) &= (W^\mu(x) + W^{\dagger\mu}(x))/\sqrt{2} & W_3^\mu &= \cos(\theta_W)Z^{0\mu}(x) + \sin(\theta_W)A^\mu(x) \\ W_2^\mu(x) &= (W^\mu(x) - W^{\dagger\mu}(x))/\sqrt{2} & B^\mu &= -\sin(\theta_W)Z^{0\mu}(x) + \cos(\theta_W)A^\mu(x) \end{aligned} \quad (1.12)$$

where θ_W is the weak mixing angle and the fields W and W^\dagger are related to the W^+ and W^- bosons. The interaction term of the Lagrangian can thus be re-expressed with

$$\begin{aligned} \mathcal{L}_{INT} &= \frac{g_W}{2\sqrt{2}} \left(J_\mu^+(x)W^\mu(x) + J_\mu^-(x)W^{\dagger\mu}(x) \right) \\ &+ eJ_\mu^{em}(x)A^\mu(x) + \frac{g_W}{2\cos(\theta_W)}J_\mu^0(x)Z^{0\mu}(x) \end{aligned} \quad (1.13)$$

where g_W is the electroweak coupling constant and being

$$J_\mu^+ = \bar{u}\gamma_\mu(1 - \gamma_5)d + \bar{c}\gamma_\mu(1 - \gamma_5)s + \bar{t}\gamma_\mu(1 - \gamma_5)b + \quad (1.14)$$

$$\bar{\nu}_e\gamma_\mu(1 - \gamma_5)e + \bar{\nu}_\mu\gamma_\mu(1 - \gamma_5)\mu + \bar{\nu}_\tau\gamma_\mu(1 - \gamma_5)\tau$$

$$J_\mu^{em} = \sum_f Q_f \bar{f}\gamma_\mu f \quad (1.15)$$

$$J_\mu^0 = \sum_f \bar{f}\gamma_\mu(v_f - a_f\gamma_5)f \quad (1.16)$$

$$v_f = \tau_3^f - 2Q_f \sin^2 \theta_W \quad a_f = \tau_3^f \quad (1.17)$$

where the sum over the index f represents the sum over all possible flavours and θ_W is defined by the relation

$$\frac{g}{\sin \theta_W} = \frac{g'}{\cos \theta_W} = e. \quad (1.18)$$

The currents $J^\pm(x)$ reflect the $(V - A)$ structure of the weak interactions associated to the W^+ and W^- bosons, while $J^{em}(x)$ accounts for the electromagnetic force and $J^0(x)$ is the neutral current associated to the Z^0 boson.

1.2 Mass generation and the Higgs mechanism

The Lagrangian introduced so far describes a theory of massless particles, which is in clear contradiction with experimental observations. In order to keep the theory renormalisable, it is necessary to introduce particle masses in a way which preserves the original gauge invariance of the Lagrangian. This can be obtained through the *Higgs mechanism* [8, 9], where the spontaneous symmetry breaking from $SU(3)_C \otimes SU(2)_I \otimes U(1)_Y$ to $SU(3)_C \otimes U(1)_Q$ provides masses to the W^\pm and Z^0 bosons, while the photon, being associated to the residual symmetry $U(1)_Q$, does not acquire any mass. The symmetry breaking is realised by introducing an isospin doublet of complex scalar fields, called the *Higgs* field Φ

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad \Phi^c = (i \cdot \tau_2) \Phi^* = \begin{pmatrix} \phi^{0*} \\ -\phi^- \end{pmatrix} \quad (1.19)$$

which, by imposing gauge invariance, can be described without loss of generality by a Lagrangian of the kind

$$\mathcal{L}_{Higgs} = D^\mu \Phi^\dagger D_\mu \Phi - \underbrace{\mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2}_{V(\Phi)} \quad \mu, \lambda \in \mathbb{R} \quad (1.20)$$

where the covariant derivative is given by Eq. 1.9. Under the hypothesis that $\mu^2 < 0$, the potential $V(\Phi)$ associated to the Higgs field has a local maximum for a value of the Higgs field of $\Phi(x) = 0$ and an absolute minimum along the circumference $\left(\frac{-\mu^2}{2\lambda}\right)^{1/2} e^{i\theta}$ ($\theta \in [0, 2\pi]$). This implies that the state of minimum energy, the vacuum state, is not unique but degenerate. The necessity to select only one of these states (which happens to describe our physical world) results in the Hilbert space not sharing anymore all symmetries of the Lagrangian and translates into a spontaneous symmetry breaking. This choice is made by assigning one of the possible non-zero expectation values to the vacuum

$$\langle 0 | \Phi(x) | 0 \rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad v = \left(-\frac{\mu^2}{2\lambda}\right)^{1/2} > 0 \quad (1.21)$$

In this way, the kinetic term $(D^\mu \Phi)^\dagger D_\mu \Phi$ generates the masses of the vector bosons W^\pm and Z^0 , corresponding to $m_W = \frac{vg}{2}$ and $m_Z = \frac{vg}{2\cos\theta_W}$, while those of the leptons and quarks can be generated by add-hoc couplings with the Higgs field through the Yukawa terms

$$Y_{ik}^u (\bar{Q}_L^i \tilde{\Phi}) u_R^k + Y_{ik}^d (\bar{Q}_L^i \Phi) d_R^k + Y_{ik}^l (\bar{L}_L^i \Phi) l_R^k + h.c. \quad (1.22)$$

where the indexes i and k refer to the quark or lepton generation. Inserting the expectation value for the Higgs field, for the quark masses a Lagrangian term is obtained of the kind

$$\mathcal{L}_M = \bar{u}^i M_{ik}^u (1 + \gamma_5) u^k + \bar{d}^i M_{ik}^d (1 + \gamma_5) d^k + h.c. \quad (1.23)$$

where

$$M_{ik}^u = \frac{Y_{ik}^u \cdot v}{\sqrt{2}} \quad M_{ik}^d = \frac{Y_{ik}^d \cdot v}{\sqrt{2}} \quad (1.24)$$

1.2.1 Higgs mechanism in a quantum field theory

In the previous sections the Higgs mechanism was considered only in terms of a classical field theory. Quantum level corrections can significantly change this picture. An elegant generalisation of the Higgs mechanism to a quantum field theory is provided by the quantisation

through the path integral method (an introduction can be found for example in Ref. [10]). In this context, the quantum analog of the potential $V(\Phi)$ is the *effective potential* $V_{eff}(\Phi_{cl})$, defined in term of the *effective action* $\Gamma[\Phi_{cl}]$ (the generating functional of the 1PI connected correlation functions) as:

$$V_{eff}(\Phi_{cl}) = -\frac{1}{VT}\Gamma[\Phi_{cl}] \quad \text{for} \quad \Phi_{cl}(x) = \text{constant} = \Phi_{cl} \quad (1.25)$$

where VT is the space-time extent of the functional integral (since $\Gamma[\Phi_{cl}]$ is an extensive quantity) and Φ_{cl} is the vacuum expectation value of the field configuration $\Phi(x)$

$$\Phi_{cl}(x) = \langle \Omega | \Phi(x) | \Omega \rangle \quad (1.26)$$

The stable quantum states of the theory are obtained by imposing a minimum for the effective action:

$$\frac{\partial}{\partial \Phi_{cl}} \Gamma[\Phi_{cl}] = 0 \quad \rightarrow \quad \frac{\partial}{\partial \Phi_{cl}} V_{eff}(\Phi_{cl}) = 0, \quad (1.27)$$

which identifies the states of minimum energy of the theory, i.e. the stable vacuum states. As in the classical picture, a system with spontaneous symmetry breaking has several minima, all with the same energy. Specifying one of them breaks the original symmetry on the vacuum. The use of the effective potential makes the link between the classical and quantum case particularly transparent, since the effective potential can be expressed in the form of a perturbative series, which can be then calculated order by order in \hbar

$$V_{eff}(\Phi_{cl}) = V(\Phi_{cl}) + \text{loop effects}, \quad (1.28)$$

where $V(\Phi_{cl})$ is the classical potential considered in the previous section. The way loop corrections affect the *effective potential* is considered in Section 1.3.1, to derive some theoretical bounds on the Higgs mass.

1.2.2 Higgs field and couplings

The particle content of the scalar sector of the theory becomes evident after choosing the unitary gauge: this consists in eliminating the unphysical degrees of freedom of the theory using gauge invariance. The Higgs doublet field introduced in (1.19) is redefined as follows

$$\Phi(x) = \frac{e^{\frac{i}{v}\vec{\chi}(x)\cdot\vec{\tau}}}{\sqrt{2}} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix} \xrightarrow{SU(2)} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}, \quad (1.29)$$

after which the scalar potential of (1.20) becomes:

$$\mathcal{L}_{Higgs} = \mu^2 H^2 - \lambda v H^3 - \frac{1}{4} H^4 = -\frac{1}{2} m_H^2 H^2 - \sqrt{\frac{\lambda}{2}} m_H H^3 - \frac{1}{4} \lambda H^4 \quad (1.30)$$

When moving to the unitary gauge, three degrees of freedom, corresponding to the $\chi^a(x)$ Goldstone bosons, are reabsorbed into the longitudinal components of the W^\pm and Z^0 weak gauge bosons. One only scalar field remains, the *Higgs boson* H , with mass $m_H^2 = -2\mu^2 = 2\lambda v^2$ and self-couplings:



$$= -3i \frac{m_H^2}{v}$$



$$= -3i \frac{m_H^2}{v^2}$$

Furthermore, the kinetic term of Eq. 1.20, after redefining the Higgs field according to Eq. 1.29, contains linear terms in the gauge bosons $W_\mu^{(\dagger)}$ and Z_μ^0 , which define the Higgs coupling to the weak gauge bosons:

$$\begin{array}{c} \text{Diagram 1: } g^{\mu\nu} \text{ (gluons)} \rightarrow H = 2i \frac{m_V^2}{v} g^{\mu\nu} \\ \text{Diagram 2: } g^{\mu\nu} \text{ (gluons)} \rightarrow H = 2i \frac{m_V^2}{v^2} g^{\mu\nu} \end{array}$$

Finally, the explicit introduction of masses for the quarks and leptons through the Yukawa couplings of (1.22) also introduces a coupling of the Higgs to the fermions:

$$f \text{ and } \bar{f} \rightarrow H = -i \frac{m_f}{v}$$

The Higgs boson tree level couplings can all be expressed in terms of just two parameters, either μ and λ appearing in the scalar potential of (1.20) or, equivalently, m_H and v , the Higgs boson mass and the scalar field vacuum expectation value. The value of v is measured in muon decays to be $v = (\sqrt{2}G_F)^{-1/2} = 246$ GeV, therefore the physics of the Standard Model Higgs boson is actually fully determined once its mass m_H is known.

1.3 Bounds on the Higgs boson mass

1.3.1 Theoretical bounds

Various theoretical bounds constrain the Higgs mass, based on the following arguments:

- Unitarity
- Triviality
- Vacuum stability

Unitarity

The unitarity argument applies to the scattering process of longitudinal gauge bosons, $V_L V_L \rightarrow V_L V_L$, where V is either a W^\pm or a Z boson. Thanks to the *electroweak equivalence theorem*, in the high energy limit ($s \gg m_V^2$) the scattering amplitude for longitudinal gauge bosons can be expressed in terms of the scattering amplitude for the corresponding Goldstone bosons:

$$A(W_L^+ W_L^- \rightarrow W_L^+ W_L^-) = A(\omega^+ \omega^- \rightarrow \omega^+ \omega^-) + O\left(\frac{m_W^2}{s}\right), \quad (1.31)$$

where ω^\pm are the Goldstone bosons corresponding to the W^\pm bosons and the scalar scattering amplitude is:

$$A(\omega^+ \omega^- \rightarrow \omega^+ \omega^-) = -\frac{m_H^2}{v^2} \left(\frac{s}{s - m_H^2} + \frac{t}{t - m_H^2} \right), \quad (1.32)$$

where s and t are the Mandelstam variables of the scattering process. The amplitude can be decomposed in partial waves of spin l (in terms of Legendre polynomials) and the overall cross section in terms of the coefficients a_l related to the single partial waves l :

$$\sigma = \frac{16\pi}{s} \sum_{l=0}^{\infty} (2l+1) |a_l|^2 \quad (1.33)$$

The coefficients a_l can be computed from the full scattering amplitude of Eq. 1.32, but here only the term with $l = 0$ is interesting. In the high energy limit ($s \gg m_H^2$) it turns out to be

$$a_0 \rightarrow \frac{m_H^2}{8\pi v^2}. \quad (1.34)$$

Using the optical theorem, it is possible to apply the unitarity constraint directly on top of (1.33):

$$\sigma = \frac{16\pi}{s} \sum_{l=0}^{\infty} (2l+1) |a_l|^2 = \frac{1}{s} \text{Im} [a(\theta = 0)] \quad (1.35)$$

Expanding also the second term in partial waves, it is easy to realise that:

$$|a_l|^2 = \text{Re}(a_l)^2 + \text{Im}(a_l)^2 = \text{Im}(a_l) \rightarrow |\text{Re}(a_l)| \leq \frac{1}{2} \quad (1.36)$$

Using this inequality, the high energy limit of a_0 in (1.34) results in a limit on the Higgs boson mass

$$m_H < 870 \text{ GeV}. \quad (1.37)$$

By applying the same argument on other scattering channels (e.g. $W_L^+ W_L^- \rightarrow Z_L Z_L$) this bound can be tightened to:

$$m_H < 710 \text{ GeV} \quad (1.38)$$

The unitarity argument can be also reverted. If no Higgs boson is found in nature, then the same scattering amplitudes above can be computed in the opposite limit $m_H \gg s$ (which is equivalent to a theory with no Higgs boson) and an upper limit is found for the scale where the scattering amplitude violates unitarity and new physics contributions are therefore expected to appear:

$$\sqrt{s} < 1.2 \text{ TeV} \quad (1.39)$$

As a consequence, either the Higgs boson exists and is below ≈ 1 TeV or new physics contribution at a scale of 1-2 TeV must appear. This ensures that at the LHC, where these energies will be well within reach, at least one of these two discoveries should be realised.

Triviality and vacuum stability

The second bound comes from loop corrections to the classical Higgs potential, due to the running of the quartic Higgs self-coupling λ [11, 12]. The dependence of λ on the energy scale Q in a ϕ^4 -theory is described by the renormalisation group equation (RGE):

$$\frac{d\lambda(Q)}{dQ^2} = \frac{3}{4\pi^2} \lambda^2(Q) \quad (1.40)$$

As a consequence, the quartic coupling λ decreases for small energies and increases for large energies. In the Standard Model, the situation is a bit more complex and the RGE for the quartic coupling is coupled to the evolution of the other coupling constants g, g' and y_t :

$$32\pi^2 \frac{d\lambda}{d(\log \frac{Q^2}{Q_0^2})} = 24\lambda^2 - (3g'^2 + 9g^2 - 24y_t^2)\lambda + \frac{3}{8}g'^4 + \frac{3}{4}g'^2g^2 + \frac{9}{8}g^4 - 24y_t^4 + \dots, \quad (1.41)$$

where $y_t = m_t/v$ is the Yukawa coupling to the top quark, which dominates over all other Yukawa couplings, and g and g' are the couplings to the weak bosons. In the limit of high energies, the first term of (1.41) dominates and the RGE equation can be solved, starting from a certain scale Q_0 :

$$\lambda(Q^2) = \lambda(Q_0^2) \left[1 - \underbrace{\frac{3}{4\pi^2} \lambda(v^2) \log \frac{Q^2}{Q_0^2}}_A \right]^{-1} \quad (1.42)$$

As a result, λ becomes infinite when $A \rightarrow 1$, i.e. at the Landau pole, and the perturbative description is no longer valid. The energy scale up to which the theory remains non perturbative depends on the exact value of the Higgs boson mass. Requiring $1 - A$ to be non-zero, setting $Q = \Lambda$ and the starting scale Q_0 equal to the electroweak scale v yields:

$$m_H^2 < \frac{8\pi^2 v^2}{3 \log \left(\frac{\Lambda^2}{v^2} \right)}. \quad (1.43)$$

This relation can be interpreted as an upper bound of m_H as a function of the scale Λ where new physics is expected to enter the game. If this condition is not met, the only way to preserve a perturbative theory valid at all scales is to impose triviality ($\lambda = 0$), which however would prevent the spontaneous symmetry breaking to take place; this is exactly why this condition goes under the name of *triviality*.

From the RGE for λ (1.41) another bound can also be obtained: in fact, in order to preserve the stability of the vacuum state, λ must remain positive. At low energies, the last term $-6y_t^4$ dominates, yielding to an evolution of $\lambda(Q)$ which is close to:

$$\lambda(\Lambda) \approx \lambda(v) - \frac{3}{4\pi^2} y_t^2 \log \left(\frac{\Lambda^2}{v^2} \right) \quad (1.44)$$

The constraint $\lambda(\Lambda) > 0$ translates into a constraint on m_H :

$$\lambda(\Lambda) > 0 \quad \rightarrow \quad m_H^2 > \frac{3v^2}{2\pi^2} y_t^2 \log \left(\frac{\Lambda^2}{v^2} \right) \quad (1.45)$$

The constraints provided by both the triviality and vacuum stability bounds are shown in Fig. 1.1.

1.3.2 Indirect bounds from electroweak precision measurements

The presence of the Higgs particle is not only detectable if produced directly, but also affects through virtual corrections several physics observables: the mass of the W boson, various leptonic and hadronic asymmetries and many other electroweak observables which are usually used to make precision tests of the Standard Model. Exploiting this dependence, a global

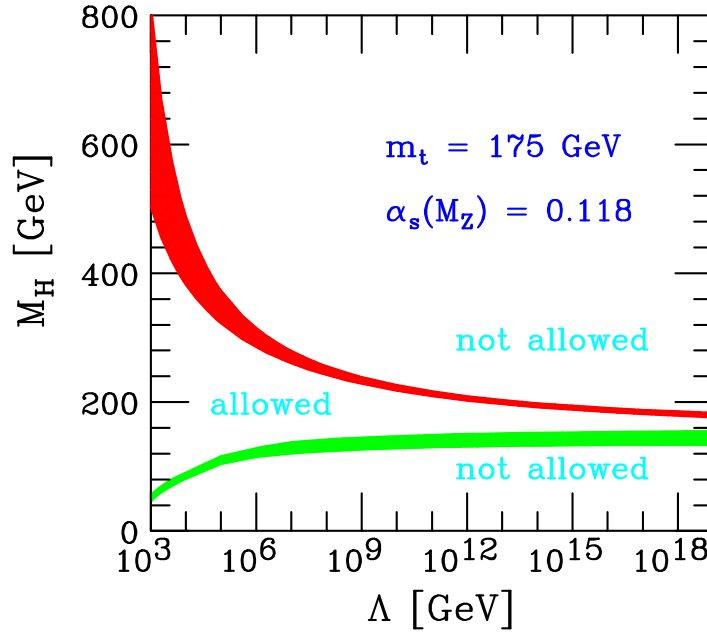


Figure 1.1: The triviality (upper) and vacuum stability (lower) bound on the Higgs boson mass as a function of the new physics or cutoff scale Λ . A value of $m_t = 175 \pm 6$ GeV and of $\alpha_s(m_Z) = 0.118 \pm 0.002$ was used to derive these limits. Taken from Ref. [13].

electroweak fit can result in an indirect measurement of the Higgs boson mass. This is analogous to the case of the top quark, whose mass was predicted before it was actually observed experimentally. However, the dependence on the Higgs boson mass in electroweak observables is weaker than for the top quark mass, typically logarithmic instead of quadratic, so the precision with which the Higgs mass can be determined in the global fit is limited. The results presented here have been obtained by the LEP electroweak working group [14].

The result of the global electroweak fit in terms of predicted top quark and W boson masses is shown in terms of a two-dimensional correlation plot in Fig. 1.2, together with the correlated Higgs boson mass values. The result of the direct measurement of the top quark and the W boson masses is also shown on top: with respect to the indirect measurements, the measured values seem to push the Higgs boson mass towards lower values. The uncertainty in the correlation plot due to the error on the determination of the QED fine-structure constant in the fit is shown in addition with a magenta arrow. Integrating the direct measurements of the top quark and W boson masses into the fit, the famous *blue band* plot can be obtained, which is shown in Fig. 1.3. The predicted Higgs boson mass is 87^{+35}_{-26} GeV (at 68% confidence level). No information from direct searches of the Higgs boson is included here.

1.3.3 Bounds from direct searches

The indirect bounds from the electroweak precision measurements can be compared with the regions of Higgs boson masses experimentally excluded by direct searches for Higgs boson decays. These regions are shown in yellow in Fig. 1.3. The region below $m_H = 114.4$ GeV was excluded at 95% confidence level by direct searches at LEP2 [15], while the region between 160 and 170 GeV was recently excluded at the Tevatron collider. The experiments at the Tevatron are starting to get sensitive to a Higgs boson with masses around the WW mass threshold [16] and the actual exclusion limits, normalised to the expected Standard Model

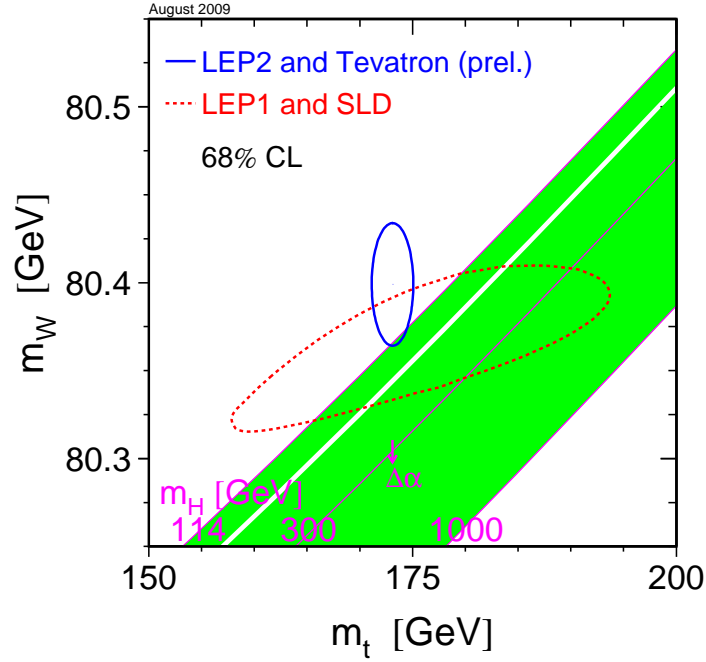


Figure 1.2: Correlation between the top quark and W boson mass in the electroweak fit, compared with the actual world average (68% confidence level curves, in red for LEP1 and SLD, in blue for LEP2 and Tevatron). The contour lines at $\pm 1\sigma$ with the corresponding values for the Higgs boson mass are also shown. Taken from the most recent update of Ref. [14].

cross section, are shown more in detail in Fig. 1.4, together with the expected exclusions and their 1 and 2σ uncertainty bands.

1.4 From theory to event generation

1.4.1 Definition of cross section

Once the Standard Model is defined in terms of its particle content and Lagrangian density, it is possible to calculate any possible transition from a state i to a state f . This is provided by the S matrix, which can be always written as $S = 1 + iT$, where the unity represents the case where no interaction occurs, while the effect of interactions is encoded in the transition matrix T . Based on the matrix T , the probability amplitude \mathcal{A} for a transition between the initial state i and the final state f to occur can be written as a function of the matrix element \mathcal{M} :

$$\mathcal{A}(i \rightarrow f) = \langle f | iT | i \rangle = (2\pi)^4 \delta^{(4)}(\sum p_i - \sum p_f) i\mathcal{M}(p_i \rightarrow p_f), \quad (1.46)$$

where p_i represent the momenta of the particles in the initial state and p_f the corresponding ones in the final state.

For $2 \rightarrow N$ processes, where the two incoming particles are defined here as a and b , the cross section can be obtained from the matrix element \mathcal{M} by considering an additional flux

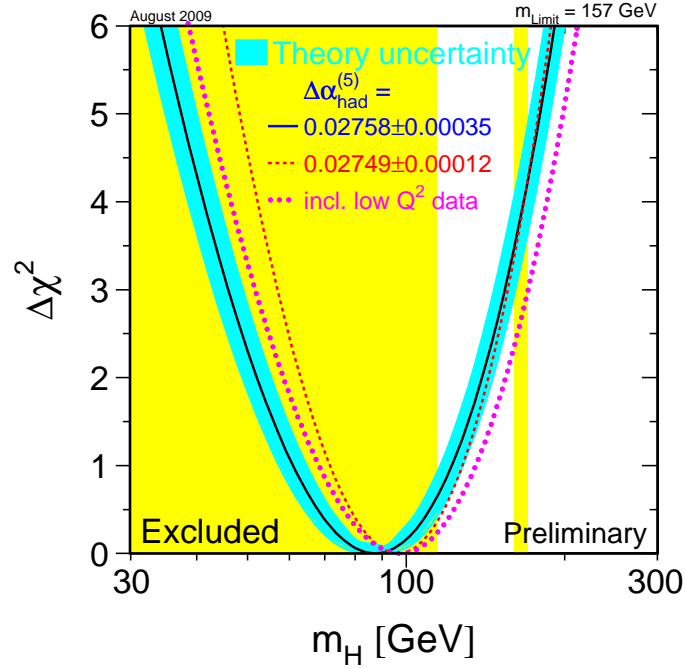


Figure 1.3: Result of the global fit to electroweak observables expressed as $\Delta\chi^2 = \chi^2 - \chi_{min}^2$ as a function of the assumed value for m_H . The theory uncertainty (missing higher order corrections) is shown as a blue-band. The dotted lines indicate the curves corresponding to a different value of $\Delta\alpha_{had}^{(5)}(m_Z^2)$ and to including low Q^2 data from the NuTeV neutrino scattering experiment. $\Delta\alpha_{had}^{(5)}(m_Z^2)$ is the dominant contribution to the uncertainty on $\alpha(m_Z^2)$ and is given by the difference between two different estimates of the contribution of light quarks to the photon vacuum polarisation. Taken from the most recent update of Ref. [14].

factor for the initial state particles a and b and the phase space factor for the final state particles in f [10]:

$$d\sigma_{a,b \rightarrow f} = \frac{1}{2E_a E_b |v_a v_b|} |\mathcal{M}(p_a, p_b \rightarrow p_f)|^2 d\Pi_N, \quad (1.47)$$

where the Lorentz invariant phase space factor is given by:

$$d\Pi_N = \left(\prod_f \frac{d^3 p_f}{(2\pi)^3} \frac{1}{2E_f} \right) (2\pi)^4 \delta^{(4)}(p_a + p_b - \sum p_f) \quad (1.48)$$

At a hadron collider as the LHC, interactions will take place between two colliding protons. The dynamics of the protons at scales close to its mass is strongly non-perturbative, so that the interaction between the protons to yield a final state f cannot be directly computed by using perturbative methods. However, thanks to the asymptotic freedom of quarks and gluons and thanks to the fact that the interactions of interest at the LHC will happen at high momentum transfer scales Q^2 , the interaction can be separated into a *hard scattering process*, which takes place between quarks and gluons to yield the final state f and can be computed in perturbation theory, and a low energy process, where the interacting quarks and gluons are extracted from the original protons, which cannot be computed by first principles.

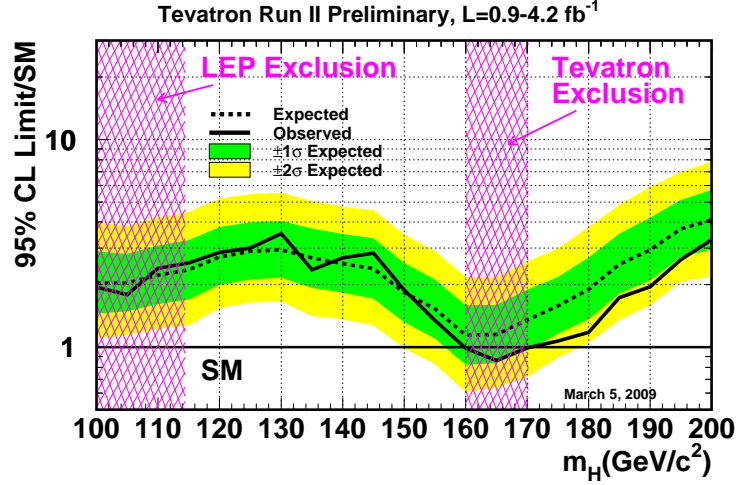


Figure 1.4: Exclusion limits on the presence of a Higgs boson at 95% confidence level, normalised to the expected Standard Model cross section, as a function of the Higgs boson mass, corresponding to the last updated results based on data collected by the Tevatron collider. The LEP2 excluded region, together with the new region excluded by Tevatron data, are also shown as shaded magenta regions in the plot.

The partonic cross section $d\hat{\sigma}_{a,b \rightarrow c}$ defining the *hard scattering process* can be computed according to Eq. 1.47, with a and b being the interacting partons, either quarks or gluons, summing over all their possible spin and colour states. The link between the partons and the original protons is provided by the *parton distribution functions* (PDFs), $f_{a/A}(x_a)$, which, before any QCD correction is taken into account, express the probability that a parton a with energy fraction x_a is extracted from the original incoming proton A . The centre-of-mass energy available for the parton-parton interaction is $x_a x_b \sqrt{s}$, depending on the momentum fraction of the original protons taken by the two partons a and b . The hadronic cross section can therefore be expressed as a convolution of the PDFs with the partonic cross section:

$$d\sigma(A, B \rightarrow f) = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_{a/A}(x_a, \mu_F^2) f_{b/B}(x_b, \mu_F^2) d\hat{\sigma}_{a,b \rightarrow c}(\mu_F^2, \mu_R^2), \quad (1.49)$$

The dependence on the factorisation scale μ_F^2 comes from QCD corrections to the naive parton model introduced by Bjorken and is described by the DGLAP evolution equations [17]. Due to QCD corrections describing quarks and gluons emitted by the initial state partons, also the partonic cross section will acquire a dependence on μ_F^2 . In addition, due to the renormalisation of the strong coupling constant, the partonic cross section will also depend on the renormalisation scale μ_R^2 . If no approximation would have to be made in the computation of the partonic cross section (e.g. neglecting higher order corrections), the dependence on μ_R^2 would disappear and the dependence on μ_F^2 would cancel when combining PDFs and partonic cross section. The μ_F^2 scale in the formula above can in fact be interpreted as the factorisation scale: all contributions at lower scales, including non-perturbative ones, have been absorbed into the PDFs, while all higher scale contributions are described by the partonic cross section. This separation is based on the *factorisation theorem* [18], which is proven only for deep inelastic scattering, but is largely believed to hold also for the case of $2 \rightarrow N$ collisions at

a hadron collider. Due to non-perturbative contributions, the PDFs for quarks and gluons cannot be computed from first principles and need to be measured experimentally: the actual knowledge of the PDFs, on which all predictions for the LHC are based, is mainly coming from the Tevatron and HERA experiments.

In many processes quarks and gluons are also among the produced final state particles. When they are produced in the hard interaction described by the partonic cross section, they are typically emitted at high energies: after radiation (which can be described by higher order QCD corrections and/or by phenomenological models as in the *parton shower* algorithms), they lose energy and, at a scale of the order of Λ_{QCD} , they enter the non-perturbative regime and hadronise. The partonic cross section cannot account for non-perturbative physics contributions: therefore, to correctly describe the transition from partons to hadrons, two possible solutions are available:

- Analogously to the use of PDFs for the initial state partons, fragmentation functions of the form $D_{p \rightarrow h}(x, \tilde{\mu}_F^2)$ can be defined, representing the probability that the final state parton p hadronises into a hadron h , keeping a fraction x of the original momentum of parton p . Again, as in the case of the PDFs, the scale $\tilde{\mu}_F^2$ separates the perturbative contribution of the partonic cross section and the non-perturbative contributions of the fragmentation functions.
- Final state partons can be reconstructed as jets in the detector. Specific algorithms (jet algorithms) are used at reconstruction level to map the final state particles produced in the radiation/hadronisation process to the original underlying final state partons. Even if the parton level picture is not unambiguous, the idea is that the jet algorithm does provide the same result if applied at hadron level (or at reconstruction level in a real detector after unfolding the detector effects) and at parton level (applying the jet algorithm directly to the parton level calculation, which provides also an effective way to regularise infrared and collinear divergences).

The first method is useful in case one is interested in observing specific final state hadrons, as for example in the case of the b -hadron production cross section, where the b -quark fragmentation function comes into play. The second method is more suitable if events with a specific number of jets are observed, regardless of the final state particles produced as a result of the hadronisation process.

Opposite to that, experimentalists usually rely on phenomenological models (*parton shower* and *hadronisation models*) to describe on a statistical basis (by Monte Carlo techniques) first the radiation of the outgoing partons and then their hadronisation into the final state particles. In this way physics events can be described in a fully exclusive way in terms of a complete set of final state particles, which is particularly useful since events of this kind can be used directly as input to the detector simulation, where every single particle needs to be traced through the ATLAS detector.

1.4.2 Luminosity

The expected rate of events for a certain process $R = \frac{dN}{dt}$ can be expressed as the product of the cross section σ and of the *instantaneous luminosity* \mathcal{L} :

$$R = \sigma \cdot \mathcal{L} \quad (1.50)$$

The instantaneous luminosity can be related to geometrical parameters of the collider setup:

$$\mathcal{L} = f \frac{n_1 n_2}{A_{eff}}, \quad (1.51)$$

where n_1 and n_2 are the number of particles contained in the two opposite orbiting bunches, f is the frequency of the collisions and A_{eff} is the effective collision area. The instantaneous luminosity is expressed in units of $[\text{cm}^{-2}\text{s}^{-1}]$. The size of the collision area depends on the beam width in the horizontal (σ_x) and vertical (σ_y) directions; under the assumption of a perfect Gaussian shape it is:

$$A_{eff} = 4\pi\sigma_x\sigma_y \quad (1.52)$$

The number of events expected for a certain process in a given dataset is:

$$N = \sigma \cdot L = \sigma \cdot \int \mathcal{L} dt \quad (1.53)$$

where L is the integrated luminosity over a certain period of data taking.

1.4.3 Higher order corrections

The partonic cross section can never be computed exactly, but only as a perturbation expansion in the coupling constants. As a consequence, the cross section results in an asymptotic series in the electroweak and strong coupling constants. When calculating higher order corrections, infinite contributions arise in the theory: they can be removed from the computation by using the technique of *renormalisation*, where these infinities are reabsorbed in the redefinition of masses, fields and couplings (introducing the dependence on the renormalisation scale μ_R^2). At the LHC, the most relevant higher order corrections to consider are usually provided by QCD corrections, since the coupling constant $\alpha_s \gg \alpha_{EW}$: the value of $\alpha_s(m_Z)$ is ≈ 0.118 , so that the perturbative series in α_s is only slowly converging.

In general the cross section can be schematically written as:

$$\sigma = \alpha_{EW}^\gamma \alpha_s^\delta \sum_{i=0}^{N_{QCD}} \alpha_s^i A_i, \quad (1.54)$$

where the lowest order exponents γ in the EW constant and δ in the QCD will depend on the process considered, while QCD corrections up to the order N_{QCD} are considered. The term with $i = 0$ is usually considered as leading order (LO), the term with $i = 1$ is the next-to-leading order term (NLO) and so on. With current techniques it is possible to calculate all $2 \rightarrow N$ processes with $N \leq 6$ in LO in a fully automatised way (as implemented in the matrix element generators of Sherpa [19] or of MadGraph [20]). Calculations beyond LO are difficult and exist only for a limited number of processes. These kind of NLO or NNLO calculation are defined as *fixed-order* calculations. However, in addition, the impact of eventual electroweak corrections needs to be evaluated process by process.

For example, in the case of the Higgs-strahlung process ($pp \rightarrow VH$, with $V = W, Z$), which will be particularly important for the second part of the present thesis, a QCD NNLO (next-to-next-to-leading order) calculation exists [21]. The leading order process is represented by the Feynman diagram of Fig. 1.5. Up to NLO this process can be factorised into the production of a virtual V boson with momentum k and then in the additional emission of a Higgs boson (if no higher order corrections in α_{EW} are considered). The Higgs-strahlung process is, in this case, absolutely identical to the simple Drell-Yan process producing a single vector boson. The NLO corrections for Drell-Yan processes were already computed a long time ago (see for example Refs. [22, 23]), while the NNLO corrections have been computed in Ref. [24]; some of the diagrams entering the computation of the NLO corrections are shown in Fig. 1.5.

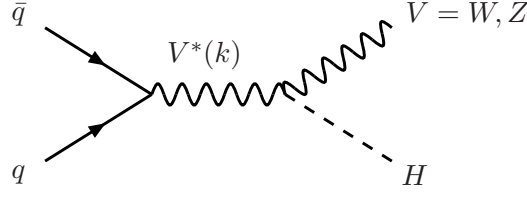


Figure 1.5: Diagram for the Higgs-strahlung process at leading order.

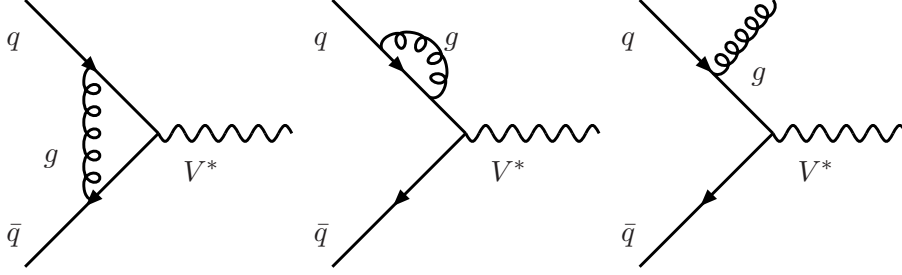


Figure 1.6: Diagrams contributing to the NLO QCD corrections to the vector boson quark-antiquark vertex.

It is common practise to define the K -factor as:

$$K_{NLO} = \frac{\sigma_{NLO}}{\sigma_{LO}} \quad (1.55)$$

and compute the NLO differential distributions as $d\sigma_{NLO} = K_{NLO} \cdot d\sigma_{LO}$. Such a procedure neglects any dependence of the K factor on the kinematics of the produced final states, which is, in the case of virtual corrections, typically a good approximation, but not in the case of large real emission contributions.

The residual dependence of the process from the factorisation and renormalisation scales μ_F^2 and μ_R^2 can be used to evaluate the stability of the perturbative expansion. In the case of μ_F^2 , at higher orders, in general a better cancellation between the dependence in the parton density functions and in the partonic cross section will take place, but also the dependence on the renormalisation scale μ_R^2 will be in general reduced, since the residual dependence on the renormalisation scale will come from terms one order higher in the coupling constant. Analogously to K_{NLO} , also the NNLO K factor can be defined. For the Higgs-strahlung process considered for illustration here, without going into further details of the NNLO computation, the result in terms of K factors for the Higgs-strahlung process and with uncertainty bands corresponding to varying the factorisation and renormalisation scales by $\frac{1}{3}m_{HV} \leq \mu_R(\mu_F) \leq 3m_{HV}$ is shown in Fig. 1.7, which is again taken from Ref. [21]. The convergence of the perturbative series when moving from LO to NLO and then finally to NNLO is remarkable.

In some cases electroweak corrections can be large for LHC processes. This is indeed the case for the Higgs-strahlung process: the order $O(\alpha)$ electroweak corrections were computed in Ref. [25] and combined with the NNLO QCD corrections in Ref. [26]. The result is shown in Fig. 1.8¹: the final uncertainty on the cross section, evaluated both by varying the factorisation and renormalisation scales μ_F^2 and μ_R^2 and by considering the impact of PDF uncertainties, is very small, below $\approx 6\%$.

¹The unphysical behaviour for $m_H \approx 2m_W$ or $m_H \approx 2m_t$ is due to not considering the finite top quark or W boson width.

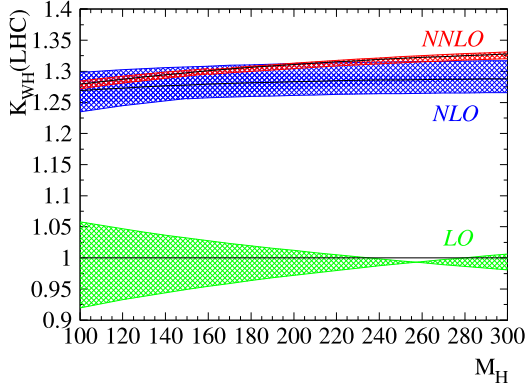


Figure 1.7: K factors corresponding to a QCD NLO and NNLO computation of the Higgs-strahlung process ($pp \rightarrow WH$), with related uncertainty bands.

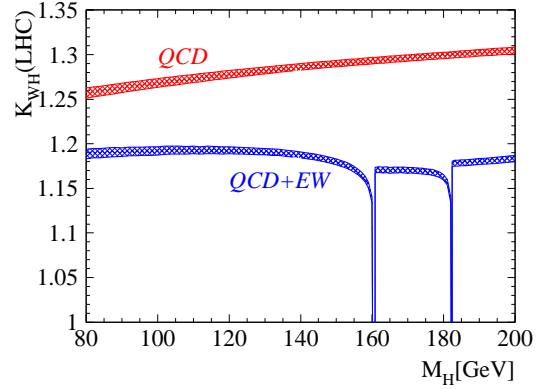


Figure 1.8: K factor corresponding to considering only QCD NNLO corrections and to considering in addition NLO electroweak corrections.

1.4.4 All-order resummation

The result of a given fixed-order calculation is only useful if the sub-leading terms in the perturbative expansion are small. In the case of strong interactions, if a final-state parton representing a gluon or a quark is soft or collinear with respect to another final state quark or gluon, in these regions of the phase space the terms of order α_s^n in the expansion will be in general logarithmically enhanced, i.e. terms of the kind:

$$\sigma = \alpha_{EW}^\gamma \left[\sum_n \alpha_s^n \left(c_{2n}^{(n)} L^{2n} + c_{2n}^{(n)} L^{2n-1} + \dots \right) \right], \quad (1.56)$$

with L being a logarithm of the form $\log \frac{Q^2}{q^2}$, where Q is the scale of the hard interaction (for example of the order of m_{WH} for the Higgs-strahlung production mentioned before), while q is, for example, either the transverse momentum of a parton with respect to another (final state radiation) or the absolute p_T of an outgoing parton (initial state radiation).

In the kinematic configurations where the difference between the two mentioned scales is large, the logarithmic terms L are enhanced and the ratio of two successive terms in Eq. 1.56, which is $\mathcal{O}(\alpha_s L^2)$, can become of order $\mathcal{O}(1)$, spoiling the convergence of the perturbative series.

In the case of b -quarks in the final state, the heavy quark mass suppresses the emission of collinear radiation. However, logarithmically enhanced terms can again arise if $\log \left(\frac{Q^2}{m_b^2} \right)$ is large. This is the case for example for heavy flavour production at a hadron collider [27], where the scale Q^2 is of the order of the p_T of the produced b -quark in the final state, and therefore, since $p_T(b) \gg m(b)$, large logarithms will again appear in the matrix element calculation. Large logarithmic corrections do also appear in one of the most important backgrounds to the Higgs-strahlung process considered in this thesis, $Wb\bar{b}$, in the kinematic regions where $p_T(b\bar{b}) \gg m(b\bar{b})$. In all these cases, the precision of the matrix element calculation can be significantly improved by either resumming analytically the large logarithms or by interfacing the matrix element calculation to a parton-shower algorithm.

Analytic resummation

The first method does rely on resumming (re-ordering) analytically the leading (LL) or next-to-leading (NLL) logarithmic terms in the perturbative expansion at all orders in perturbation

theory [28, 29]. This procedure has the advantage that it contains no approximation, but at the same time it usually does not allow to obtain fully differential cross section predictions and is limited to specific observables.

Parton-shower approach

The parton-shower approach is an alternative to analytic LL resummation. The starting point is again a fixed-order matrix element calculation, which results in n_{ME} final state partons. The starting hypothesis is that the $(n_{ME} + n_{PS})$ -parton final state cross section, in the region of phase space where additional radiation is enhanced (kinematic configurations with logarithmically enhanced collinear and soft radiation), does factorise into the n -parton final state cross section times n_{PS} single $1 \rightarrow 2$ emissions of additional collinear or soft partons. The $1 \rightarrow 2$ emission probability is provided by the *Altarelli-Parisi splitting functions* [17, 30]. One of the shortcomings of the parton-shower approach is that it neglects interference terms between emissions of additional partons coming from two different final state partons: it can however be shown (as derived in Ref. [31]) that such interference terms result in a suppression of radiation happening at angles larger than the angle between the two emitting partons and can be taken into account by ordering the additional radiation in decreasing emission angles.

Starting from the n_{ME} final state partons of the matrix element calculation, the parton shower acts iteratively by causing successive $1 \rightarrow 2$ branchings, driven by an *evolution variable*, which can be either the virtuality of the incoming parton (for Q^2 ordered parton showers), the relative transverse momentum of the two partons (p_T ordered) or the angle between the two partons. The latter two approaches are implemented respectively in recent versions of PYTHIA [32] and in HERWIG [33] and both guarantee the angular ordering of the shower evolution. The scale at which the parton-shower evolution starts is not uniquely determined, but is usually set equal to a scale characteristic of the hard scattering process. The iterative splittings which take place are governed by the so-called *Sudakov factor*, which expresses the probability that no splitting takes place between a starting scale Q_{start} and the splitting scale Q_{split} : based on this probability (which depends on α_s and on the unregularized Altarelli-Parisi splitting functions), a Monte Carlo random number generator is used to generate a possible value of Q_{split} , on top of which the successive branching will be applied. The evolution ends at a cutoff scale Q_{cutoff} (typically 1-2 GeV), below which the physics is expected to enter the non perturbative regime, and the four momenta of the final state partons are passed to the hadronisation algorithm, which maps final state partons into final state hadrons.

This is the procedure which is applied on top of the n final state partons and is therefore called *final state radiation*. In addition, also the two incoming partons do radiate, but in this case the evolution does not proceed forward in time (*forward evolution*), but is traced back from the hard scattering process to the original extraction of the partons from the protons. During the *backward evolution* a correction to the *Sudakov factor* needs to be taken into account, which suppresses branches at large x (fraction of momentum kept after branching) and makes explicit use of the parton density function: the reason for this is that partons at high x are less likely to have undergone branching [31].

The parton-shower algorithm provides an effective way to resum leading logarithmic contributions to all orders in perturbation theory. At the moment no parton-shower algorithm exists which is able to effectively resum next-to-leading logarithmic terms. The parton-shower cannot replace an analytic resummation: it only provides an approximate resummation and it does violate 4-momentum conservation. In a parton-shower approach the 4-momenta are in fact adjusted after the parton-shower has been applied in order to restore 4-momentum conservation. On the other side parton-shower algorithms have a fundamental advantage

with respect to analytic resummation: as already stressed they allow, in combination with a hadronisation model, to obtain fully differential and exclusive predictions by generating a large ensemble of events which asymptotically reflect the expected differential cross section.

It is important to stress that analytic resummation and parton-showers are only correct in describing the kinematic regions where partons tend to be in the collinear or soft limit. Even if these approaches have been in many cases proven to be reliable even beyond the strictly collinear and soft kinematic regions, they will typically break down if an additional hard and well separated parton (jet) is required to be present in the event. In this case the additional parton needs to be described by the matrix element itself and the only solution is to extend the matrix element computation one step further in the perturbative expansion in α_s .

1.4.5 Interfacing higher fixed-order calculations with a parton-shower

One solution to solve this problem is to use the NLO parton level calculation. However interfacing a NLO matrix element calculation with a parton-shower algorithm is not straightforward, since the contribution with real emission diagrams in the matrix element calculation resulting in the terms with $n + 1$ final state partons and the parton shower acting on top of the n final state matrix element calculation (either the LO the NLO virtual correction term) will partially overlap, which leads to double counting. In addition, as in any NLO calculation, divergences arise which need to be cancelled between the real emission and the virtual corrections. Two different prescriptions exist in the literature to perform a proper matching of the NLO calculation with a LL parton-shower algorithm and correctly treat the divergences: they are implemented respectively in the MC@NLO Monte Carlo program [34] and in the new POWHEG method [35]. The main advantage of the latter method over the former is that the matching procedure does not depend on the details of the parton-shower implementation and that no events with negative weights are produced. A NLO generator allows to obtain a more accurate prediction of the total cross section (which reflects QCD NLO accuracy) and, at the same time, it avoids the need of using the parton-shower approximation for one additional jet, so that the differential distributions for the $n + 1$ jet final state is expected to be correct to NLO accuracy. However, a further additional jet in the event, even if hard and well separated from the other jets, will still need to be described by the parton-shower algorithm. A matching procedure between a QCD NNLO parton level calculation and a LL parton-shower would correctly describe $n + 2$ hard and well separated jets and so on. A NNLO Monte Carlo generator of this kind is beyond today's reach.

If one gives up on the NLO accuracy, it is possible to obtain a correct description (in terms of differential cross section) of up to a higher number (at the moment ≤ 6) of hard and well separated additional jets, plus additional soft and collinear radiations from them, by using the CKKW [36] and MLM [37] matching prescriptions. The idea is that, starting from the leading order cross section where n final state partons are produced, final states with $n+1$ hard jets are correctly described by the real diagrams with one additional parton, final states with $n+2$ hard jets are described by real diagrams with two additional partons and so on. Diagrams with no virtual corrections (real emission diagrams with up to 6 final state partons) can be nowadays generated in a completely automated way, within a reasonable CPU time. The consequences of ignoring virtual corrections are not considered, but this affects only the accuracy in the prediction of the rates, which remain LO. Another consequence of not considering virtual corrections is that the real diagrams diverge when the additional produced partons are soft or collinear to other partons: to avoid these singularities, a matching prescription is implemented, so that the region of phase space with hard and well separated partons (where no singularity arises) is described by the matrix element, while soft and collinear emissions are generated by

the parton-shower.

In the case of the Higgs-strahlung process studied in this thesis, the interesting signal and background processes² are typically in a $2 \rightarrow 2$ kinematic configuration and the corrections which are eventually needed (to cross-check the parton-shower approximation or to account for the NLO corrections of the rates) come from $2 \rightarrow 3$ processes, so that a NLO Monte Carlo, based either on MC@NLO or POWHEG, would be the ideal tool to accurately describe these processes. Unfortunately no such NLO Monte Carlo exists for one of the main backgrounds³ processes, which is the reason why the study presented in this thesis relies on LO Monte Carlo generators only. Due to the jet veto implemented in the analysis, high multiplicity final states are not expected to play an important role and therefore performing a matrix element / parton shower matching using the CKKW or MLM prescriptions is not expected to improve the description of the process.

1.4.6 Hadronisation

The output of a parton-shower Monte Carlo produces a set of quarks and gluons, with virtuality of the order of the cutoff scale Q_{cutoff} . At this scale the perturbative calculation is matched to a phenomenological description of the non-perturbative transition from partons to hadrons. A fixed scale Q_{cutoff} is typically used for the matching and for the tuning of the hadronisation model. The idea is, however, that the hadronisation model is nearly universal and independent on the perturbative physics above Q_{cutoff} .

The two most used hadronisation models are the *cluster model* and the *string model*: both describe the experimental data very well (after appropriate tuning of the model parameters), accounting correctly for the correlations between the different colour-connected partons and are implemented respectively in the HERWIG and in the PYTHIA Monte Carlo generators.

The basic idea of the string model [38] is that a quark antiquark pair is connected by a colour string. When the quark and antiquark move further apart, the potential energy of the colour field between them increases linearly with their distance and the colour field collapses in a string-like configuration. While the energy increases, $q\bar{q}$ pairs are spontaneously generated out of the string, which then form the final state hadrons. If gluons are present, they produce kinks along the strings, which initially preserve the gluon momenta. A string is therefore characterised by its endpoints (the quarks) and a certain number of kinks (the gluons). In principle the assignment of strings to the final state partons is not uniquely defined: however it can be shown, at least in the leading colour N_c approximation, that the leading contribution comes from strings stretching between the closest opposite sign partons, whereas further string connections are suppressed by a factor $\frac{1}{N_c^2}$.

The cluster model relies on the formation of colour-singlet clusters of partons, which then subsequently decay into the observed hadrons. This happens most easily by non-perturbative splitting of gluons into $q\bar{q}$ pairs. Colour-singlet clusters with masses of up to a few GeV can then be seen as superpositions of meson resonances and therefore decay into mesons and baryons, where their relative fraction is typically well reproduced by the clusterization dependence on the available energy density and phase space.

1.4.7 Underlying event

The underlying event is defined as the outcome of additional parton interactions taking place among components of the same protons which give rise to the hard scattering process of

²These processes are the signal (W/ZH), the diboson ($WW/ZZ/WZ$), top ($t\bar{t}$) and $W/Z + jj$ background (dominated by $W/Zg \rightarrow W/Zb\bar{b}$).

³This background is the $Wb\bar{b}$ or, more in general, the Wjj process.

interest. The problem is both that the additional parton-parton interactions result in very low p_T scattering processes, which are difficult to account for by using perturbative methods only. Different multiple parton interaction (MPI) models exist on the market. One of these models, implemented in PYTHIA, is described in Ref. [39]. This model assumes that multiple interactions take place in a substantially independent way (except for the effect of proton momentum conservation), so that Poisson statistics can be used, and that they can be described by using perturbative QCD above a certain scale p_T^{\min} (since the perturbative cross section will diverge for $p_T \rightarrow 0$) and by a simple two string model at lower p_T . The chosen scale p_T^{\min} is one of the main parameters of the model. A dependence on the proton-proton impact parameter is also introduced (as an option), since a lower impact parameter is equivalent to an increased overlap of the two protons in the collision and therefore increases the amount of multiple parton interactions. An alternative model is implemented in the JIMMY Monte Carlo program, which is described in detail in Ref. [40]. This model is usually used in combination with the HERWIG Monte Carlo program.

1.4.8 Pile-up

At LHC the hard scattering events of interest will originate from a single proton-proton interaction out of the collision of two bunches filled with protons. While it is extremely unlikely that more than one colliding proton-proton pair will produce a high p_T hard scattering event when the collision of two proton bunches takes place, in general several additional low p_T interactions will take place among other proton-proton pairs in the same bunches. These are usually called *minimum bias* events (or pile-up events), since they can be most easily experimentally detected under minimal trigger requirements: under these conditions in general no hard scattering event will take place (since it is extremely unlikely) and all interactions will come from low p_T scattering processes.

The minimum bias interactions are completely independent from the hard scattering process and are therefore easier to describe than the underlying event, even if they also suffer from the perturbative approach breaking down at very low p_T . The scattering process at very low p_T can be distinguished in single diffractive, double diffractive and non diffractive inelastic events [41]. Even if this is only an approximation, only the last two were considered in ATLAS so far (see Ref. [42]) and a cross section of ≈ 70 mb was assumed, which is affected by sizable theoretical uncertainties, due to the fact that the extrapolation from Tevatron energies to the LHC is based on model assumptions. If the size of the interacting proton bunches is sufficient with respect to the experimental resolution, the additional interactions from minimum bias events can be spatially distinguished from the hard scattering event: this is the reason why the experimental event-by-event reconstruction of the primary interaction vertices out of their charged decay products, which will be the object of the first part of the present thesis, is an important experimental handle to suppress the contamination from pile-up events.

In addition to the *in-time* pile-up events considered here, where additional pile-up interactions are taking place nearly at the same time of the hard scattering process (out of the interactions of the same proton bunches), also *out-of-time* pile-up events need to be considered. In fact, as will be discussed in the next section, the time interval between two consecutive bunch collisions at the LHC is, at design, only 25 ns, which is shorter than the average response time of many of the subsystems of the ATLAS detector, so that the time information alone is not sufficient to separate these events.

2 Higgs phenomenology

One of the main aims of the LHC is to elucidate whether the mechanism of symmetry breaking postulated by the Standard Model of fundamental interactions and elementary particles is really realised in nature. This question can be answered by either finding the Higgs boson and measuring its properties or by excluding that it does exist over the full mass range to which its existence is limited by the actual theoretical and experimental bounds described in Section 1.3.

2.1 Higgs decays

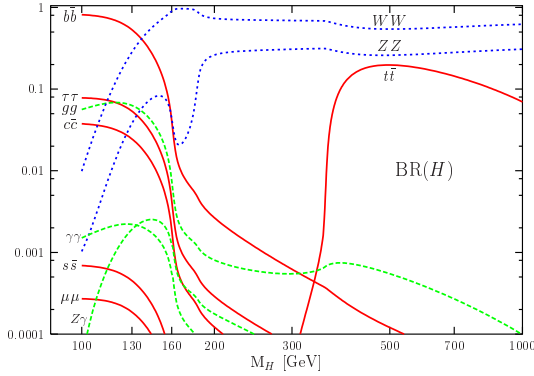


Figure 2.1: Branching ratio of the Higgs boson decaying in the different possible channels, as a function of the Higgs boson mass M_H . Taken from Ref. [13].

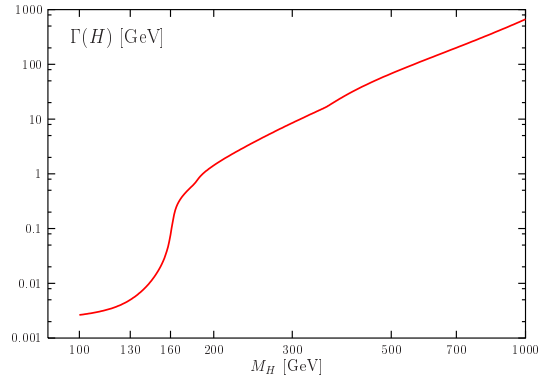


Figure 2.2: Total Higgs boson width as a function of the Higgs boson mass M_H . Taken from Ref. [13].

The Higgs boson is an unstable particle, which can be detected only through its decay products. The Higgs boson couples directly to different particles, opening up the possibility of having several decay channels. The production and the decay of a Higgs boson can be factorised under the *narrow width approximation*. Under this hypothesis the Higgs boson width is sufficiently small with respect to its mass that non-resonant contribution can be neglected and the mass of the outgoing Higgs boson can be considered on-shell, which leads to the above mentioned factorisation. The fraction of Higgs bosons decaying into a certain decay channel is called *branching ratio* and depends on the partial and total decay widths:

$$\text{BR}(H \rightarrow f) = \frac{\Gamma_f}{\Gamma_{\text{TOT}}}, \quad (2.1)$$

where Γ_{TOT} is the total decay width. The branching ratios of the various channels depend on the Higgs boson mass m_H and the result of the computation of the branching ratios is shown in Fig. 2.1, together with the total decay width, which is illustrated in Fig. 2.2. All decay widths are obtained by including the available QCD and electroweak corrections (a good overview is provided by Ref. [13]), as contained in the HDECAY program [43].

Given the Higgs couplings introduced in Section 1.2.2, the Higgs boson decays at tree level into pairs of electroweak gauge bosons ($H \rightarrow W^+W^-, ZZ$) and into pairs of quarks and leptons ($H \rightarrow q\bar{q}, l^+l^-$). Due to loop corrections, it can also decay into two photons ($H \rightarrow \gamma\gamma$), two gluons ($H \rightarrow gg$) or a γZ pair ($H \rightarrow \gamma Z$). The branching ratios of the Higgs boson in these channels have a significant different behaviour for a light Higgs boson ($m_H < 130 - 140$ GeV) and a heavy Higgs boson ($m_H \geq 130 - 140$ GeV).

At low Higgs boson masses the $H \rightarrow b\bar{b}$ decay channel dominates ($\text{BR} \approx 67\%$ with $m_H = 120$ GeV), followed by the $H \rightarrow \tau\tau$ channel ($\text{BR} \approx 6.9\%$ with $m_H = 120$ GeV), which is around one order of magnitude lower. The $H \rightarrow gg$ decay plays no role at a hadron collider, since it will be impossible to separate it from the huge background from multi-jet QCD events. The $H \rightarrow \gamma\gamma$ decay channel has a very low branching ratio ($\text{BR} \approx 0.2\%$ with $m_H = 120$ GeV), but the very good experimental mass resolution allowed by the $\gamma\gamma$ final state and the relatively low rate of quark or jets fragmenting into high energy isolated photons allows the extraction of a signal in the presence of very large continuum backgrounds.

At high Higgs boson masses the dominant decays are $H \rightarrow W^+W^-$ and $H \rightarrow ZZ$, in particular for Higgs boson masses above the respective $2m_W$ and $2m_Z$ thresholds, while, above $2m_t$, also the $H \rightarrow t\bar{t}$ decay channel opens up. There is also an intermediate region, below the $m_H = 2m_W$ mass threshold, where the three-body decays into WW^* or ZZ^* become important, due to the large HWW and HZZ couplings, even if they are suppressed by the fact that one of the two W or Z bosons are off-shell.

2.2 Production modes at the LHC

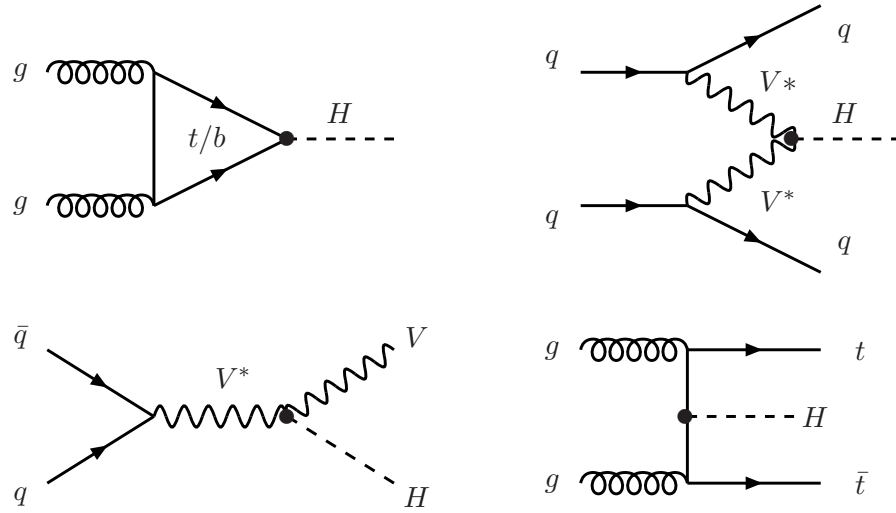


Figure 2.3: The dominant Standard Model Higgs boson production mechanisms at the LHC: gluon fusion (top left), vector boson fusion (top right), Higgs-strahlung (bottom left) and associated production with a top antitop quark pair (bottom right).

The parton level processes which allow the production of a Higgs boson at a hadron collider and, more specifically, at the LHC are shown in Fig. 2.3. Their cross section, assuming the nominal centre-of-mass energy of 14 TeV, are shown in Fig. 2.4.

The leading production mode is gluon-gluon fusion, $gg \rightarrow H$, which, even if loop induced, is enhanced due to the heavy top-quark loop, followed by the weak boson fusion, $q\bar{q} \rightarrow q\bar{q}H$,

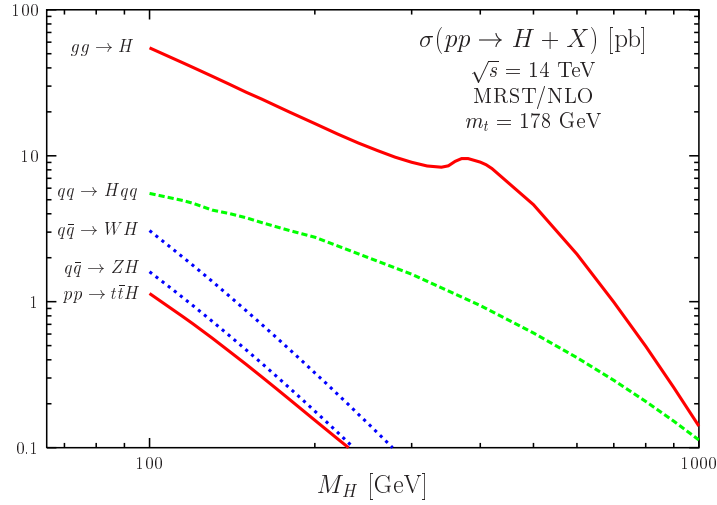


Figure 2.4: Cross sections for the main Higgs boson production modes at the LHC computed at NLO, as a function of the Higgs boson mass (taken from Ref. [13]).

where the Higgs boson is produced in association with two forward jets, with large rapidity gap between them. Other interesting processes, in decreasing order of cross sections, are the Higgs-strahlung process, where the Higgs boson is produced in association with either a W or Z boson ($q\bar{q} \rightarrow VH$ with $V = W, Z$), and finally the associated production with a pair of top antitop quarks ($gg \rightarrow t\bar{t}H$).

Since the centre-of-mass of the parton-parton interaction producing the Higgs boson must be of the order of the Higgs boson mass and the centre-of-mass energy of the interacting protons is $\sqrt{s} = 14$ TeV, in general very small values of the parton momentum fractions x are probed. In this region the probability to extract a gluon out of a proton (gluon PDF) is significantly higher than the corresponding probability to extract a quark or antiquark. This is the reason why in general processes initiated by gluons, as the gluon fusion process, are significantly enhanced with respect to processes initiated by quarks or antiquarks.

2.3 ATLAS Discovery potential

In principle the detection of a Higgs boson in one of the two multipurpose detectors at the LHC (CMS or ATLAS) can be carried out in any combination of Higgs boson production mode and decay channel. For each of these combinations, the expected rate is given by the product of the cross section and the branching ratio ($\sigma \times \text{BR}$), times the instantaneous luminosity. The rate alone however does not provide any indication of whether the specific Higgs boson search channel considered is sensitive to the eventual presence of the Higgs boson: the sensitivity is instead condensed in the expected discovery significance. The significance of a specific search channel does mainly depend on how cleanly the signature provided in the detector can be extracted from very similar signatures provided by other, typically much more copiously produced, background events. This has been studied in detail during the last 10-15 years for many possible search channels, using increasingly realistic Monte Carlo simulations (both from the point of view of event generation and detector simulation): the most recent results produced by the ATLAS Collaboration are shown in terms of discovery significance for the various search channels in Fig. 2.5.

For Higgs boson masses above $m_H \approx 130$ GeV and below $m_H \approx 450$ GeV, assuming a centre-

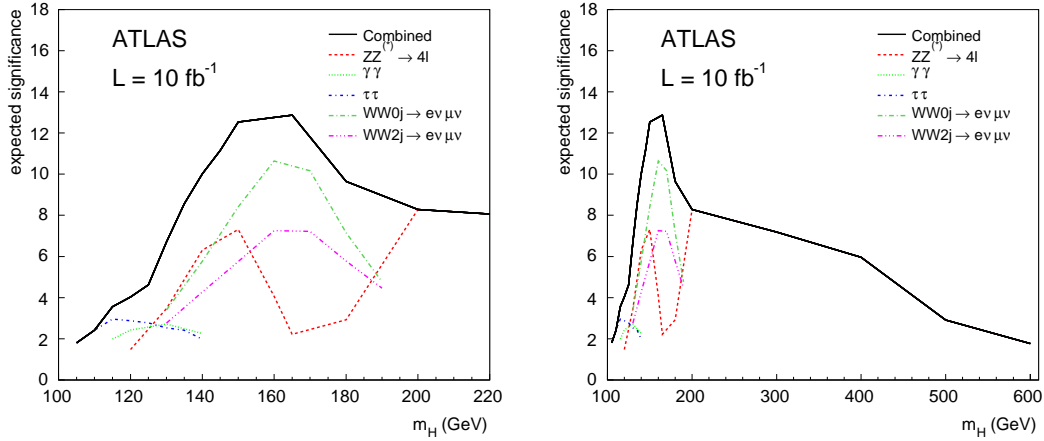


Figure 2.5: ATLAS discovery potential of a Standard Model Higgs boson as a function of the Higgs boson mass, corresponding to an integrated luminosity of 10 fb^{-1} (taken from Ref. [44]).

of-mass energy of $\sqrt{s} = 14 \text{ TeV}$, a Standard Model Higgs boson can be discovered with the ATLAS Detector after having collected at least 10 fb^{-1} of data. The most significant channels are $H \rightarrow W^+W^-$ both in gluon and in vector boson fusion and $H \rightarrow ZZ$. The $H \rightarrow ZZ$ is also the most suitable channel to extend the Higgs searches to Higgs boson masses of up to very high masses. It should be noticed that two important Higgs boson channels, namely $H \rightarrow WW \rightarrow (e\nu e\nu, \mu\nu\mu\nu)$, have not yet been considered in the combination and will be added in a second step.

Below $m_H \approx 130 \text{ GeV}$, the situation is more difficult and the discovery sensitivity to a Standard Model Higgs boson needs to rely on the combination of different channels, in particular $H \rightarrow \gamma\gamma$ and $q\bar{q}H \rightarrow q\bar{q}\tau^+\tau^-$. For the $H \rightarrow \gamma\gamma$ channel only the inclusive analysis (based mainly on the gluon fusion production mode) was considered for the combination. The low mass region is certainly where the LHC will have the hardest time to discover a Higgs boson and a higher amount of integrated luminosity (but less than 30 fb^{-1}) may be needed to claim a 5σ discovery.

Excluding the presence of a Higgs boson corresponding to a certain mass is nearly as important as discovering that a Higgs boson of a certain mass exists. The amount of integrated luminosity required to exclude a Higgs boson of a certain mass, with a certain confidence level, is shown in Fig. 2.6, for Higgs boson masses up to 140 GeV .

2.4 Higgs decays to b -quarks

Given that the discovery of a Higgs boson with a mass close to the LEP2 limit of 114.4 GeV is more challenging as compared to higher Higgs boson masses, it would be desirable to be able to rely on additional search channels, both to increase the discovery potential and to make the overall search for a light Higgs boson more robust.

At low Higgs boson masses, as already shown in Fig. 2.1, most of the Higgs bosons decay into a pair of b -quarks, which are reconstructed as heavy flavour jets in the detector. Thanks to the experimental ability to identify b -quark jets (b -tagging), these decays may become accessible, however several problems arise:

- The presence of b -quark jets in the event cannot be used as a criterion for the on-line

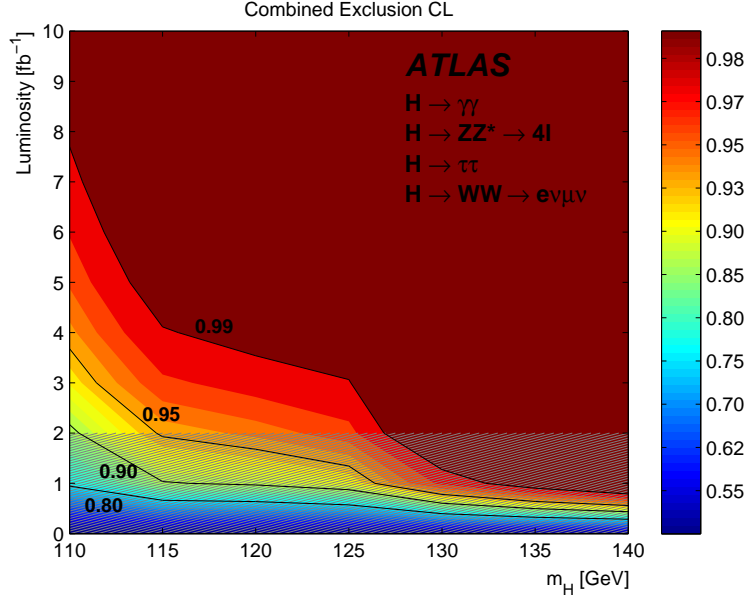


Figure 2.6: Expected integrated luminosity required for ATLAS to exclude a Higgs boson mass m_H at a confidence level given by the corresponding colour. The area below 2 fb^{-1} indicates the region where the large statistics limit, on which these exclusion limits are based, is not expected to hold (taken from Ref. [44]).

selection of events (at least not at the first trigger level). As a result only the production modes with additional signatures, such as a top-antitop quark pair in the $t\bar{t}H$ production or a W or Z boson in the VH (with $V = Z, W$) production mode, resulting in leptonic decays, can be really efficiently triggered on.

- Even in the $t\bar{t}H$ and VH search channels, the search for the Higgs boson suffers from large backgrounds from events with light quark (u, d, s) or gluon jets, like respectively $t\bar{t}jj$ and Vjj . In order to reject them efficiently, the b -jet identification algorithm needs to reject light- or gluon jets with a misidentification efficiency of the order of 10^{-2} ($\approx 10^{-4}$ for a pair of jets), keeping at the same time a reasonable fraction of b -jets.
- Background events with real b -jets produced by strong interactions will be copiously produced at the LHC. These provide an irreducible background which cannot be rejected by applying the b -jet identification algorithm. A good $b\bar{b}$ di-jet invariant mass resolution is the most important handle against this background. For the $t\bar{t}H$ and VH search channels these backgrounds are respectively $t\bar{t}b\bar{b}$ and $Vb\bar{b}$.

Due to these reasons, it is clear that the $gg \rightarrow H \rightarrow b\bar{b}$ search channel is impossible to access at the LHC: the backgrounds from QCD di-jet background events are too large. However also the $q\bar{q} \rightarrow q\bar{q}H \rightarrow q\bar{q}b\bar{b}$ search channel was not mentioned above: in this search channel the vector boson fusion signature does in principle provide a handle to significantly reduce the large backgrounds from multi-jet QCD events. A realistic parton level study was performed for the first time in Ref. [45]. Since this search channel relies entirely on hadronic final states, tight selection criteria need to be applied on the forward jets produced in association with the Higgs boson in order to be able to select these events at trigger level. The main source of background contamination comes from irreducible QCD 4-jet events, $b\bar{b}jj$. The discovery potential of the channel with an integrated luminosity below 30 fb^{-1} is therefore found to be very marginal. At higher integrated luminosities, the effect of high luminosity pile-up should

be considered, before any realistic statement can be made. Recently a new search channel in vector boson fusion, $q\bar{q} \rightarrow q\bar{q}H \rightarrow q\bar{q}b\bar{b}\gamma$, with the addition of a photon in the final state, has also been proposed [46]. Even if the signal is reduced by around two order of magnitudes, the signal-to-background ratio does considerably improve, due to the destructive interference between various terms with central photon emission in the QCD 4-jet plus photon background ($b\bar{b}jj\gamma$). Experimental studies are on-going (they rely on the photon for the trigger selection of events), but, even according to the parton level study, a high amount of integrated luminosity ($\approx 100 \text{ fb}^{-1}$) will be needed to get a sensitivity of up to 3σ .

The remaining two (more promising) search channels with Higgs bosons decaying into b -quarks are $t\bar{t}H$ and W/ZH . The $t\bar{t}H$ channel was found to be very promising in the first physics potential studies performed at the time of the ATLAS Technical Design Report (TDR) [42]. However with a more realistic estimate of the two main backgrounds ($t\bar{t}jj$ and $t\bar{t}b\bar{b}$), and with a more realistic detector simulation [44], the expected statistical significance with an integrated luminosity of 30 fb^{-1} has decreased to $\approx 2\sigma$. Even worse, when a realistic uncertainty on the background level is considered, the significance is found to be marginal ($\approx 0.5\sigma$). On the theory side a new parton level study has just been performed in order to try to rescue the sensitivity in this channel, by requiring the Higgs boson and one of the two top quarks to have high transverse momenta [47]. The result is that with 30 fb^{-1} a statistical significance of $2.2\text{-}2.6\sigma$ can be obtained, slightly higher than in inclusive analysis, but with an improved signal-to-background ratio; however, the main advantage of this analysis is that, due to a significantly reduced combinatorial background from the signal and smoother $b\bar{b}$ mass distributions from the backgrounds, it should be possible to estimate the background contribution using the $b\bar{b}$ mass sidebands and the systematic uncertainty on the background level, therefore, should be strongly reduced.

A detailed discussion of the W/ZH search channels will be the subject of the second part of the present thesis, in Chapters 7 and 8, where a novel idea about how to improve the sensitivity in these channels is exploited [1] and a first complete experimental study of these channels is performed, in order to demonstrate that the W/ZH search channels can indeed be used to recover a real sensitivity to a light Higgs boson decaying to b -quarks.

One of the crucial ingredients of these low mass Higgs boson searches is the ability to identify the b -quark jets with high efficiency and low misidentification rates. Quite some effort has been spent in the course of the present thesis to improve the b -jet identification performance, by developing new algorithms based on the explicit detection of secondary vertices, as discussed in detail in Chapter 6.

2.5 Measurement of Higgs couplings

Once the Higgs boson has been discovered, it will be important to measure its properties in order to make sure that what has been observed does really correspond to the Standard Model Higgs boson. These properties include a more precise determination of the Higgs boson mass, of the width of the Higgs boson resonance (at least in the mass range where the experimental resolution does allow to resolve it), the measurement of the spin and CP eigenstate, the determination of the Higgs boson couplings and, even if it seems not to be possible at the LHC, the Higgs Boson self-couplings. In this section only the measurement of the Higgs boson couplings for a light Higgs boson will be briefly discussed, for which the $H \rightarrow b\bar{b}$ search channels play a fundamental role, while a detailed discussion of the determination of the remaining properties can be found for example in Refs. [13, 48].

A recent detailed study of the determination of the Higgs boson couplings for a Higgs boson with $m_H = 120 \text{ GeV}$ has been presented in Ref. [49], which partially updates (for this specific

Higgs boson mass) the results presented in Refs. [50] and [51]. The result of this study will be briefly discussed here.

In general it is possible to translate the observed number of signal events for a certain Higgs boson production and decay mode directly into a measurement of the corresponding $\sigma \cdot \text{BR}$, by taking into account a proportionality factor which depends on the integrated luminosity and on the signal acceptance. Up to a good approximation,

$$\sigma \cdot \text{BR} \propto \frac{\Gamma_i \Gamma_f}{\Gamma_{\text{TOT}}}, \quad (2.2)$$

where Γ_i and Γ_f are the partial decay widths involving respectively the Higgs couplings at production and decay and Γ_{TOT} is the total Higgs boson decay width. By exploiting all search channels which are actually considered promising in ATLAS or, more in general, at the LHC, it is possible to constrain the Higgs couplings g_{jjH} to fermions or vector bosons ($j = f, v$), whose expression was introduced in Section 1.2.2.

Since the main idea is to look for deviations from the values of the Higgs couplings as predicted by the Standard Model, the Higgs couplings are parametrised as:

$$g_{jjH} = g_{jjH}^{SM}(1 + \Delta_{jjH}), \quad (2.3)$$

so that a measurement resulting in $\Delta_{jjH} = 0$ expresses perfect compatibility with the Standard Model. Due to the fact that the couplings in most of the cases enter the partial widths in squared form, a sign ambiguity can arise, so that, in the case the Standard Model, a second solution at $\Delta_{jjH} = -2$ can be found. The total Higgs width is allowed to change roughly in the same way:

$$\Gamma_{\text{TOT}} = \Gamma_{SM}(1 + \Delta\Gamma) \quad (2.4)$$

	σ
Δ_{WWH}	± 0.24
Δ_{ZZH}	± 0.46
Δ_{ttH}	± 0.59
Δ_{bbH}	± 0.64
$\Delta_{\tau\tau H}$	± 0.57
$\Delta_{\gamma\gamma H}$	± 0.55
Δ_{ggH}	± 0.80

Table 2.1: Errors expected on the Higgs couplings after collecting 30 fb^{-1} of integrated luminosity, based on a single experiment. Symmetric errors are assumed here. Effective ggH and $\gamma\gamma H$ couplings are included in the analysis. Taken from Ref. [49].

The result of the the global fit to the Higgs couplings, in terms of errors achievable on the deviations of the couplings from the Standard Model, is illustrated in Table 2.1, which shows that a sufficiently accurate determination of the most important Higgs couplings to fermions and bosons is indeed possible already with 30 fb^{-1} , based on a single experiment.

One of the most important assumptions made in this analysis is that the Higgs boson coupling to b -quarks can be accessed via the $W/ZH \rightarrow W/Zb\bar{b}$ search channels. This assumption is based on the hadron level study of Ref. [1]. If this would not be the case, the ability to constraint not only the Higgs boson coupling to b -quarks, but also the remaining Higgs couplings, would be strongly decreased, as shown in Fig. 2.7. In fact, after removing this search channel, there is essentially no constraint anymore on the $g_{Hb\bar{b}}$ coupling and also the precision

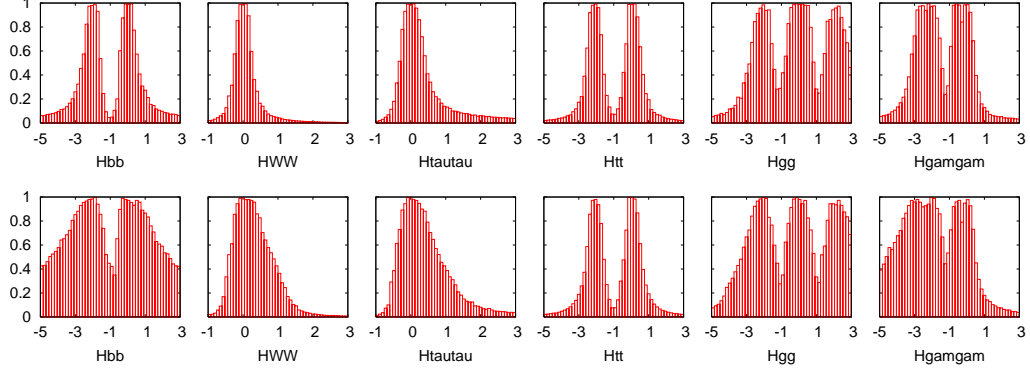
$1/\Delta\chi^2$


Figure 2.7: Profile likelihood probability distributions in terms of $1/\Delta(\chi^2)$ for the determination of the Higgs couplings, obtained first by including also the WH/ZH channels (upper row) and then by including only the $t\bar{t}H$ channel (lower row), both with a decay $H \rightarrow b\bar{b}$. The variable $\Delta(\chi^2)$ expresses the difference of the profile likelihood ratio value as a function of the Higgs coupling of interest with respect to the value which maximises the profile likelihood ratio, using the χ^2 approximation in the vicinity of the maximum. All experimental and theory errors are included. Taken from Ref. [49].

with which also the other couplings (like g_{HWW}) can be determined is significantly reduced. The reason for this is that all observed rates, as illustrated in Eq. 2.2, are sensitive to the product of two partial widths divided by the overall Higgs boson total width. The total width is dominated by the partial decay width into b -quarks, so that, if this is not constrained, it is possible to arbitrarily scale down or up all observed rates, which makes the system of equations substantially less constrained.

This confirms that, if a Higgs boson with a mass around 120 GeV is present in nature, the WH/ZH channels with $H \rightarrow b\bar{b}$ are absolutely crucial to measure the Higgs boson couplings and make sure the couplings agree with the Standard Model predictions. The detector level study presented in this thesis can be used to obtain a more realistic estimate of the parameters to be used as input to the measurement of the Higgs boson couplings.

3 The ATLAS Experiment and its environment

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [52] is a hadron accelerator and collider installed in the existing 26.7 km long tunnel that was previously constructed to host the LEP machine. Being a particle-particle rather than a particle-antiparticle collider, the LHC collider consists of two rings with counter-rotating beams. Due to the limited size of the LEP tunnel, a *two-in-one* design was chosen for the superconducting magnets needed to keep the accelerated particles along the curved trajectory, so that the magnets provide a magnetic field in opposite directions for the two nearby lying accelerated particle beams. The LHC is mostly a proton-proton collider, but dedicated runs are also foreseen with lead-lead ion collisions. The ATLAS and CMS experiments are located at two opposite points of the LEP tunnel, which are called respectively Point 1 and Point 5. Their operation will depend in an essential way on the parameters of the beams the LHC collider will deliver.

The maximum beam energy that the LHC can deliver depends on the maximum magnetic field of the dipole magnets needed to keep the particles along the trajectory: the nominal field is 8.33 T, which corresponds to a beam energy of 7 TeV. This corresponds to collisions at an unprecedented centre of mass energy of $\sqrt{s} = 14$ TeV, which will permit to significantly extend the physics reach in terms of sensitivity to the Higgs boson and to eventual new physics phenomena with respect to previous experiments.

Apart from the beam energy, a second fundamental parameter is the delivered instantaneous luminosity. This depends on the parameters of the LHC machine and on the configuration of the magnets (mainly quadrupoles) in the proximity of the experiments, which have to focus the beams into the point where the collisions take place. The protons will be accelerated along the LHC rings in both directions in the form of spatially distinct bunches, the nominal distance being 25 ns, which corresponds to 7.5 m. When they are brought to collision, they are not collided head on, but a small crossing angle of the order of 150-200 μrad is adopted to avoid the occurrence of parasitic collisions. The formula expressing the luminosity as a function of the geometrical characteristics of the colliding bunches and of the machine parameters (Eq. 1.51 in Section 1.4.2) can be expressed in more detail as:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F, \quad (3.1)$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, β^* the beta function at the collision point and F the geometrical luminosity reduction factor due to the crossing angle at the interaction point (IP):

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right)^{-\frac{1}{2}}, \quad (3.2)$$

where θ_c is the full crossing angle at the IP, σ_z the RMS of the bunch length and σ^* is the RMS of the transverse beam size at the IP. The expected peak design luminosity, together with the main accelerator parameters are shown in Table 3.1. The parameters are shown for injection and collisions at design luminosity. In addition, even if the evolution of the instantaneous luminosity will depend on how the commissioning of the machine will proceed, some of the machine parameters foreseen for the first physics run at $E_{beam} = 3.5$ TeV are also shown.

In addition to the beam energy and to the delivered luminosity, the space-distribution of the interaction region at the centre of the ATLAS Detector, which is usually denoted as *beam spot* (BS), is very important, in particular when additional low p_T parton-parton interactions from pile-up are considered. In fact these additional pile-up interactions will distribute randomly across the interaction region. Given a head-on collision of the two beams (with zero crossing angle) and under the hypothesis of Gaussian distributed protons inside a bunch, the space-distribution of the interaction region is approximately the product of the spacial distributions of the protons in the two bunches:

$$P(\vec{r}_{int}) = P_{bunch1}(\vec{r}) \cdot P_{bunch2}(\vec{r}), \quad (3.3)$$

so if the protons in the two bunches are Gaussian distributed in the same way, the product will result in another Gaussian distribution, with:

$$\sigma_{BS} = \frac{\sigma_{bunch}}{\sqrt{2}}. \quad (3.4)$$

Crossing angles of the order of 300 μrad , as in the nominal running scenario, leave the transverse size of the beam spot essentially unaffected, but slightly shorten the longitudinal beam spot size. The interaction region is therefore described by a Gaussian probability distribution, which at nominal luminosity has a standard deviation of $\sigma_{x,y}^{BS} = 12 \mu\text{m}$ in the transverse plane and $\sigma_z^{BS} = 45 \text{ mm}$ in the longitudinal direction (along the beams). During the first physics run at $\sqrt{s} = 7$ TeV expected during end of 2009 - autumn 2010 (see Table 3.1), the beam spot is expected to be significantly different, with a standard deviation of $\sigma_{x,y}^{BS} = 32 \mu\text{m}$ in the transverse plane and of 42 mm in the longitudinal plane.

In all studies presented in this thesis the beam energy will be assumed to be $\sqrt{s} = 7$ TeV, as in the nominal scenario. The beam spot in the simulation is assumed to have a size of $\sigma_{x,y}^{BS} = 15 \mu\text{m}$ and $\sigma_z^{BS} = 55 \text{ mm}$.

The amount of pile-up events crucially depends on the instantaneous luminosity and on the bunch crossing spacing. The number of expected pile-up events per bunch crossing, based on the rate definition of Eq. 1.50 of Section 1.4.2:

$$R = \sigma_{MB} \cdot \mathcal{L} \cdot \frac{\Delta t_{BC}}{1 - f_{empty}}, \quad (3.5)$$

where Δt_{BC} is the time interval between two bunch crossings and f_{empty} is the fraction of empty bunches, which is assumed to be 20%. As a result, the average number of pile-up interactions expected per bunch crossing at design luminosity, under the assumption that $\sigma_{MB} = 70 \text{ mb}$, is $\mu \approx 23$. During the first years of LHC running, after the start-up phase, a long run at a luminosity of $\approx 2 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ is expected. This is usually referred to as the *low luminosity* scenario, on which most of the physics analyses to be performed in the first years of ATLAS, including the Higgs analysis presented in this thesis, are based. In this case, only an average of $\mu \approx 4.6$ pile-up events per bunch crossing are expected.

During the start-up phase of the LHC, when the instantaneous luminosity will be progressively increased, it might be easier to first increase the number of particles per bunch than

		Injection	Collision	First physics run
Energy		450 GeV	7000 GeV	3500 GeV
Luminosity	nominal		$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$?
	ultimate		$2.5 \cdot 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	-
Number of bunches		2808		?
Bunch spacing		25 ns		$\geq 50 \text{ ns}$
Number of particles per bunch	nominal	$1.15 \cdot 10^{11}$?
	ultimate	$1.7 \cdot 10^{11}$		-
Beam current	nominal	0.58 A		?
	ultimate	0.86 A		-
Transverse emittance (rms,normalised)		$3.5 \mu\text{m}$	$3.75 \mu\text{m}$	$3.75 \mu\text{m}$
Longitudinal emittance		1.0 eVs	2.5 eVs	$\approx 1 \text{ eVs}$
Bunch length, total (4σ)		1.7 ns	1.0 ns	$\approx 0.8 \text{ ns}$
Synchrotron radiation loss per turn		0	7 keV	?
Relativistic gamma		479.6	7461	3730.26
β at IP1 and IP5		18 m	0.55 m	1-3 m
Half crossing angle at IP1 and IP5 ($\theta_c/2$)		$\pm 160 \mu\text{rad}$	$\pm 142.5 \mu\text{rad}$	0 or $\pm 150 \mu\text{rad}$
Geometrical luminosity reduction factor F		-	0.836	0.98 – 1
RMS bunch length		11.24 cm	7.55 cm	5.41 cm
RMS beam size at IP1 and IP5 (σ^*)		$375.2 \mu\text{m}$	$16.7 \mu\text{m}$	$45 \mu\text{m}$

Table 3.1: Main machine parameters of the LHC collider in the proton proton collision mode.

to reduce the time interval between two following bunches. While this will probably permit to reach higher luminosities in a shorter time, it will also substantially increase the number of expected pile-up interactions. This means that a significant contamination from pile-up events may be expected even for instantaneous luminosities below $\mathcal{L} = 10^{32} \text{ cm}^{-2}\text{s}^{-1}$. One of these scenarios will be considered for the primary vertex finding studies presented in this thesis (Chapter 5).

3.2 The ATLAS detector

The high energy and luminosity delivered by the LHC collider will produce collisions in the ATLAS detector at high interaction rates, with very high radiation doses, in particular in the proximity of the interaction region, where processes with high particle multiplicities and energies will be produced. In spite of this difficult environment, the search for the Higgs boson and for new physics phenomena poses the following requirements:

- large acceptance in pseudo-rapidity with almost full azimuthal angle coverage;
- good charged particle resolution and reconstruction efficiency in the inner detector;
- good impact parameter resolution for b -jet or τ -lepton identification (vertex detectors close to interaction region, despite the high radiation doses)
- a very good electromagnetic (EM) calorimeter for electron and photon identification and measurements, complemented by a full-coverage hadronic calorimeter for accurate jet and missing transverse energy measurements;

- precise muon identification and momentum resolution over a wide range of momenta;
- a flexible and fast trigger system to select physically interesting events, even in the presence of low transverse-momentum objects and large backgrounds from uninteresting events.

The nominal interaction point is defined as the origin of the coordinate system, the beam direction defines the z -axis and the x - y plane is transverse to the beam direction. The positive x direction is defined as pointing from the interaction point to the centre of the LHC ring. The azimuthal angle ϕ is measured around the beam axis, while the polar angle θ is the angle from the beam axis. Usually the angle θ is more conveniently expressed in terms of the pseudo-rapidity $\eta = -\log(\tan(\theta/2))$, whose differential for massless objects (as, up to a good approximation, electrons and jets from up, down or strange quarks) is invariant under a longitudinal Lorentz boost.

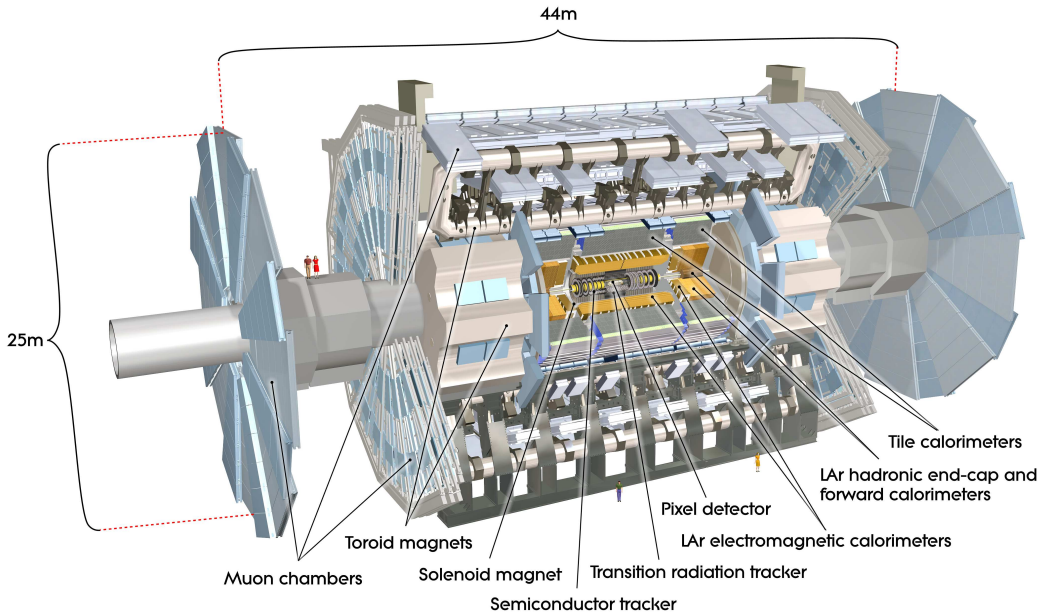


Figure 3.1: A three-dimensional view of the ATLAS Detector.

The overall ATLAS detector layout is shown in Fig. 3.1. The main sub-detectors extend radially, starting from the interaction point.

The detector subsystem closest to the interaction point is the inner detector, which is immersed in a 2 T solenoidal field. The inner detector has a coverage up to $|\eta| < 2.5$. Outside the inner detector, the electromagnetic calorimeter, based on LAr technology, covers the pseudo-rapidity range up to $|\eta| < 3.2$. Behind it, the hadronic calorimetry in the range $|\eta| < 1.7$ is provided by a scintillator-tile calorimeter, while in the end-caps ($|\eta| > 1.5$) LAr technology is used also for the hadronic calorimeters. The LAr forward calorimeters provide both EM and hadronic energy measurements and extend the pseudo-rapidity coverage up to $|\eta| = 4.9$. The muon system surrounds the calorimeter and has three layers of high precision tracking chambers, which extend up to $|\eta| < 2.7$, slightly beyond the inner detector acceptance.

A detailed description of the ATLAS detector can be found in the technical design report (TDR) [42], which is updated by Ref. [53], on which the present chapter is largely based.

3.2.1 Inner Detector

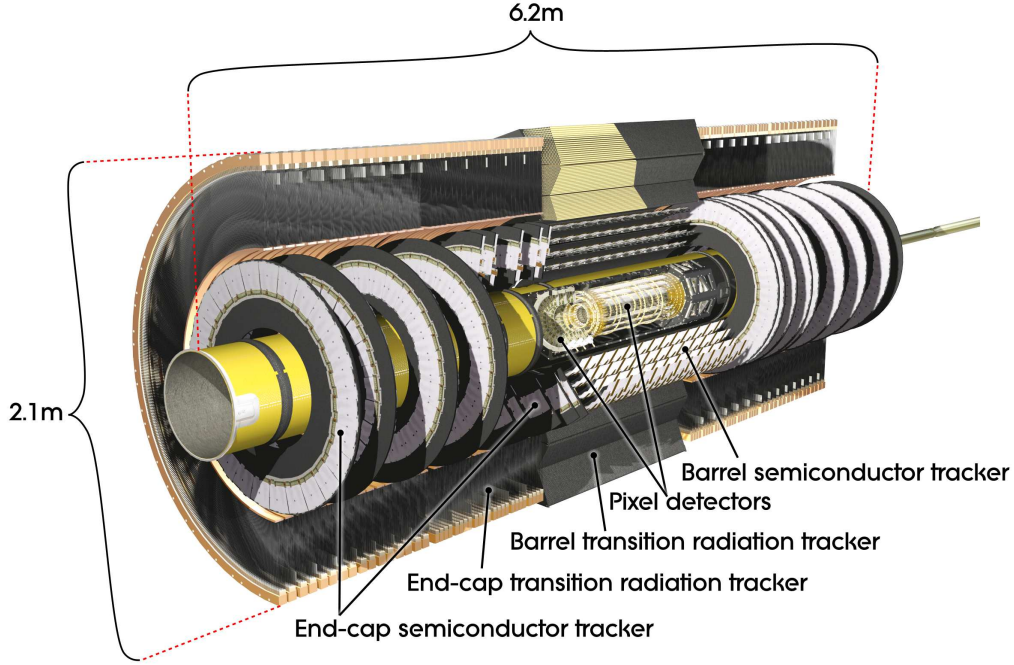


Figure 3.2: Cut-away view of the ATLAS inner detector. Taken from Ref. [53].

The reconstruction of charged tracks, in terms of momentum and impact parameters, is achieved by combining the energy hits deposited in the discrete high-resolution semiconductor pixel and silicon microstrip detectors (SCT) and, at higher radii with respect to the interaction point, the information provided by the straw-tube tracking detectors, which are also able to detect transition radiation and are therefore called Transition Radiation Trackers (TRT). The layout of the inner detector (ID) is illustrated in Fig. 3.2, while a more detailed plan view of the single sub-detector layers is shown in Fig. 3.3.

Pixel detector

The precision tracking detectors (pixels and SCT) are arranged on concentric cylinders around the beam axis, while in the end-cap regions they are located on disks perpendicular to the beam axis. The highest granularity, which is crucial to achieve a robust pattern recognition in the high occupancy and pile-up environment ATLAS will be subjected to, is achieved around the interaction region using silicon pixel detectors, which are segmented in $R - \phi$ and z and have a minimum pixel size in $R - \phi \times z$ of $50 \times 400 \mu\text{m}^2$, which results in an intrinsic accuracy in the barrel region of $10 \mu\text{m}$ ($R - \phi$) and $115 \mu\text{m}$ (z) and in the end-cap region of $10 \mu\text{m}$ ($R - \phi$) and $115 \mu\text{m}$ (R). In general, a charged track will release an energy deposition in three pixel layers. The first of these layers is positioned as close as possible to the interaction region (in the barrel at $R = 50.5 \text{ mm}$), in order to achieve the best possible impact parameter resolution: since this is extremely important for an efficient b -jet identification, the first layer is usually called the b -layer.

Silicon strip detector

For the SCT, eight strip layers (corresponding to four space points) are crossed by each track. Both in the barrel and end-cap regions, this sub-detector uses small-angle (40 mrad) stereo strips to measure both coordinates: in the barrel region one of these strips is parallel to the beam direction, while in the end-cap region one of them is running radially, in both cases measuring directly $R - \phi$. The silicon strip detectors have a mean strip pitch of $80 \mu\text{m}$. The intrinsic accuracy in the barrel region is $17 \mu\text{m}$ ($R - \phi$) and $580 \mu\text{m}$ (R), while in the end-cap region they are $17 \mu\text{m}$ ($R - \phi$) and $580 \mu\text{m}$ (z). The total number of readout channels of the pixel detector is approximately 80.4 million readout channels, while the total readout channels of the SCT is approximately 6.3 million.

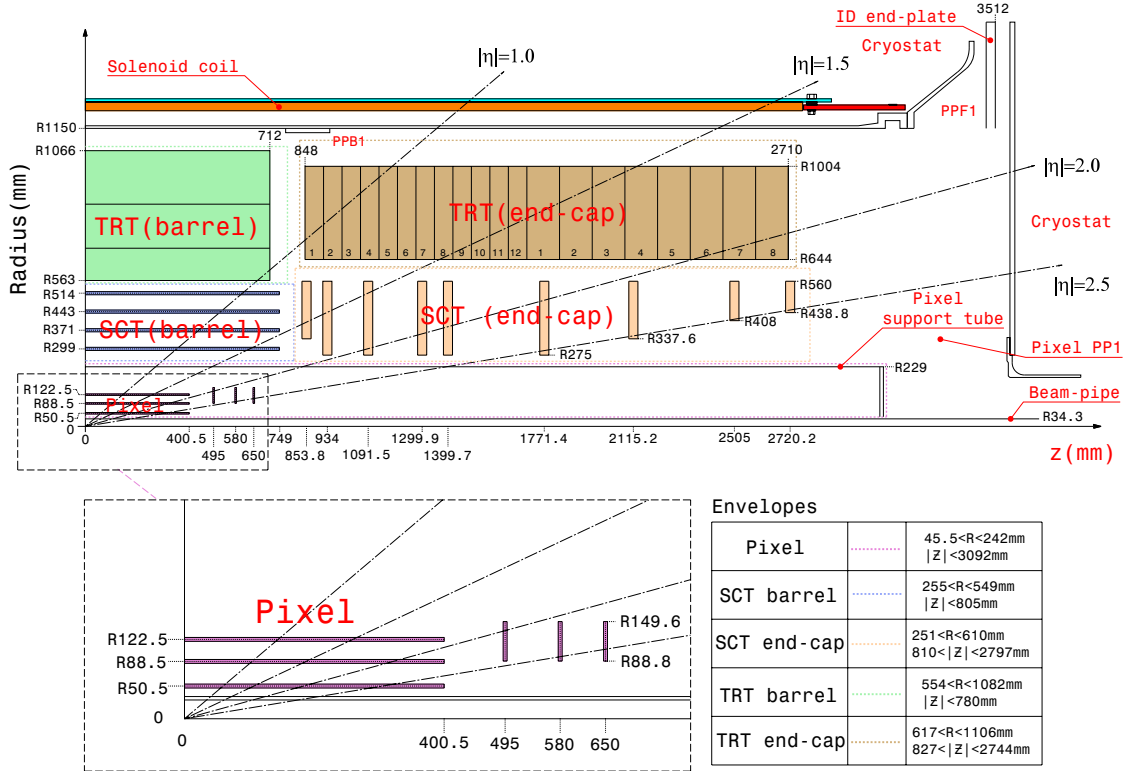


Figure 3.3: Plan view of a quarter-section of the ATLAS inner detector showing each of the major detector elements with its active dimensions and envelopes. Taken from Ref. [53].

Transition radiation tracker

In addition, a charged particle track leaves on average 36 number of hits in the 4 mm diameter straw tubes of the TRT, which enables to cover charged particle tracks up to $|\eta| = 2.0$. The TRT only provides $R - \phi$ information, for which it has an intrinsic accuracy of $130 \mu\text{m}$ per straw. In the barrel region, the straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves, approximately at $\eta = 0$. In the end-cap region, the 37 cm long straws are arranged radially in wheels. While the intrinsic resolution of the TRT cannot compete with the resolution of the silicon technology based detectors, the high number of measurements and the long lever arm with respect to the measurements in the silicon layers makes the TRT significantly contribute to the momentum resolution. In addition, the low detector granularity is compensated by the larger radial positions of the straw tubes, so that

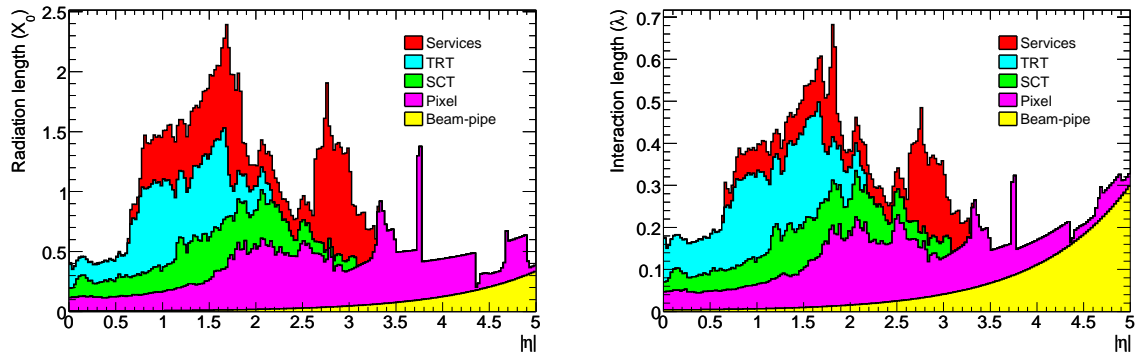


Figure 3.4: Material distribution (X_0 , λ) at the exit of the ID envelope, including the services and thermal enclosures. The distribution is shown as a function of η and averaged over ϕ . The breakdown indicates the contributions of external services and of individual sub-detectors, including services in their active volume.

a standalone pattern recognition is still possible for instantaneous luminosities up to and not beyond the nominal luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$: this is very important to correctly reconstruct tracks originating from secondary (late) interactions, as conversions, which leave no or only few energy hits in the silicon detectors. Transition radiation photons produced by electrons interacting in the xenon-based gas mixture of the straw tubes can be detected and contribute to an improved electron identification performance. The total number of TRT readout channels is approximately 351,000.

Impact of detector material

The need of a high granularity tracking detector, in particular in the vicinity of the interaction region, with related electronics, readout services and cooling, results in a relatively heavy inner detector (weighting ~ 4.5 tonnes), with a quite significant material budget (which can be expressed in terms of radiation lengths X_0 or interaction lengths λ). This material has been accurately mapped and introduced in the ATLAS simulation, as shown in Fig. 3.4, since the impact on the physics performance of the ATLAS detector is expected to be large, with the following main effects:

- A significant fraction of low-energy charged pions will undergo inelastic hadronic interaction inside the inner detector volume.
- Approximately 40% of the photons convert into an electron-positron pair before reaching the LAr cryostat and the electromagnetic calorimeter
- Many electrons lose a good part of their energy through bremsstrahlung before reaching the calorimeter.

The first two effects will be an important factor limiting the b -jet identification (b -tagging) performance. This relies in fact on distinguishing higher impact parameter tracks (from b -decays) from prompt tracks. Hadronic interactions and photon conversions constitute a source of charged tracks with high impact parameters and are thus a severe background for b -tagging. A strategy to reduce the influence of these tracks needs to be adopted and will be discussed in detail in Chapter 6.

3.2.2 Calorimeter

The requirement of hermeticity, which is a necessary condition to achieve a good resolution of the measurement of the missing transverse momentum, is one of the key design components

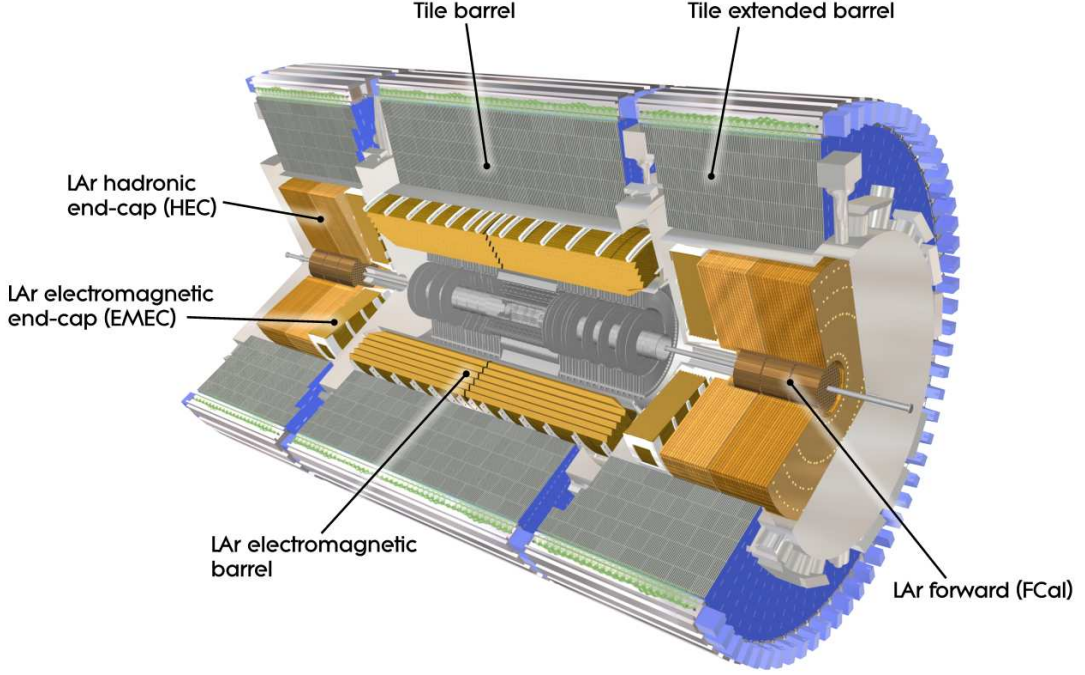


Figure 3.5: Cut-away view of the ATLAS calorimeter system.

of the ATLAS calorimeter, which in fact has a coverage up to $|\eta| = 4.9$. Different technologies are used across different regions in pseudo-rapidity, as the different calorimeter sub-detectors in Fig. 3.5 show. Over the $|\eta|$ range where the calorimeter is surrounding the inner detector, the EM calorimeter is finely segmented for precision measurements of electrons and photons, while the rest of the calorimeter is segmented more coarsely, since it is mainly aimed at reconstructing jets and at measuring the missing transverse momentum.

Another important design criterion was the need of containing electromagnetic and hadronic showers of particles of energies around the TeV scale, since energy escaping the calorimeter results both in a significantly reduced energy resolution and in punch-throughs into the muon system. The total thickness of the EM calorimeter is $> 22 X_0$ in the barrel and $> 24 X_0$ in the end-caps. The approximately 10 interaction lengths (λ) both in the barrel and in the end-caps are sufficient to provide very good resolutions for high-energy jets. The total thickness, including the outer support, is 11λ at $\eta = 0$: this has been shown by simulation and confirmed by test beam data to be sufficient to reduce punch-throughs into the muon system well below the irreducible level of prompt or in-flight decays into muons.

The pseudo-rapidity coverage, granularity and segmentation in depth of the calorimeters are summarised in Table 3.2.

LAr electromagnetic calorimeter

The EM calorimeter is divided into a barrel ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). Each end-cap calorimeter is mechanically divided into two coaxial wheels: an outer wheel covering the region $1.375 < |\eta| < 2.5$, and an inner wheel covering the region $2.5 < |\eta| < 3.2$. The EM calorimeter is based on a lead-LAr detector with accordion-shaped kapton electrodes and lead absorber plates over its full coverage. The liquid argon was chosen as

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

Table 3.2: Main parameters of the calorimeter system.

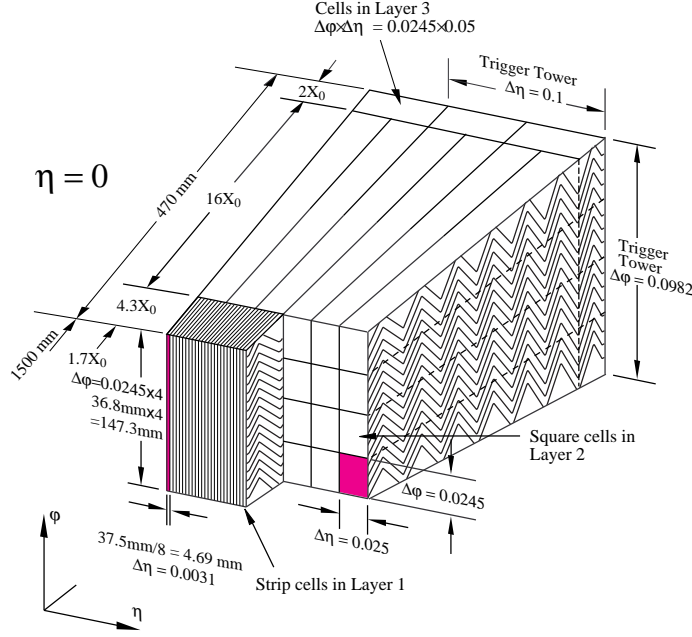


Figure 3.6: Sketch of a barrel module where the different layers are clearly visible with the ganging of electrodes in ϕ . The granularity in η and ϕ of the cells of each of the three layers and of the trigger towers is also shown. Taken from Ref. [53].

an active medium because of its intrinsic radiation hardness and good energy resolution. The advantage of the accordion geometry is that it provides complete ϕ symmetry without azimuthal cracks: a disadvantage of this geometry is the long time which the propagation and interaction of particles through the accordion shaped calorimeter takes in the detector simulation (the calorimeter simulation is therefore around an order of magnitude slower than the inner detector and muon system simulations).

Over the region which is intended to be used for precision physics $|\eta| < 2.5$, the EM calorimeter is segmented in depth in three sections. In addition, a presampler is used to recover the energy lost in dead material in front of the calorimeter. The layout of the barrel is shown in more detail in Fig. 3.6. The readout granularity of the different layers is shown in Table 3.2. The first layer of the calorimeter, called the η -strip layer, is finely granulated in η in order to allow for a better separation between photons (which results in a single energy deposition) and neutral pions, which results into two very close deposits of energy (from the $\pi^0 \rightarrow \gamma\gamma$ decay). The resolution achievable in the barrel EM calorimeter, according to test beam data, is:

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E(\text{GeV})}} \oplus 0.17\% \quad (3.6)$$

where 10% is the stochastic term and 0.17% is the constant term. The energy response is also linear within $\pm 0.1\%$. Similar results have been obtained for the end-cap EM calorimeter.

At the transition between the barrel and the end-cap calorimeters, at the boundary between the two cryostats, the amount of material in front of the calorimeter reaches a localised maximum of about $7 X_0$. For this reason, the region $1.37 < |\eta| < 1.52$ is not used for precision physics measurements involving photons and electrons.

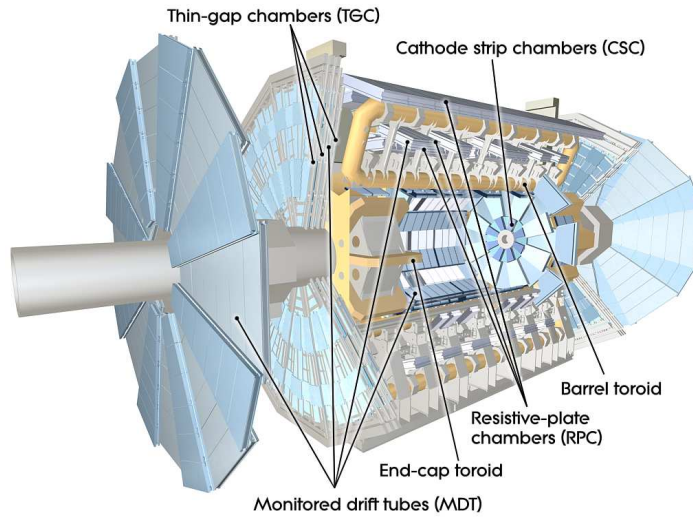


Figure 3.7: Cut-away view of the ATLAS muon system.

Hadronic calorimeters

The hadronic calorimeters are subdivided in the tile calorimeter, whose barrel covers the region $|\eta| < 1.0$ and whose extended barrels cover the region $0.8 < |\eta| < 1.7$, in the LAr hadronic end-cap calorimeters, which extends from $|\eta| = 1.5$ up to $|\eta| = 3.1$ and finally the LAr forward calorimeter, which covers the pseudo-rapidity range up to $|\eta| = 4.9$. The tile calorimeter uses steel as the absorber and scintillating tiles as active material. Two sides of the scintillating tiles are read out by wavelength shifting fibres into two separate photomultiplier tubes. The energy response to isolated charged pions of the combined LAr and tile calorimeter was tested with test beam data and turns out to be $\frac{\sigma(E)}{E} = \frac{52\%}{\sqrt{E}} \oplus 3\%$, very close to design specifications. For the end-cap hadronic calorimeter LAr technology is used, analogously to the EM calorimeter in the barrel region, but copper is used instead of lead as a passive material and a flat-late geometry was chosen. The energy response to isolated pions can be condensed in the energy resolution $\frac{\sigma(E)}{E} = \frac{71\%}{\sqrt{E}} \oplus 1.5\%$. Finally, the forward calorimeter is based again on LAr active material and uses copper as passive absorber material for the first layer and tungsten for the second and third layer. As a result of test beam data the energy response to pions is expressed by the relative energy resolution $\frac{\sigma(E)}{E} = \frac{94\%}{\sqrt{E}} \oplus 7.5\%$.

3.2.3 Muon system

The detection of muons in ATLAS can count on a completely independent sub-detector, whose layout is schematically shown in Fig. 3.7. Muons are among the most important signatures for interesting processes at the LHC and have the clear advantage that they are the only charged particles which are not stopped in the calorimeter (apart in the rare case of punch-throughs) and can be cleanly detected in the muon system. This can be ideally exploited also for the on-line selection of events.

A system of three large air-core toroids generates the magnetic field for the muon spectrometer. In the end-cap region, end-cap toroids are inserted in the barrel toroid at each end and lined up with the central solenoid. Each of the three toroids is made of eight coils assembled radially and symmetrically around the beam axis. Contrary to what happens in the inner detector, muons are therefore bent outside the inner detector in the $R - z$ plane.

The muon detector is based on two kind of sub-detectors, one for precision measurements, which is needed to determine the muon momenta with high precision, and very fast sub-detectors with coarser granularity, which are used for the online triggering of muon events and are needed to uniquely associate the muons to a certain bunch crossing. The precision measurements are performed by the Monitored Drift Tube chambers (MDT's), which cover the pseudo-rapidity region up to $|\eta| = 2.7$ (except in the innermost end-cap layer where they are limited to $|\eta| < 2$), and, in the forward region ($2 < |\eta| < 2.7$), by Cathode-Strip Chambers (CSC), which are used in the innermost tracking layer due to their higher rate and better time resolution. The MDT chambers only constrain the muon track in the bending plane (z coordinate), with a precision of $35 \mu\text{m}$, while the CSC's, being multiwire proportional chambers with cathode planes segmented into strips in orthogonal direction, provide a measurement both in the R direction of $40 \mu\text{m}$ precision and in the ϕ direction of 10 mm . These chambers are complemented (both for the measurement of the missing ϕ coordinate in the case of the MDT's and for the online selection of events) in the barrel region ($|\eta| < 1.05$) by Resistive Plate Chambers (RPC's) and in the end-cap ($1.05 < |\eta| < 2.4$) by Thin Gap Chambers (TGC). The intrinsic time resolution of these detector components (1.5 ns for RPC's and 4 ns for TGC's) is appropriate for triggering and permits to identify the correct bunch crossing within an accuracy of 99%.

Given the three chambers layout of the muon system, the momentum measurement of a high p_T track will depend on the resolution by which its sagitta (deviation in the $R - z$ plane in the middle chamber with respect to a straight line) can be determined. For a high p_T track of 1 TeV , this sagitta will be about $500 \mu\text{m}$. The muon chamber resolution provided in ϕ by the MDT chambers allows to reach a precision of better than 10% in the momentum measurements of muons up to 1 TeV , which corresponds to the design goals of ATLAS. For the measurement of low p_T muons, the measurements in the muon system typically need to be complemented by the measurements in the inner detector, since in general not all muon stations will be reached due to the stronger bending of low p_T particles in the magnetic field.

3.2.4 Trigger system

The trigger system has three different levels of event selection: Level-1 (L1), Level-2 (L2) and the event filter (EF).

The first level is implemented completely in custom hardware modules, to provide a latency of around $2 \mu\text{s}$. Since a decision must be taken every 25 ns , during the transit and processing time of around $2.5 \mu\text{s}$ the detector data must be time stamped and held in the buffers of the front-end circuits. The only signatures which can be selected in the L1 trigger are high- p_T muons, electrons/photons, jets, τ -leptons decaying into hadrons and missing transverse energy. No information coming from the tracking devices can be used at L1, so for example selecting b -jets at this level is not possible. Based on these signatures a decision is taken, in order to reduce the rate by around a factor of 500 and pass the selected events to the next trigger levels.

The L2 and EF trigger levels form the High-Level Trigger (HLT) and are both based on commercially available computer and networking hardware. The main difference between L2 and the EF is that the first is seeded by so-called Regions-of-Interest (RoI's), which are regions of the detector which the L1 trigger has identified as interesting, while the second is based on the full detector granularity. In this way the L2 trigger is much quicker and yields a first rate reduction from around 75 KHz to below 3.5 KHz , with an average event processing time of approximately 40 ms , while the EF brings this further down to 200 Hz , with a processing time per event of around 4 seconds . The higher granularity, tracking information and improved

reconstruction make a refined decision with respect to Level 1 possible. The final rate can be sustained by the actual storage technology. A detailed description can be found in Ref. [53].

4 Event reconstruction

When a physics event passes the online trigger selection, it is stored on tape and kept for later processing. The offline event reconstruction software, which in ATLAS is implemented in the software framework ATHENA, then processes the event, starting from the raw data obtained from the various sub-detectors (energy deposits and hits), processing it in many different stages and finally interpreting it as a set of charged tracks, electrons, photons, jets, muons and, in general, of possible kinds of final state objects with related four momenta. In a complex detector and physics environment as ATLAS, this task is not trivial and a lot of effort has been put into trying to improve the reconstruction algorithms in the various different detector and physics domains.

In the first part of this thesis (Chapters 5 and 6) some new reconstruction algorithms will be presented, which are aimed at improving the ability of ATLAS to reconstruct the primary interaction vertex and the identification of b -quark jets. These algorithms are mainly based on the reconstruction of charged particle tracks in the inner detector, and, for the specific case of the b -jet identification, on the reconstruction of jets in the calorimeter. A brief introduction to the reconstruction of charged particle tracks in the inner detector and to the jet reconstruction will be given in this chapter.

In addition, the analysis of the $pp \rightarrow WH \rightarrow \ell\nu b\bar{b}$ Higgs boson search channel (Chapters 7 and 8) does strongly rely on the algorithms needed to reconstruct its various different final state objects: electrons, muons, jets, missing transverse energy and b -jets. While the algorithm adopted in the analysis for the b -jet identification will be presented later in the course of the present thesis, the reconstruction of the other final states relies on the respective reconstruction algorithms as implemented in the ATLAS reconstruction software and will be therefore briefly introduced in this chapter.

The expected identification performance for the various different physics objects presented here is based on detailed Monte Carlo full simulation studies of the ATLAS detector response, based on the GEANT4simulation package. More details about these studies can be found in Ref. [44].

4.1 Reconstruction of charged particle tracks

The Inner Detector layout and the characteristics of its main sub-detectors with the related intrinsic resolutions were presented in Section 3.2.1 of Chapter 3. The tracking algorithm is based on a modular software framework, which is described in more detail in Ref. [54]. The main steps of the tracking algorithm are the following:

- a pre-processing stage, in which the raw data from the pixel and SCT detectors are converted into clusters, while the TRT raw timing information is turned into calibrated drift circles. The SCT clusters need to be further transformed into space-points, by combining the cluster information from opposite sides of a SCT module (stereo strip layers).
- a track-finding stage, in which the pattern recognition and a global χ^2 minimisation procedure is implemented as a default, but other tracking algorithms can also be chosen

for specific applications (e.g. to try to recover energy losses due to bremsstrahlung of electron tracks).

In the track-finding stage, track seeds are found in the first three pixel layers and in the first SCT layer. These are extended throughout the SCT to form track candidates and a first track fit is performed: during this step ambiguities in the cluster-to-track association are resolved, outlying measurements are removed and fake track candidates are rejected, if possible, by applying quality cuts. Afterwards, the track candidates are extended into the TRT (which covers up to $|\eta| = 2.$) and left-right ambiguities in the association of the tracks to drift circles are solved. The final track candidate is refitted with the full information from the three tracking sub-detectors.

A complementary strategy does also exist, called *back-tracking*, which searches for unused track segments in the TRT. Tracks are formed by trying to extend such segments into the SCT and pixel detectors. This recovers tracks for which the first hits in the pixel layers are missing (e.g. because they originate from secondary interactions).

A track trajectory is usually defined at its closest point to the interaction region on the transverse plane by its impact parameters in the transverse plane and in the longitudinal direction, respectively called d_0 and z_0 , and by its momentum, typically expressed in azimuthal angle ϕ , polar angle θ and inverse momentum $1/p$. Usually, the inverse momentum is multiplied by the track charge, to yield a more compact representation (q/p). More details about the track parametrisation will be given in Section 5.3.3 of Chapter 5.

In the performance plots shown in the following, after the tracks are reconstructed, some standard quality criteria are applied:

- transverse momentum $p_T > 1$ GeV;
- impact parameters: $|d_0| < 2$ mm and $|(z_0 - z_{PV}) \cdot \sin(\theta)| < 10$ mm (where the z_{PV} is the z coordinate of the primary vertex position);
- at least seven precision (pixel or SCT) hits.

A second tighter set of quality cuts is also defined, which is mainly intended for application to the impact parameter based *b*-jet identification algorithm:

- the standard quality criteria already mentioned;
- tighter impact parameter cuts: $|d_0| < 1$ mm and $|(z_0 - z_{PV}) \cdot \sin(\theta)| < 1.5$ mm;
- at least two hits in the pixel layers, one of them in the *b*-layer.

The efficiency of reconstructing isolated muons and charged pions is illustrated in Fig. 4.1. While the muons can be reconstructed with very high efficiency, the pion efficiency is mainly limited by hadronic interactions (e.g. spallation processes) in the detector material, which is present in higher amounts at higher pseudo-rapidities.

The resolution of any track parameter X can be approximately expressed as a function of p_T as:

$$\sigma_X(p_T) = \sigma_X(\infty) \left(1 \oplus \frac{p_X}{p_T} \right), \quad (4.1)$$

where $\sigma_X(\infty)$ is the asymptotic resolution expected at infinite momentum and p_X is a constant representing the value of p_T at which the multiple scattering contribution equals that from the detector resolution. This expression is approximate, but is expected to work well at low

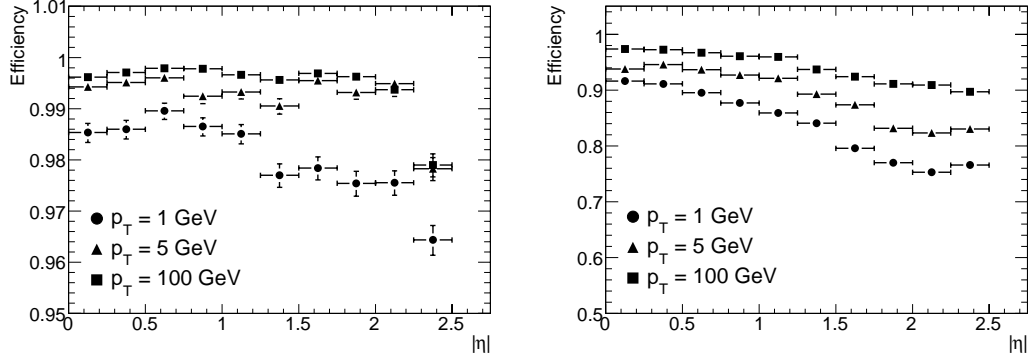


Figure 4.1: Track reconstruction efficiencies as a function of $|\eta|$ for muons (left) and pions (right) with $p_T = 1, 5$ and 100 GeV. Taken from Ref. [44].

momenta, where the multiple scattering dominates the resolution, and at high momenta, where the resolution is limited by the bending power of the solenoid field and by the intrinsic detector resolution. The resolution is taken as the RMS evaluated over a range which includes 99.7% of the data (corresponding to $\pm 3 \sigma$ of a Gaussian distribution). The momentum resolution for muons and the impact parameter resolutions for pions as a function of p_T and η of the track are shown in Fig. 4.2. The results in terms of the parameters $\sigma_X(\infty)$ and p_X for the various different track parameters are listed in Table 4.1.

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

Table 4.1: Expected track parameter resolutions (RMS) at infinite transverse momentum, $\sigma_X(\infty)$, and transverse momentum, p_X , at which the multiple-scattering contribution equals that from the detector resolution. The values are shown for two η -regions, one in the barrel inner detector where the amount of material is close to its minimum and one in the end-cap where the amount of material is close to its maximum. Taken from Ref. [44].

The pattern recognition and track fitting task is more difficult in a dense jet environment, in particular because of the occurrence of shared hits between different tracks. This decreases the track reconstruction efficiency and increases the fake rate for tracks reconstructed inside jets, which are both shown in Fig. 4.3 as a function of the track pseudo-rapidity and in Fig. 4.4 as a function of the distance $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$ of the track with respect to the jet axis, in proximity of which the track density is expected to be the highest.

4.2 Electron identification

The electron identification algorithm is crucial to identify electrons and reject the very large backgrounds typically originating from jets faking electrons with rejection factors up to 10^5 . The standard algorithm starts with an electron seed in the electromagnetic (EM) calorimeter, provided by a cluster with transverse energy above 3 GeV and a matching track among all

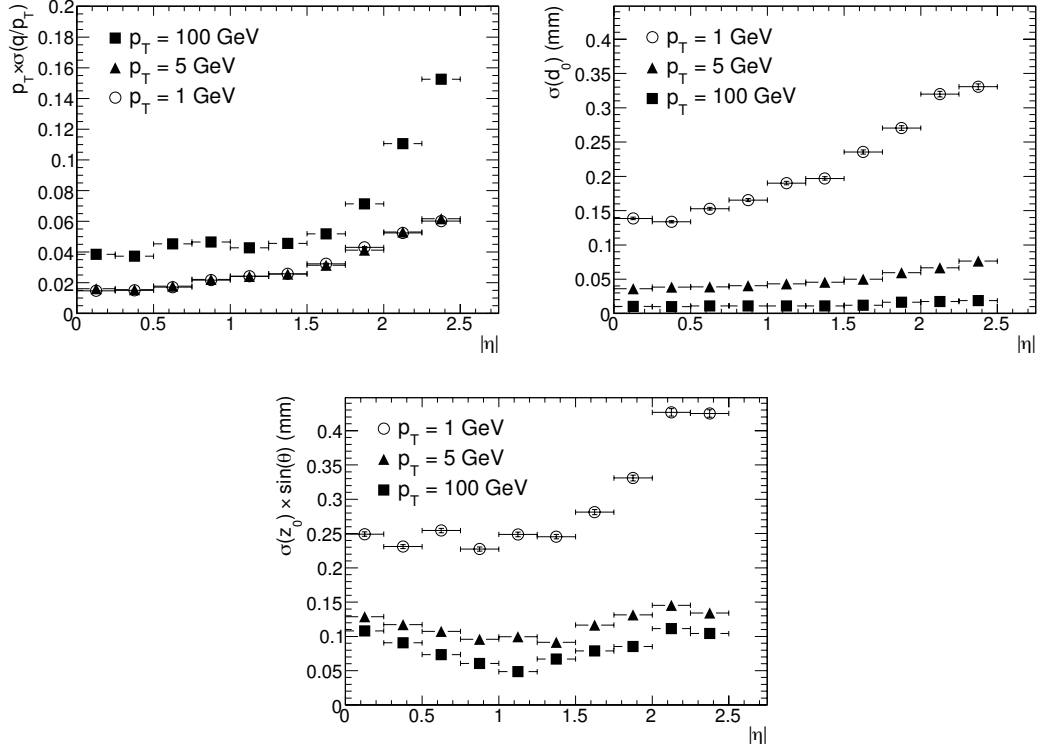


Figure 4.2: Relative transverse momentum resolution for muons (top left), transverse (top right) and longitudinal (bottom) impact parameter resolution for pions as a function of $|\eta|$ for muons/pions with $p_T = 1$ GeV (open circles), 5 GeV (full triangles) and 100 GeV (full squares). Taken from Ref. [44].

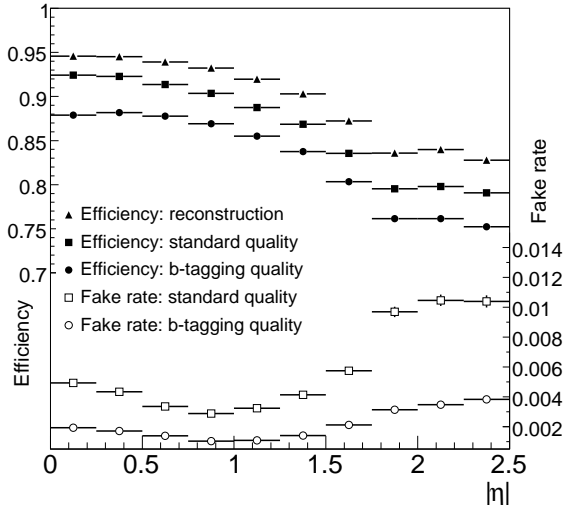


Figure 4.3: Track reconstruction efficiencies and fake rates as a function of $|\eta|$, for charged pions in jets in $t\bar{t}$ events and for different quality cuts (as described in the text). Taken from Ref. [44].

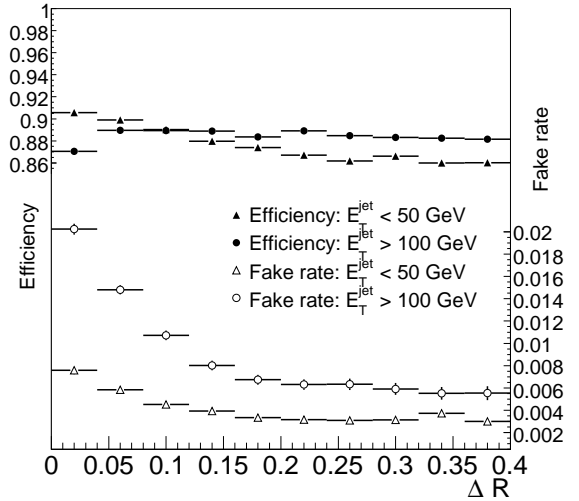


Figure 4.4: Track reconstruction efficiencies and fake rates as a function of the distance ΔR (defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) of the track to the jet axis, using the standard quality cuts and integrated over $|\eta| < 2.5$. Taken from Ref. [44].

those reconstructed in the inner detector and not used in the reconstruction of any $\gamma \rightarrow e^+e^-$ conversion vertices, satisfying the following criteria:

- the EM cluster and the track extrapolated to the calorimeter must be within a window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.10$;
- the ratio E/p of the energy of the cluster to the momentum of the track is required to be lower than 10.

The measurement of the momentum of the electron track is mainly limited by the severe energy losses by bremsstrahlung in the detector material; special tracking algorithms (Gaussian Sum Filter [55] or Dynamic Noise Adjustment [56]) are used to partially recover these energy losses.

The preselection cuts keep 93% of true isolated electrons with $E_T > 20$ GeV and $|\eta| < 2.5$. The standard identification algorithm for high- p_T electrons is then based on a combination of cuts, with three reference sets of cuts, *loose*, *medium* and *tight*, providing different electron efficiency versus jet rejection working points.

Loose electron identification

The loose identification criteria are based on a limited number of discriminating variables from the calorimeters and deliver high electron identification efficiency, but low background rejection. Cuts are applied on the hadronic leakage (ratio of E_T of the first sampling of the hadronic calorimeter to EM cluster E_T) and on shower-shape variables based on the EM middle layer (lateral shower shape and lateral shower width).

Medium electron identification

With respect to the loose identification, in the medium identification selection the fine longitudinal segmentation of the first layer of the EM calorimeter is exploited, in particular to reject photons from neutral pion decays in jets, which result into two very close deposits of energy in the calorimeter, and some quality cuts on the electron track are applied.

In more detail, to reject $\pi^0 \rightarrow \gamma\gamma$ decays, a second maximum in energy is looked for in the cells of the first EM layer in a window of $\Delta\eta \times \Delta\phi = 0.125 \times 0.2$ around the cell with the highest E_T . Then cuts are applied on the following variables:

- the difference between the first and second maximum;
- the difference in energy between the second maximum and the η strip with the lowest energy between the two local maxima;
- the second maximum normalised to the cluster energy;
- the total shower width in the first layer;
- the shower width over three strips around the maximum;
- the fraction of energy outside these three core strips.

In addition to that, further cuts are applied to the track matched to the EM calorimeter cluster, in particular on the number of hits in the pixel sub-detector, on the number of silicon hits (pixel + SCT sub-detectors) and on the transverse impact parameter. The medium selection criteria increase the jet rejection by a factor 3 – 4 with respect to the loose identification, reducing the electron selection efficiency by around ≈ 10 %.

Tight electron identification

In addition to the medium identification selection criteria, additional cuts are applied on the number of silicon layer hits (to reject electrons from conversions), on the number of hits in the TRT, on the fraction of high threshold hits in the TRT (which correspond to transition radiation in the TRT and therefore higher probability of having electrons than charged pions), on the difference between η and ϕ of the EM cluster and the extrapolated track position and on the ratio of cluster energy to track momentum. Two slightly different sets of tight selection criteria are available: in the first (denominated *tight isol*), an isolation criterion is also considered, additionally relying on the energy deposited in the calorimeter in a cone of $\Delta R = 0.2$ around the electron EM cluster, while in the second this cut is substituted by tighter cuts on the TRT high threshold hit information.

Performance

The expected performance of the three mentioned identification criteria is shown in Table 4.2. The first two criteria will be used for the Higgs search analysis presented in this thesis: the electron identification efficiency will be considered in more detail for the specific kinematic range of electrons selected in the analysis, but no further investigation will be carried out for the jet rejection, which is, however, a fundamental ingredient to reduce the background from QCD multi-jet events to a negligible level. In the analysis also further isolation requirements are applied to reject background from QCD multi-jet events with leptons from heavy flavour decays.

Cuts	$E_T > 17 \text{ GeV}$		
	Efficiency (%)		Jet rejection
	$Z \rightarrow ee$	$b, c \rightarrow e$	
Loose	87.9 ± 0.1	50.8 ± 1	567 ± 1
Medium	77.3 ± 0.1	30.7 ± 1	2184 ± 13
Tight (TRT)	61.7 ± 0.1	22.5 ± 0.4	$(8.9 \pm 0.3)10^4$
Tight (isol.)	64.2 ± 0.1	17.3 ± 0.4	$(9.8 \pm 0.4)10^4$

Table 4.2: Expected efficiencies for isolated and non-isolated electrons and corresponding jet background rejections for the three standard levels of cuts used for electron identification. The results are shown for simulated inclusive jet samples corresponding to E_T -thresholds of the electron candidates of 17 GeV. Taken from Ref. [44].

4.3 Jet reconstruction

The jet algorithms try to map final state hadrons onto their underlying final state partons. The calorimeter is ideally suited for the reconstruction of jets, since it provides an effective way to reconstruct the four momenta of both charged and neutral particles.

4.3.1 Input to jet finding

Due to the finite granularity of the calorimeter, it is, however, not possible to always clearly disentangle the contributions arising from the individual particles. Usually the input to the jet clustering algorithms in ATLAS is provided either by a set of four-momenta representing the energy deposited in calorimeter towers of 0.1×0.1 in $\Delta\phi \times \Delta\eta$ (summing up the contribution of all EM and hadronic calorimeter layers), or, in alternative, by trying to build up more complex three-dimensional clusters of energy, called *topoclusters*. The last approach has been

adopted as the new standard for jet reconstruction and will be therefore briefly described here. The clustering starts with seed cells with a signal-to-noise ratio $E_{\text{cell}}/\sigma_{\text{noise,cell}}$ above 4. All directly neighbouring cells of these seed cells, in all three dimensions, are collected into the cluster. Neighbours of neighbours are considered if the added cells have a signal to noise ratio above 2. Finally, a ring of guard cells with signal-to-noise ratio above a basic threshold of 0 is added to the cluster. After the initial clustering, the found three-dimensional clusters are analysed for the presence of multiple local maxima and, in this case, a splitting between these maxima is performed. On di-jet QCD events Monte Carlo simulations show that the number of particles can be resolved in a sufficiently accurate way in the pseudo-rapidity region $1.5 < |\eta| < 2.5$, while in the complementary regions the number of real particles per reconstructed topocluster on average is ≈ 1.6 .

4.3.2 Jet finding algorithms

The topoclusters are then used as input to the jet finding algorithm. In ATLAS two main types of algorithms are implemented, fixed cone algorithms and sequential recombination algorithms. A good jet algorithm should be defined at any order in perturbation theory, yield the same result if applied at detector level, at hadron level and directly at parton level and be sufficiently insensitive to the details of the hadronisation. The theoretical motivations and limitations of the various jet finding algorithms are for example reviewed in Ref. [57].

Fixed cone jet finder in ATLAS

In ATLAS the standard jet finding algorithm for a long time was an iterative fixed-cone jet finder. First, all input four momenta are ordered in decreasing order in transverse momentum p_T . If the object with the highest p_T is above the seed threshold (set to 1 GeV), all objects within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < R_{\text{cone}}$, where R_{cone} is the fixed cone radius, are combined with the seed. Then an iterative procedure is started: the new direction is used to draw a new cone around it and the four momentum is again updated. This is repeated until the direction stabilises and the resulting cone is considered as a jet candidate. The same procedure is applied to all seeds. The jets found this way can share part of their constituents. Therefore, jets which share constituents corresponding to more than a fraction $f_{sm} = 0.5$ of the p_T of the less energetic jet are merged, while they are split in the opposite case. Two cone sizes are usually used ($R_{\text{cone}} = 0.4, 0.7$). From a theoretical point of view, the splitting and merging procedure makes this algorithm partially infrared safe, but the algorithm remains well defined only up to leading order of perturbation theory.

Sequential recombination algorithms

The most common sequential recombination algorithm is the inclusive k_T algorithm. All pairs of input objects are analysed with respect to their distance measure defined as:

$$d_{ij} = \min(p_{T,i}^2, p_{T,j}^2) \frac{\Delta R_{ij}^2}{R^2}, \quad \Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2}, \quad (4.2)$$

where R is a cone-radius like parameter which determines the size of the jets. In addition, also the distances of the objects i with respect to the beam are considered, since divergences in QCD branching also show up between initial state and final state partons and they are defined as $d_{iB} = p_{T,i}^2$. The minimum d_{\min} of all d_{ij} and d_{iB} is determined. If d_{\min} is among one of the d_{ij} , then the corresponding objects i and j are merged into a single new object using four-momenta recombination, while if d_{\min} is one of the d_{iB} , the object i is considered to be

a jet by itself and removed from the list. This procedure is repeated iteratively, updating all distances at each iteration step. The four-momenta recombination scheme is usually chosen to be the simple sum of the two four momenta of the objects i and j . The advantage of the k_T algorithm is that it is by construction collinear and infrared safe, however, it is also more sensitive to picking up soft contributions from pile-up and the underlying event.

The k_T algorithm can be generalised by introducing the following particle-particle and particle-beam distance measures:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad \Delta R_{ij} = \sqrt{\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2} \quad (4.3)$$

$$d_i = p_{T,i}^{2p} \quad (4.4)$$

where p is a parameter which is 1 for the k_T algorithm. Two different algorithms can be obtained: the C/A algorithm by choosing a value of $p = 0$ and the anti- k_T algorithm by choosing a value of $p = -1$. The C/A algorithm recombines objects close in ΔR iteratively and reflects the angular ordering of QCD radiation: it is hence ideally suited to reconstruct and decompose the various decay components of heavy objects like Higgs bosons [1] or top quarks [58] using subjet structure and will therefore be used in the course of the present thesis to study boosted Higgs boson decays into a pair of b -quarks. Contrary to the k_T algorithm, the anti- k_T algorithm does first cluster hard objects together, which results in more regular jets with respect to the k_T and C/A algorithms: it is considered as a good replacement for the ATLAS (collinear unsafe) standard fixed-cone algorithm and was therefore recently made the new standard algorithm for jet finding in ATLAS.

4.3.3 Jet calibration

Once the jets are found, they need to be calibrated to compensate for the e/π energy response ratio of the non-compensating calorimeters, for energy losses in cracks, due to particles escaping the calorimeter and due to out-of-cone contributions. The standard calibration routine adopted in ATLAS, which will also be used for the clustering of the boosted $H \rightarrow b\bar{b}$ candidates needed in the analysis presented in the present thesis, is based on cell signal weighting, applied to the calorimeter cells forming the three-dimensional clusters of energy (topoclusters) contained in the found jets. These weights are often referred to as *H1-style weights*, since this approach was adopted originally in the H1 experiment [59]. The basic idea is that low signal densities in the calorimeter cells indicate a hadronic signal and thus the need of a signal weight for compensation of the order of the electron/pion signal ratio e/π , while high signal densities are more likely generated by electromagnetic showers, which do not require additional re-weighting. After the cell information corresponding to a certain jet has been extracted, the signal in each cell i is weighted by a function depending on the cell location \vec{X}_i and on the cell signal density $\rho_i = E_i/V_i$, with E_i being the electromagnetic energy signal of the cell and V_i being its volume. The weighting factor is ≈ 1 for high density cells and rises up to 1.5, the typical e/π ratio for the ATLAS calorimeters, with decreasing cell signal densities. The calibrated jet four-momentum $(E_{jet,calo}, \vec{p}_{calo})$ is then obtained from the weighted cell signals, which are treated as massless four-momenta (E_i, \vec{p}_i) :

$$(E_{jet,calo}, \vec{p}_{calo}) = \left(\sum_i^{N_{cells}} w(\rho_i, \vec{X}_i) E_i, \sum_i^{N_{cells}} w(\rho_i, \vec{X}_i) \vec{p}_i \right) \quad (4.5)$$

The weighting functions are universal and have been determined by using seeded fixed size cone jets ($R_{cone} = 0.7$) in fully simulated QCD di-jet events by fits of the reconstructed jet

energies to the corresponding Monte Carlo truth particle jet energies. In addition, residual non-linearities (as a function of p_T) and non-uniformities (as a function of η) have to be corrected: this final jet energy scale correction depends on the jet finding algorithm considered (mainly because of out-of-cone corrections) and needs to be obtained for every specific jet finding algorithm. Alternative calibration procedures do also exist but are not considered in the course of the present thesis.

4.4 Missing transverse energy

The measurement of missing transverse energy (E_T^{miss}) in ATLAS is mainly based on the energy deposited in the calorimeter and on the reconstructed muon tracks (which escape the calorimeter). Missing energy at a hadron collider can be measured only in the transverse plane, since an unknown fraction of the longitudinal momenta of the incoming protons is taken away from the proton remnants leaving the detector outside its acceptance region after the hard scattering has occurred. The energy in the calorimeter is classified and calibrated according to the reconstructed objects (jets, electrons, photons,...) to which it belongs. The thickness of the cryostat (dead material) between the LAr barrel EM calorimeter and the tile barrel hadronic calorimeter requires a specific correction to be implemented. Simulation studies on typical physics events show that the missing transverse energy can be reconstructed with a response which is linear within 5%, even for very low missing E_T values and that the energy resolution is well described by the simple formula $\sigma = a \cdot \sqrt{\sum E_T}$, with $a \approx 0.55$ and where the sum runs over all interacting particles produced in the event. At high energies the angular resolution is better than 100 mrad, while a significant degradation is seen at missing transverse energies below 40 GeV.

4.5 Muon reconstruction

Two main strategies are employed to identify and reconstruct muons. Only the baseline algorithms adopted as standard in the reconstruction will be described here. The direct approach is to reconstruct *standalone* muons by finding tracks in the muon spectrometer and then, as a second step, extrapolate them to the beam line. The other approach, which yields *combined* muons, is to match standalone muons to inner detector tracks and then combine the two measurements. In addition, some efficiency is recovered (in particular for low p_T muons up to few GeV, which do not reach the outer muon stations) by extrapolating all tracks reconstructed in the inner detector and looking for unused hits in the muon system along their extrapolated path: these are called *tagged* muons.

The reconstruction of standalone muons starts from building track segments in each of the three muon stations (described in Section 3.2.3 of Chapter 3). Then these segments are linked together and the track is extrapolated to the beam line. This extrapolation must consider the enlargement of the track parameter errors due to taking into account multiple scattering in the detector material. Standalone muons can be reconstructed up to $|\eta| = 2.7$, extending the inner detector coverage of up to $|\eta| = 2.5$, but have some acceptance holes in proximity of $\eta = 0$ and $|\eta| = 1.2$.

Combined muons are formed by trying to combine standalone muons to inner detector tracks: the combination relies on a χ_{match}^2 which condenses the probability that the track parameters of the inner detector and muon system standalone tracks, defined at the point of closest approach to the beam line, are compatible with each other within their experimental errors. A simple statistical combination (by re-weighting the track parameters by their re-

spective weight matrix) of the two tracks is then performed¹. Combined muons significantly reduce the rate of fake muons, originating mainly from muons produced by π and K decays in the calorimeter.

¹The ideal combination would rather rely on matching the two tracks at a common point in between the inner detector and muon system and then perform a real complete refit of the tracks in both inner detector and muon system.

5 Reconstruction of the primary interaction vertex

5.1 Introduction

When the LHC will be in operation, bunches of protons orbiting in opposite directions in the LHC accelerator will be forced to collide in proximity of the centre of the ATLAS detector. The precise size of the interaction region (usually called *beam spot*) or more precisely the spatial distribution of the interacting proton proton pairs will depend on the mode of operation of the LHC machine, as described at the beginning of Chapter 3. In the simulation this distribution is assumed to be approximately Gaussian both in the transverse plane and in the longitudinal direction, with parameters assumed to be $\sigma_{x,y} \approx 15 \mu m$ and $\sigma_z \approx 55 mm$, respectively, regardless of the instantaneous luminosity.

The position of the hard scattering event of interest, which is typically selected by the online trigger system according to a distinctive high p_T signature left in one of the ATLAS sub-detectors, is therefore tightly constrained in the transverse plane, while it is essentially left unconstrained in the longitudinal direction. This can be considered as an *a priori* knowledge of the primary interaction vertex position, before any reconstruction algorithm is applied: in order to be able to make use of such information, the stability of the interaction region must be carefully monitored during a single period (*run*) of data taking.

Most of the final state charged particles produced in the high p_T physics event of interest are reconstructed in the inner detector as primary tracks. These tracks have a common intersection in a single origin, which is the primary interaction vertex where all primary particles, including neutral ones, originated. The determination of the primary interaction vertex depends crucially on the impact parameter resolution of the primary tracks, for which the first layers of the pixel detector are absolutely crucial, as described in Section 3.2.1. The impact parameter resolution as a function of track transverse momentum and pseudo-rapidity were illustrated in Section 4.1.

Taking a $pp \rightarrow t\bar{t}$ event as a reference for a typical high p_T event of interest, with an average number of 43 reconstructed primary tracks at an average p_T of ≈ 3 GeV (as shown in Fig. 5.6 and Fig. 5.7), the primary event vertex can be reconstructed with a resolution of around $10 \mu m$ in the transverse plane and $\approx 30 \mu m$ in the longitudinal direction, as will be demonstrated in the course of this chapter. The accuracy the primary vertex can be determined with, depends mainly on the amount of outgoing charged particles and on their transverse momenta. The determination of the primary event vertex is more challenging in $Higgs \rightarrow \gamma\gamma$ events, where the only primary charged particle tracks in the event are produced as a consequence either of initial state radiation or of the underlying event: in this case the resolution of the reconstructed primary vertex drops to $\approx 20 \mu m$ in the transverse plane and to $\approx 50 \mu m$ in the longitudinal direction, without including any information from the knowledge of the beam spot position and size.

The transverse resolution of the primary vertex is comparable to the transverse size of the beam spot: this means that, if the beam spot position and size is accurately determined averaging the primary vertex positions over a large number of events, the improvement which

the explicit reconstruction of the primary event vertex on an event basis can provide on its transverse position is small, while the reconstruction of the primary vertex position along the longitudinal direction is essential, since it permits to improve its longitudinal resolution from the nominal 55 mm of the longitudinal beam spot¹ width to $\approx 30\text{-}50\ \mu\text{m}$.

It is worth highlighting that, although the transverse resolution of the charged particle tracks in the inner detector has a marginal impact on the determination of the primary vertex position, it is however extremely important to distinguish whether or not a specific track was produced directly in the primary interaction vertex (primary track) or in a secondary displaced decay vertex of a short-lived particle ($c\tau \simeq 100\ \mu\text{m}\text{-}1\ \text{mm}$) produced in the primary interaction (secondary track): this will be the most essential ingredient for the b -tagging algorithms which will be discussed in Chapter 6.

The distinction of primary and secondary tracks is of utmost importance not only for b -tagging, but also for τ -tagging, photon and electron identification, reconstruction of conversions and many other applications. Several vertex reconstruction tools originally developed for primary vertex finding and b -tagging in the course of this thesis were also adopted in ATLAS for example for τ -tagging and photon/lepton identification, or even directly in Higgs boson searches as in the $H \rightarrow ZZ \rightarrow 4\ell$ search channel: some examples for such cases will be briefly mentioned in the results section (Section 5.5).

5.2 Primary vertices in pile-up environment

The longitudinal size of the interaction region of several tenths of centimetres is of crucial importance to create the conditions to distinguish the high p_T event of interest from additional physics events produced by the interaction of additional proton-proton pairs during a single bunch crossing, which are usually defined as *pile-up* (or minimum bias) events.

These events will be produced randomly spread across the interaction region. Given that the longitudinal size of the beam-spot is much larger than the longitudinal primary vertex resolution of typical minimum bias events ($\approx 55\ \text{mm} \gg \approx 80\ \mu\text{m}$), several primary vertices can be separated along the z axis and reconstructed on an event by event basis; one of them will correspond to the high p_T event of interest.

As already mentioned in Section 3.1, assuming the cross section for minimum bias events to be 70 mb and an instantaneous luminosity of $2 \times 10^{33}\ \text{cm}^{-2}\text{s}^{-1}$ with a bunch crossing spacing of 25 ns, as planned for the first *low luminosity* running scenario, around 4.6 additional interactions are expected to be overlaid to the high p_T event of interest on average.

While *in-time* pile-up events, as just mentioned, have an important effect on tracking and thus on vertexing, *out-of-time* pile-up events, as defined in Section 1.4.8, have a smaller effect, thanks to the fast response in time of both the pixel and SCT sub-detectors, which is such that hits corresponding to energy deposits outside a time window compatible with the bunch crossing of interest are rarely associated to tracks of the event of interest.

The spatial separation between the primary vertex corresponding to the high p_T event of interest and additional vertices from pile-up events is a necessary but not sufficient condition to really isolate the signal primary vertex from the remaining ones. If one would be only interested in a specific analysis and apply the analysis specific selection on all primary vertices reconstructed in the event², the probability for a pile-up event to fulfil the same requirements would be extremely low and one would be essentially able to always select the correct vertex. However, the ATLAS reconstruction software needs to remain very generic, so that the

¹The beam spot is defined in this thesis as the three-dimensional space-distribution of the interaction region.

²Here it is assumed that the final state objects of the signal process in the physics analysis of interest can be uniquely related to a specific primary vertex: some exception exist, as in the case of the $H \rightarrow \gamma\gamma$ process.

identification of the primary vertex where the high p_T event of interest originated needs to be based on generic event properties as for example the transverse momenta of the charged particle tracks associated to the vertices or their track multiplicity. Based only on these criteria and in particular for certain specific physics processes (as in the case of $H \rightarrow \gamma\gamma$), it is not straightforward, after all primary vertices corresponding to a single recorded bunch crossing are reconstructed, to identify the vertex containing the physics event of interest.

The primary vertex reconstruction algorithm, in the presence of additional interactions from pile-up, needs to be designed and optimised in such a way to correctly and efficiently separate spatially the single interaction vertices originating from the high p_T event of interest and from pile-up events. In a second step, based on the properties of the physics events corresponding to the reconstructed interaction vertices, selection criteria need to be defined in order to select the high p_T event of interest. These two tasks constitute an important part of the *vertex finding* algorithms.

The algorithms which are responsible for the reconstruction of the primary interaction vertex can be usually distinguished in two different stages:

- Vertex finding
- Vertex fitting

The vertex finding algorithms usually deal with the task of correctly clustering tracks into multiple vertices, doing the tracks to vertices association properly. The main problem of the finding algorithms, as already mentioned in this introduction, is, on one side, the optimal distinction of primary from secondary tracks and, on the other side, the optimal separation of tracks originating from different but closely positioned vertices from the signal or additional pile-up events. The vertex fitting algorithms, instead, deal with the problem of getting the best estimate of the vertex position, once a certain set of tracks is defined.

Some new *adaptive* methods for vertex fitting, however, deal intrinsically also with a part of the vertex finding problem, creating a very close interplay between the finding and fitting tasks. In order to systematically describe the relevant algorithms for the present thesis work, first the vertex fitting algorithms will be presented (Section 5.3), and then a description will follow about how their use can be adapted to solve the primary vertex finding problem (Section 5.4).

5.3 Mathematical determination of the vertex position

The determination of the vertex position, usually defined as *vertex fit*, is a well defined task: it consists in taking N input tracks and in determining their intersection, which results in the estimation of the vertex position and its related error. The vertex fit itself does not usually care whether the hypothesis that the N input tracks intersect a single point of the three-dimensional space is correct or not.

The input, given by the charged particle tracks as reconstructed by the ATLAS tracking algorithm in the inner detector, is provided in the so called *perigee* representation, which defines the track helix (expected due to the bending of a charged particle in the ATLAS magnetic field) at the position of closest approach in the transverse plane of the helix to a reference point, which is usually chosen to be the origin of the ATLAS detector $(0, 0, 0)$, as shown in Fig. 5.1. In this representation the trajectory is fully specified by five parameters \vec{q} : the transverse and longitudinal impact parameters with respect to the reference point (d_0 and z_0) and the track momentum (in terms of azimuthal angle ϕ , polar angle θ and charge

times the inverse momentum q/p). The helix parametrisation will be described in more detail in Section 5.3.3.

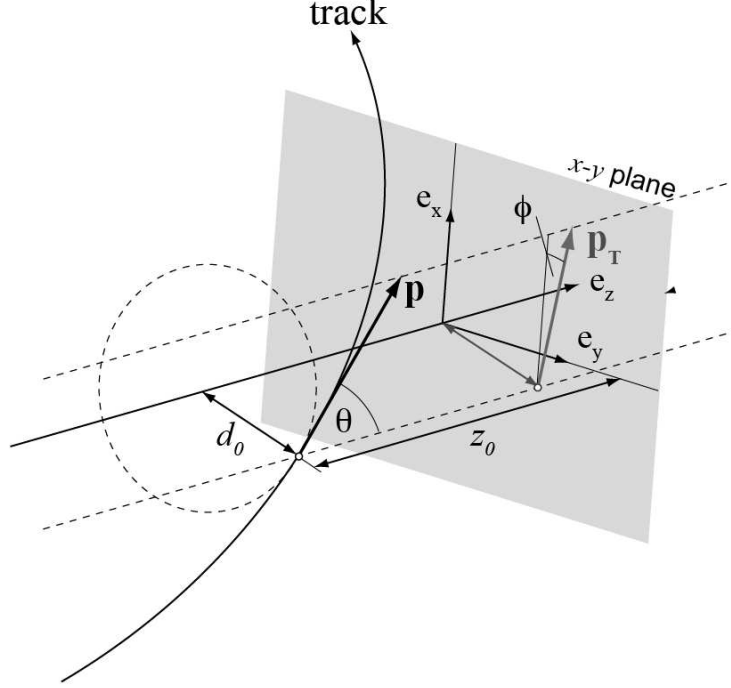


Figure 5.1: Illustration of helix parameters of a charged track defined in the *perigee* representation in the ATLAS convention, from Ref. [60].

Given a set of parameters \vec{q} , each track can be represented as a probability density distribution in the three-dimensional space, of the form:

$$P(\vec{r}) = \int d\phi_p \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}(\phi_p))^T \mathbf{COV}_{3 \times 3}^{-1}(\phi_p) (\vec{r} - \vec{r}(\phi_p)) \right], \quad (5.1)$$

where the additional parameter ϕ_p is defined in such a way to represent a specific position on the trajectory itself in an arbitrary coordinate system and must be integrated over. The covariance matrix $\mathbf{COV}(\phi_p)$ represents the projection of the covariance matrix in the space of the perigee parameters \vec{q} on the three-dimensional space, given a specific position ϕ_p along the trajectory³. Geometrically this can be visualised as a Gaussian probability tube along the track.

In general the distribution of the probability for the measurements provided by N tracks to be compatible with the hypothesis that all tracks intersect in a common space position \vec{r} will be:

$$P(\vec{r}) = \int d\phi_{p,1} d\phi_{p,2} \dots d\phi_{p,N} \prod_{i=1}^{N_{trk}} \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}_i(\phi_{p,i}))^T \mathbf{COV}_{3 \times 3,i}^{-1}(\phi_{p,i}) (\vec{r} - \vec{r}_i(\phi_{p,i})) \right], \quad (5.2)$$

³The covariance matrix $\mathbf{COV}_{3 \times 3}$ is ill-defined, because the error is infinite along the track itself, but its inverse, the weight matrix, $\mathbf{COV}_{3 \times 3}^{-1}$, is well defined.

where this expression still contains an integral over all tracks trajectories, which makes its evaluation very expensive. However, this expression can be significantly simplified by two simple considerations:

- The N track PDF is a product of the single track Gaussian probability tubes along the tracks. With two or more non-degenerate tracks, the intervals of parameters ϕ_p of interest along the involved tracks (for which the product of probabilities is non zero) gets reduced to a very small range, where the position along the tracks is sufficiently close to the intersection of all tracks.
- The constraint provided by each track in the form of its covariance matrix on the vertex position \vec{r} is always necessarily perpendicular to the track itself, which can be expressed mathematically by the condition $\nabla_{\phi_p} \vec{r}(\phi_p) \cdot \text{COV}_{3 \times 3}^{-1}(\phi_p) = 0$.

As a consequence, if the variables ϕ_p are chosen to be around the value which maximises $P(\vec{r})$, i.e. close to where the intersection of all tracks is expected, the dependence on ϕ_p of the integrand of Eq. 5.2 is only very weak since it depends only on variations of the covariance matrix COV_i as a function of $\phi_{p,i}$. So, if the values of $\vec{\phi}_p$ are chosen to correspond to points along the track not too far from the expected vertex, the integrand is nearly independent of them and, after integrating out these additional degrees of freedom, the vertex PDF can be simplified to:

$$P(\vec{r}) = \prod_{i=1}^{N_{trk}} \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i}))^T \text{COV}_{3 \times 3,i}^{-1}(\hat{\phi}_{p,i}) (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i})) \right]. \quad (5.3)$$

The maximization of $P(\vec{r})$ then corresponds now to the use of the following χ^2 based estimator:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i}))^T \text{COV}_{3 \times 3,i}^{-1}(\hat{\phi}_{p,i}) (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i})) \quad (5.4)$$

The easiest way to estimate the vertex position \vec{r} and its related error is to minimise this χ^2 function numerically. The auxiliary parameters $\vec{\phi}_p$ can be dealt with by repeating the fit again in case some of the initial parameters $\vec{\phi}_p$ correspond to positions along the tracks trajectories which are significantly displaced with respect to the final estimated vertex position \vec{r} .

This method is usually referred as global χ^2 minimisation. It can be significantly sped up by linearising the residuals in the χ^2 function and implementing all its derivatives analytically, but in general the speed of this approach is limited by the inversion of square matrices whose size is proportional to the number of tracks in the vertex fit.

5.3.1 Billoir method

Methods significantly faster than the global χ^2 based fit are available, relying on the *Billoir* method [61] or on a *Kalman* Filter [62]. They will be briefly described here.

Since the covariance matrices of the measured tracks are usually expressed in the reference frame of the helix parameters, it is convenient to rewrite Eq. 5.4 changing the reference frame from the three-dimensional space to the space of the helix parameters in the perigee definition:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} (\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i})^T \text{COV}_{5 \times 5,i}^{-1}(\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i}) \quad (5.5)$$

where now $\vec{q}_{meas,i}$ and $COV_{5 \times 5,i}$ are the track parameters in the perigee representation and the related covariance matrix of track number i , while $\vec{q}_{pred,i}$ represents the parameters the track would have if it would pass exactly through the vertex position \vec{r} .

While Eq. 5.4 involves a projection of the 5-dimensional helix parameter space to a three-dimensional space, Eq. 5.5 involves in the function $\vec{q}_{pred,i}$ a projection of the three-dimensional space position \vec{r} to the 5-dimensional helix parameter space: since in the second case such a projection is clearly not possible, the solution typically adopted for vertexing applications is to add N additional unknowns to the fit, namely the momenta of the tracks at the vertex position \vec{r} : $\vec{p}_1, \dots, \vec{p}_{N_{trk}}$. The reason is: given a vertex position and the track momentum at this position, the corresponding track parameters \vec{q} are uniquely defined. The momenta $\vec{p}_1, \dots, \vec{p}_{N_{trk}}$ can be then extracted in the fit itself, simultaneously with the vertex position, and will represent the best estimate of the track momenta at the vertex position under the hypothesis that the tracks pass through the vertex. This can be achieved by minimising the following modified χ^2 function:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} (\vec{q}_{pred,i}(\vec{r}, \vec{p}_i) - \vec{q}_{meas,i})^T \text{COV}_{5 \times 5,i}^{-1} (\vec{q}_{pred,i}(\vec{r}, \vec{p}_i) - \vec{q}_{meas,i}), \quad (5.6)$$

The functions $\vec{q}_{pred,i}(\vec{r}, \vec{p}_i)$ need to be linearised around a first estimate of the vertex position \vec{r}_0 and track momenta \vec{p}_i :

$$\begin{aligned} \vec{q} &= \vec{q}_0 + A [\vec{r} - \vec{r}_0] + B [\vec{p} - \vec{p}_0] = \\ &= \vec{q}(\vec{r}, \vec{p}) = \vec{q}(\vec{r}, \vec{p})_{\vec{r}=\vec{r}_0, \vec{p}=\vec{p}_0} + \underbrace{\frac{\delta \vec{q}(\vec{r}, \vec{p})}{\delta \vec{r}}}_{A} \Big|_{\vec{r}=\vec{r}_0} (\vec{r} - \vec{r}_0) + \underbrace{\frac{\delta \vec{q}(\vec{r}, \vec{p})}{\delta \vec{p}}}_{B} \Big|_{\vec{p}=\vec{p}_0} (\vec{p} - \vec{p}_0) \\ &= \vec{c}_0 + A \vec{r} + B \vec{p} \end{aligned} \quad (5.7)$$

where A and B are the Jacobians of the transformation from the space \vec{q} to (\vec{r}, \vec{p}) , evaluated in (\vec{r}_0, \vec{p}_0) . While in Ref. [61], these quantities are computed in form of a first order approximation around the inverse of the track curvature (which is typically expected to be a small quantity), in the next subsection of this thesis an exact computation at all orders of these quantities will be presented.

Eq. 5.7 is also called the *measurement* equation, since it represents the relation between the input variables (track parameters) and the fit output variables (position and track momenta at vertex position). Putting the linearised equation back into Eq. 5.6, the χ^2 function gets the form:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} (\vec{c}_{0,i} + A_i \vec{r} + B_i \vec{p}_i - \vec{q}_{meas,i})^T \text{COV}_{5 \times 5,i}^{-1} (\vec{c}_{0,i} + A_i \vec{r} + B_i \vec{p}_i - \vec{q}_{meas,i}). \quad (5.8)$$

As computed in Ref. [61], minimising the χ^2 by imposing the first derivatives with respect to \vec{r} and all \vec{p}_i to be zero, the following solution for \vec{r} is obtained:

$$COV(\vec{r}) = \left[\sum_i A_i^T \tilde{C}_i^{-1} A_i \right]^{-1} \quad (5.9)$$

$$\vec{r} = COV(\vec{r}) \left[\sum_i A_i^T \tilde{C}_i^{-1} (\vec{q}_{meas,i} - \vec{c}_{0,i}) \right] \quad (5.10)$$

$$\tilde{C}_i^{-1} = C_i^{-1} - C_i^{-1} B_i [B_i^T C_i^{-1} B_i]^{-1} B_i^T C_i^{-1}, \quad (5.11)$$

where for simplicity we have defined $\text{COV}_{5 \times 5} [\vec{q}_{meas,i}] = C_i$.

More in general, both the vertex position \vec{r} and the track momenta at the vertex position $(\vec{p}_1, \vec{p}_2, \dots, \vec{p}_{N_{trk}})$ can be obtained, together with their related full $3 + 3 \cdot N_{trk}$ covariance matrix.

The biggest advantage of the *Billoir* method is that it requires the final inversion of a matrix of size equal to the vector of variables to be determined in the fit, which, for a single vertex fit, is three. This implies a numerical complexity which scales merely with the number of tracks N_{trk} , rather than with N_{trk}^3 , as in the case of more conventional global χ^2 minimisation approaches. The solution of the vertex fit was formulated here in such a way to make the comparison with the Kalman filtering method easier: in the next section it will be demonstrated that the *Billoir* method is formally equivalent to a Kalman filter in the *weighted* mean formalism.

5.3.2 Kalman filter

The first application of a Kalman filter to track and vertex fitting tasks was presented in Ref. [62]. Since then a wide use of Kalman filters was made in the high energy physics community. Here only the basic principles will be recalled to introduce some of the topics developed in this thesis.

A Kalman filter is usually applied to linear dynamic systems, which at every step k are described by a state vector \vec{r} . The evolution of the state vector from step $k - 1$ to step k is described by a linear transformation:

$$\vec{r}_k = F_{k-1} \cdot \vec{r}_{k-1} + \vec{w}_{k-1}, \quad (5.12)$$

where F_{k-1} is the propagation matrix and w_{k-1} is a random noise contribution to the system during the propagation step. At every step a new measurement constrains the system, where the measurements \vec{m}_k needs to be described as a linear function of the state vector \vec{x}_k with associated Gaussian noise \vec{e}_k :

$$\vec{m}_k = H_k \vec{x}_k + \vec{e}_k \quad (5.13)$$

Given such a linear dynamic system, the Kalman filter provides:

- Filtering: the optimal recursive estimator of the *present* state vector \vec{x} given all previous measurements.
- Prediction: the optimal estimation of the state vector at a future time, given all past measurements.
- Smoothing: an improved estimation of the state vector at some time in the past, given all measurements up to the *present* time.

Due to these properties, Kalman filters were first applied for tracing air crafts in radar applications, but they are indeed very interesting also for being applied in high energy physics, in particular for track and vertex fitting applications. In the case of track fitting, the various steps k need to be interpreted as the various hits to be added to a track along the track path, the state vector represents the track helix as parametrised on the measurement surface corresponding to the track hit to be added, the measurement equation (Eq. 5.13) represents the relation between the measured hit and the track helix on the corresponding measurement surface: at each step k , the track helix is propagated to a new measurement surface, through

the propagation matrix F_k and adding both the noise eventually corresponding to multiple scattering and eventual energy losses due to radiation, and the new measurement is added in linearised form, according to the linearisation matrix H_k . After all hits of a track have been considered, at the last measurement surface the best estimation of the track trajectory is obtained. In order to propagate this knowledge back to the first measurement surface, the technique of smoothing is used. While the *progressive* fitting method for tracking was already known at the time of Ref. [63], it is the introduction of the Kalman filtering techniques in the tracking domain that allowed smoothing to be applied.

In the case of vertexing the application of a Kalman filter is even more straightforward. The single step k is interpreted as the addition of a track to the vertex. The state vector is given by the vertex position and track momentum corresponding to the track added at step k :

$$\vec{x}_k = (\vec{r}, \vec{p}_k) \quad (5.14)$$

So, after a track is iteratively included in the fit, the Kalman filter provides the estimation of the vertex position and related covariance matrix corresponding to the minimisation of the vertex χ^2 of Eq. 5.8 including all tracks considered up to that point. After all tracks are added to the fit, the problem of estimating the vertex position is solved. The same does not hold for the track momenta evaluated at the estimated vertex position, which should provide an improved estimate of the track momenta given by constraining the tracks to pass through the vertex: since these momenta are updated only at a certain step k , they do not take into account all changes in the vertex position taking place at all steps after k . In order to get a complete estimate of them, the final result needs to be propagated back to earlier Kalman update steps, which is what is usually done by the *smoothing* step. The smoothing procedure is also needed to get a correct estimation of the χ^2 contribution given by a single track k , which needs again to rely on the final estimate of the vertex position.

As an alternative, the state vector could in principle include the vertex position plus the momenta at the vertex position of all tracks included in the fit: in this way the definition of the state vector would remain the same for all steps k . However, this would make the fit unnecessary complex, increasing the dimensionality of the matrices to invert from $3 + 3$ to $3 + N_{trk} \times 3$ ⁽⁴⁾; at every single step k , there is no direct relation in the measurement equation between the track parameters \vec{q}_k of track k with the momenta p_k at the vertex of all other tracks in the fit and all the correlations between the momenta at the vertex of different tracks come from variations of the position of the vertex itself. So there is no approximation in restricting the state vector at step k only to the vertex position and the momentum of the track at the vertex being added to the fit. This principle will be also exploited by the JetFitter algorithm described later in this thesis.

In the case of a vertex fit the evolution equation becomes trivial:

$$\vec{x}_k = \vec{x}_{k-1}, \quad (5.15)$$

since no propagation and no noise addition needs to be performed, while the measurement equation gives the linear relation between the track measurement in terms of helix parameters and the vertex position and track momentum defined at the vertex:

$$\vec{m}_k = H_k \vec{x}_k = \vec{c}_0 + A\vec{r} + B\vec{p}_k, \quad (5.16)$$

and was already presented in Eq. 5.7.

⁴This is only the case if the *weighted mean* formalism is adopted, which will be described later.

Since the Kalman filter is an iterative procedure, the components of the state vector (the vertex position and track momenta at vertex) need to be initialised: when the initial estimate is very poor, it is sufficient to inflate the error matrix in order to make the influence of the starting state vector negligible in the fit:

$$\vec{x}_0 = (\vec{r}_0, \vec{p}_{k,0}) \quad (5.17)$$

$$C_0 = COV(\vec{x}_0) = \left(\frac{1}{\delta_1} I_3, \frac{1}{\delta_2} I_3 \right) \quad (5.18)$$

where δ_1 and δ_2 are small numbers, while I_3 is the unit matrix of size 3.

At each step k , the tracks are iteratively added to the fit through a Kalman update step. There are two alternative approaches, the *weighted mean* formalism and the *gain* formalism. In the first one, the new state vector is obtained by combining the knowledge of the state vector \vec{x}_{k-1} at step $k-1$ with the new contribution coming from the new measurement at step k , re-weighted by the weight matrix C_{k-1}^{-1} of step $k-1$ and by the weight matrix of the new measurement V_k^{-1} , respectively:

$$C_k^{-1} = C_{k-1}^{-1} + H_k^T V_k^{-1} \vec{m}_k \quad (5.19)$$

$$\vec{x}_k = C_k [C_{k-1}^{-1} \vec{x}_{k-1} + H_k^T V_k^{-1} \vec{m}_k] \quad (5.20)$$

This result is quite intuitive and can be easily derived by just multiplying a multi-Gaussian distribution of probability in the state vector with a multi-Gaussian distribution of probability of the constraint on the state vector given by the additional measurement at the step k , and solving for the new state vector provided by maximising their combined probability.

In the *gain* formalism the new state vector is obtained in an alternative way and the difference between the previous and new state vector is obtained applying the gain matrix on top of the measurement residual:

$$\vec{x}_k = \vec{x}_{k-1} + K_k [\vec{m}_k - H_k \vec{x}_{k-1}], \quad (5.21)$$

where the *gain* matrix K_k is defined is:

$$K_k = C_{k-1} H_k^T [V_k + H_k C_{k-1} H_k^T]^{-1} \quad (5.22)$$

and the covariance matrix C_k is obtained by:

$$C_k = (1 - K_k H_k) C_{k-1} \quad (5.23)$$

There are two main differences between the two mentioned approaches:

- The *weighted mean* formalism requires at each step k the inversion of a matrix of size equal to the dimensionality of the state vector, while the *gain* formalism requires the inversion of a matrix of size equal to the dimensionality of the measurement to be added to the fit.
- The *weighted mean* formalism is expressed in terms of the weight matrices both of the state vector at step $k-1$ and of the weight matrix of the measurement to be added at step k , so that it can accommodate an initial state vector with infinite error. On the contrary, in the *gain* formalism the *gain* matrix is expressed in terms of covariance matrices, so that it can accommodate a measurement with zero associated error, which corresponds to adding an exact constraint to the fit (for a more formal derivation see Ref. [64]).

In general the *weighted mean* formalism is faster if the state vector is smaller than the measurement vector at a certain step k , while the *gain* formalism is faster in the other case. In the case of the single vertex fit, the *weighted mean* formalism is usually used, because it allows to compute the Kalman update step of the state vector in the limit of assigning an infinite error to the initial track momentum \vec{p}_k at the vertex at step k . However, in the course of the present thesis, also the *gain* formalism will be very useful, when applying exact constraints on the dedicated Kalman filter implemented in the JetFitter algorithm.

In the following the formulae to apply the Kalman filter in the *weighted mean* formalism to a single vertex fit will be briefly re-derived. This consists of applying the Kalman update step of Eq. 5.20, substituting the measurement matrix H_k with the expression of Eq. 5.16 and separating the state vector and its covariance matrix into two components. Before performing the Kalman update at a generic step k the state vector and related covariance matrices are defined as:

$$\vec{x}_{k-1} = (\vec{r}_{k-1}, \vec{p}_{k-1}) \quad (5.24)$$

$$C_{k-1} = \left(\text{COV}(\vec{r})_{k-1}, \frac{I_3}{\delta} \right), \quad (5.25)$$

where I_3 is a unit matrix of size 3 and δ is an arbitrarily small number. The update step can be performed as:

$$C_k = \begin{bmatrix} \text{COV}(\vec{r})_{k-1}^{-1} + A^T \text{COV}(\vec{q})_k^{-1} A & A^T \text{COV}(\vec{q})_k^{-1} B \\ B^T \text{COV}(\vec{q})_k^{-1} A & \delta \cdot I_3 + B^T \text{COV}(\vec{q})_k^{-1} B \end{bmatrix}^{-1} \quad (5.26)$$

$$\begin{bmatrix} \vec{r}_k \\ \vec{p}_k \end{bmatrix} = C_k \begin{bmatrix} \text{COV}(\vec{r})_{k-1}^{-1} \vec{r}_{k-1} + A^T \text{COV}(\vec{q})_k^{-1} (\vec{q}_{meas} - \vec{c}_0) \\ \delta \cdot \vec{p}_{k-1} + B^T \text{COV}(\vec{q})_k^{-1} (\vec{q}_{meas} - \vec{c}_0) \end{bmatrix} \quad (5.27)$$

After some matrix algebra manipulation and considering the limit $\delta \rightarrow 0$, the result for the vertex position and related covariance matrix is:

$$\text{COV}(\vec{r})_k = [\text{COV}(\vec{r})_{k-1}^{-1} + A^T \tilde{\text{C}}(\vec{q})_k^{-1} A]^{-1} \quad (5.28)$$

$$\vec{r}_k = \text{COV}(\vec{r})_k [\text{COV}(\vec{r})_{k-1}^{-1} \vec{r}_{k-1} + A^T \tilde{\text{C}}(\vec{q})_k^{-1} (\vec{q}_{meas} - \vec{c}_0)], \quad (5.29)$$

where $\tilde{\text{C}}(\vec{q})_k^{-1}$ is for simplicity defined as:

$$\tilde{\text{C}}(\vec{q})_k^{-1} = \text{COV}(\vec{q})_k^{-1} - \text{COV}(\vec{q})_k^{-1} B [B^T \text{COV}(\vec{q})_k^{-1} B]^{-1} B^T \text{COV}(\vec{q})_k^{-1}. \quad (5.30)$$

Comparing Eq. 5.29 with the solution provided by the Billoir method in Eq. 5.11, it is straightforward to realise that the Kalman formalism in the weighted mean approach represents the iterative version of the Billoir solution: this shows that the two approaches are completely equivalent. Actually the comparison is useful, since it clearly shows that, if one can release the requirement of computing the vertex position iteration after iteration, computing everything iteratively in terms of $\text{COV}^{-1}(\vec{r})_k \cdot \vec{r}_k$, one can reduce the number of required inversions of the covariance matrix of the vertex position by a factor N_{trk} . This has been considered when implementing the Kalman filter formalism in the vertexing framework of the ATLAS software ATHENA: the inversion is performed only on request, when the vertex position needs finally to be obtained.

The computation of the momenta of the tracks at the vertex and of the χ^2 contribution of all tracks is then usually done during the final smoothing steps, since for their correct

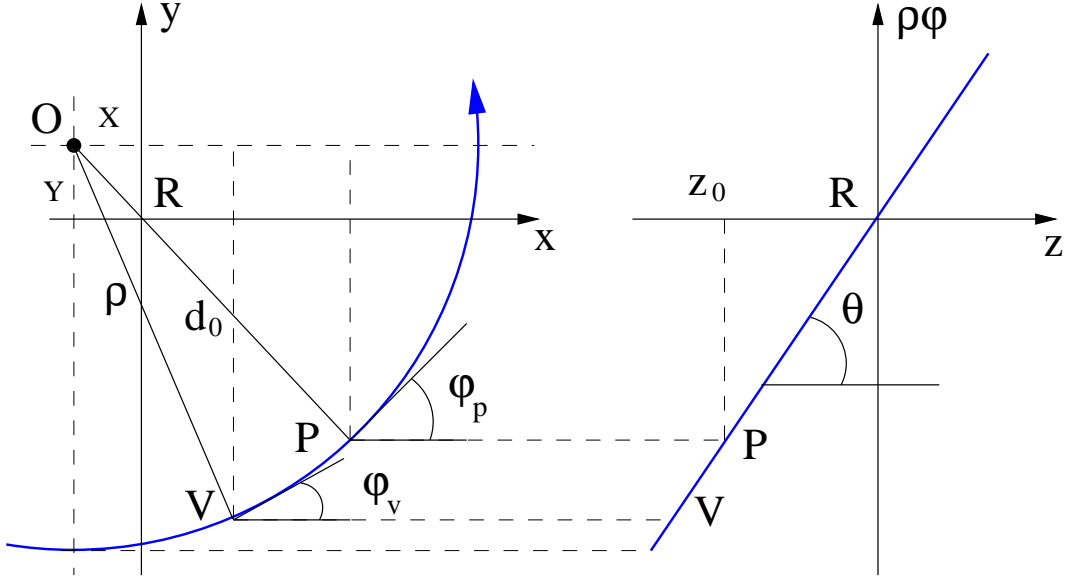


Figure 5.2: Track helix in the transverse (left) and in the $\rho\phi$ plane (right). The track is defined in the *perigee* representation at the point of closest approach P to the reference position R .

estimation the final fitted vertex position is needed. Their explicit formulation can be found in Ref. [62].

While the Kalman filter approach in the *weighted mean* formalism as illustrated here was implemented for primary vertex fitting applications as presented in the present chapter, a mathematical extension of it was worked out to define the JetFitter algorithm for b -tagging, as presented in Section 6.6.

5.3.3 Linearised measurement equation: computation of Jacobians

An important input to the vertex fit is the parametrisation of the track trajectories. For each track, the trajectory, following the equation of motion of a charged particle in a constant magnetic field, needs to be linearised around a certain position of space \vec{r}_0 and around a certain value of the track momentum \vec{p}_0 .

In the *perigee* representation the trajectory is defined at the point of closest approach P along the track with respect to a reference position R , as shown in Fig. 5.2, according to the ATLAS reference frame introduced in Section 3.2. At the point P five helix parameters \vec{q} are defined:

- d_0 : signed transverse impact parameter (the distance from point R to P in the transverse plane);
- z_0 : longitudinal impact parameter (the distance from point R to P along the z axis);
- ϕ_p : azimuthal angle of trajectory at P , defined in $[-\pi; \pi]$;
- θ_p : polar angle of trajectory at P , defined in $[0; \pi]$;
- q/p : ratio of the charge of the particle to its full momentum.

The sign of the impact parameter d_0 is defined in such a way that, if ϕ_0 is the azimuthal angle expressing the direction of the vector \vec{OP} in the transverse plane as shown in Fig. 5.2,

the sign is positive if $\phi_0 - \phi_p = \frac{\pi}{2} + 2\pi n$ ($n \in \mathbb{Z}$) and it is negative otherwise. In addition, the signed curvature is defined as:

$$\rho = \frac{\sin(\theta)}{\frac{q}{p}B_z}. \quad (5.31)$$

Given the ATLAS reference frame, in the inner detector the magnetic field is in good approximation entirely directed along the z axis, a positive curvature will correspond to a clockwise rotation and a negative curvature to a counter-clockwise rotation.

A generic position V on the trajectory can be parametrised in terms of the trajectory parameters, plus an additional parameter, which defines the point along the track to be considered: for simplicity, the transverse direction ϕ_V of the momentum at the point V of the trajectory is chosen. The spatial coordinates of this point V can then be expressed as:

$$\begin{aligned} x_V &= x_R + d_0 \cos\left(\phi_p + \frac{\pi}{2}\right) + \rho \left[\cos\left(\phi_V + \frac{\pi}{2}\right) - \cos\left(\phi_p + \frac{\pi}{2}\right) \right] \\ y_V &= y_R + d_0 \sin\left(\phi_p + \frac{\pi}{2}\right) + \rho \left[\sin\left(\phi_V + \frac{\pi}{2}\right) - \sin\left(\phi_p + \frac{\pi}{2}\right) \right] \\ z_V &= z_R + z_0 - \frac{\rho}{\tan(\theta)} [\phi_V - \phi_p] \end{aligned} \quad (5.32)$$

where (x_R, y_R, z_R) defines the reference position R .

The linearisation of the measurement equation requires the calculation of the matrices of derivatives $A = \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(x_V, y_V, z_V)}$ and $B = \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(\phi_V, \theta, q/p)}$. In order to compute them, the perigee parameters $(d_0, z_0, \phi_P, \theta_P, (q/p)_P)$ need to be expressed as a function of $(x_V, y_V, z_V, \phi_V, \theta_V, (q/p)_V)$:

$$\begin{aligned} d_0 &= \rho + \text{sgn}(d_0 - \rho) \sqrt{(x_V - x_R - \rho \cos(\phi_V + \frac{\pi}{2}))^2 + (y_V - y_R - \rho \sin(\phi_V + \frac{\pi}{2}))^2} \\ \phi_P &= \arctan\left(\frac{y_V - y_R - \rho \sin(\phi_V + \frac{\pi}{2})}{x_V - x_R - \rho \cos(\phi_V + \frac{\pi}{2})}\right) \\ z_0 &= z_R + z_V + \frac{\rho}{\tan(\theta)} [\phi_V - \phi_P(x_V, y_V, \phi_V, \theta, q/p)] \\ \left(\frac{q}{p}\right)_P &= \left(\frac{q}{p}\right)_V \\ \theta_P &= \theta_V \end{aligned} \quad (5.33)$$

To simplify the above formulae the following expressions are defined:

$$\begin{aligned} X &= x_V - x_R - \rho \cos\left(\phi_V + \frac{\pi}{2}\right) \\ Y &= y_V - y_R - \rho \sin\left(\phi_V + \frac{\pi}{2}\right) \\ S &= \sqrt{X^2 + Y^2} \end{aligned} \quad (5.34)$$

where X and Y can be interpreted geometrically as shown in Fig. 5.2. In addition, for all tracks considered, $|\rho| \gg |d_0|$, so that $\text{sgn}(d_0 - \rho) = -\text{sgn}(\rho)$.

In order to get the Jacobians A and B to feed into Eq. 5.16, the derivatives of $(d_0, z_0, \phi_P, \theta_P, (q/p)_P)$ with respect to $(x_V, y_V, z_V, \phi_V, \theta_V, (q/p)_V)$ need to be computed, as explicitly derived in the Appendix in Section A.1.1.

Given the result of this computation, it turns out that it is useful to define:

$$R = Y \sin(\phi_V) + X \cos(\phi_V) \quad (5.35)$$

$$Q = X \sin(\phi_V) - Y \cos(\phi_V)$$

Applying the identities of Eq. 5.35, the final result for the Jacobian matrices can be obtained in a very compact form:

$$A = \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(x_V, y_V, z_V)} = \begin{pmatrix} -h \frac{X}{S} & -h \frac{Y}{S} & 0 \\ \frac{\rho}{\tan \theta} \frac{Y}{S^2} & -\frac{\rho}{\tan \theta} \frac{X}{S^2} & 1 \\ -\frac{Y}{S^2} & \frac{X}{S^2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (5.36)$$

and

$$B = \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(\phi_V, \theta, q/p)} = \begin{pmatrix} -\frac{h\rho}{S} R & \frac{\rho}{\tan \theta} [1 - \frac{h}{S} Q] & -\frac{\rho}{q/p} [1 - \frac{h}{S} Q] \\ \frac{\rho}{\tan \theta} [1 - \frac{\rho}{S^2} Q] & \rho [\Delta\phi + \frac{\rho}{S^2 \tan^2 \theta} R] & \frac{\rho}{q/p \tan \theta} [\Delta\phi - \frac{\rho}{S^2} R] \\ \frac{\rho}{S^2} Q & -\frac{\rho}{S^2 \tan \theta} R & \frac{\rho}{S^2 q/p} R \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where the quantities h and $\Delta\phi$ are defined as:

$$\begin{aligned} h &= \text{sgn}(\rho) \\ \Delta\phi &= \phi_P - \phi_V. \end{aligned} \quad (5.37)$$

Using the explicit formulae for the Jacobians just derived, the matrices A and B and the constant term \vec{c}_0 of Eq. 5.7 need to be computed individually for every track, in the vicinity of the linearisation point, which usually represent the best estimate of the vertex position at the time the track is linearised. The first essential ingredient is defining the point \vec{r}_0 along the track, with respect to which the track needs to be linearised. In the actual implementation this point is defined at the point of closest approach to the estimated vertex position in the transverse plane⁵. Once the position \vec{r}_0 along the track is defined, also the track momentum corresponding to this position, \vec{p}_0 , is computed and the Jacobians A and B are evaluated, by choosing the coordinates x_V, y_V, z_V of the point V at which the Jacobian is computed to be equal to \vec{r}_0 and the track momentum at the point V to be equal to $\phi_V, \theta_V, q/p$. The parameters $d_0, z_0, \phi_P, \theta_P$ and q/p used in the evaluation of the Jacobians are just the original helix parameters of the track. Finally the constant term \vec{c}_0 needs to be computed. This is achieved exploiting the definition of \vec{c}_0 implicitly assumed in Eq. 5.7:

$$\vec{c}_0 = \vec{q}_0 - A \cdot \vec{r}_0 - B \cdot \vec{p}_0 \quad (5.38)$$

If the reference point R at which the track parameters are defined (see Fig 5.2) is too far away from the linearisation point to be used, then the track is extrapolated directly to the

⁵This choice is motivated by the fact that the impact parameter resolution in the transverse plane is significantly better than the longitudinal impact parameter resolution and thus using the point of closest approach in the transverse plane yields better results than using the closest point to the linearisation point along the track in full three dimensions. If the reference point R of the track is chosen to be the actual estimated vertex position, this choice is equivalent to putting $\phi_V = \phi_P$ and therefore $\Delta\phi = 0$.

linearisation point using the ATLAS tracking extrapolation tool [65]. The equations of motion used for the computation of the Jacobians are based on the hypothesis that the magnetic field is always directed only in the z direction and that it is unchanged in different positions of the track helix, which is only correct on a sufficiently short range. As opposed to that, the ATLAS tracking extrapolation tool makes use of a complete propagation of the track through the ATLAS magnetic field, relying on the Runge-Kutta technique [66], and takes the influence of the detector material on the extrapolated parameters and errors into account. In this case the linearisation procedure, as outlined before, is applied directly to the new track parameters after extrapolation.

In Section A.1.2 of the Appendix the Jacobians A and B computed in this sections are first expanded in terms of the inverse of the track curvature ρ and the first order result is compared with the approximated result for large curvatures available in Ref. [61], as a first cross-check of the new computation: perfect agreement is found.

5.3.4 Jacobians for a neutral track

For particular applications, feeding *neutral tracks* into a vertex fit can be very useful, as will be described in the present thesis work when applying vertexing to b -tagging, where a neutral track can represent for example a K^0 meson or a Λ baryon, reconstructed at their decay vertex position through their charged decay products. This requires the definition of a neutral trajectory and the computation of its Jacobians. However, the linearisation of a neutral trajectory can be obtained from the charged case, taking the approximate result for large curvatures (presented in Section A.1.2), and then setting $\frac{1}{\rho} = 0$, which corresponds to keeping only the constant term of the Taylor expansion in $\frac{1}{\rho}$. In addition, the parameter q/p is substituted by $1/p$, since the sign of the curvature has no relevance in the limit for $\rho \rightarrow \infty$.

The result for the position and momentum Jacobian is reported here:

$$A = \frac{\partial(d_0, z_0, \phi_P, \theta_P, 1/p)}{\partial(x_V, y_V, z_V)} \quad (5.39)$$

$$= \begin{pmatrix} -\sin(\phi_V) & \cos(\phi_V) & 0 \\ -\cos(\phi_V)\cot(\theta) & -\sin(\phi_V)\cot(\theta) & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$B = \frac{\partial(d_0, z_0, \phi_P, \theta_P, 1/p)}{\partial(\phi_V, \theta, 1/p)} \quad (5.40)$$

$$= \begin{pmatrix} -\tilde{R} & 0 & 0 \\ \cot(\theta)\tilde{Q} & \frac{\tilde{R}}{\sin^2(\theta)} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where \tilde{R} and \tilde{Q} are as defined as:

$$\tilde{R} = (y_V - y_R)\sin(\phi_V) + (x_V - x_R)\cos(\phi_V) \quad (5.41)$$

$$\tilde{Q} = (x_V - x_R) \sin(\phi_V) - (y_V - y_R) \cos(\phi_V), \quad (5.42)$$

in such a way that they remain finite in the limit $\rho \rightarrow \infty$.

As a cross-check, the same result has been also obtained directly, considering the equation of motions of a neutral particle and computing the derivatives directly.

5.3.5 Adaptive methods for vertex fitting

The mathematical determination of the vertex position is usually based on the minimisation of the track trajectories residuals with respect to the estimated vertex position, based on the χ^2 already defined in Eq. 5.5, which, for the purpose of the present discussion, can be simplified to:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} \sum_{j=1}^5 \frac{[\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i}]_j^2}{[\vec{\sigma}_i]_j^2}, \quad (5.43)$$

where the covariance matrices, to simplify the notation, have been assumed to be diagonal.

If the measurements of all track trajectories contributing to the vertex fit are Gaussian distributed around their true position in space, the minimisation of the χ^2 function above, independently on the method used to accomplish this scope, provides an efficient and unbiased estimator for the vertex position. However, since the χ^2 function depends quadratically on the trajectory residuals with respect to the primary vertex position, the fit can be significantly biased by the presence of outlying measurements (*outliers*). These can be of two kinds:

- Badly reconstructed tracks, for example because of wrong hit to track assignments during the tracking pattern recognition step.
- Correctly reconstructed tracks, but representing charged particles originating in different and sufficiently displaced decay vertex position.

Tracks corresponding to outlying measurements can be significantly displaced with respect to the vertex to be determined in the fit. A measurement of this displacement is provided by the contribution to the overall vertex fit χ^2 of each individual track. This follows a χ^2 distribution with two degrees of freedom⁶ and can be easily translated in terms of Gaussian significances.

The (in)sensitivity of a vertex fit to such outlying measurements is usually defined as *robustness*.

One way to reduce such a dependence on outlying measurements is to deal with it during the vertex *finding* step, repeating the minimisation of the vertex fit χ^2 of Eq. 5.43 more than once and using an estimate of the compatibility of the track to the present vertex position as a criterion to include or exclude a certain track in the fit itself. This is however part of the *finding* procedure, which is described in the next section.

An alternative approach to accomplish the same goal is to modify directly the vertex fitting procedure itself, in such a way to make it *adaptively* reduce the influence of outliers. A

⁶The helix parameters contain five degrees of freedom, but three of them are eaten up to obtain the track momentum at the vertex.

modified χ^2 function can be defined, where the single track contributions are re-weighted by a factor which depends on their compatibility with the last estimated vertex position:

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} \omega(\hat{\chi}_i^2(\vec{r})) \sum_{j=1}^5 \frac{[\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i}]_j^2}{[\vec{\sigma}_i]_j^2} \quad (5.44)$$

The weight factor for a single track is typically defined as a function of the χ^2 contribution defining the probability of the track to be compatible with the vertex position. This makes the new function to minimise non-linear with respect to the sum of squared residuals of the uncorrected χ^2 function of Eq. 5.43. A practical way to minimise the new function is to repeat the χ^2 minimisation iteratively, with the weight factors ω computed with respect to the residual track χ^2 contribution from the previous iteration (denoted $\hat{\chi}^2$), until convergence is reached. This also justifies the interpretation of the weight factor ω as the a-priori probability of the track to be correctly assigned to the vertex to fit.

However there is a severe limitation in this approach. During the first fit iterations, where the fit is still potentially largely biased by outlying measurements, the estimated vertex position can be significantly wrong: if the track weight factors ω to be used for later fit iterations are based on this estimate, they can either lead to an early exclusion of tracks which would be perfectly compatible with the real vertex or to an inefficient down-weighting of real outliers. A possible solution to this problem is to make use of a thermodynamic annealing procedure, where the weight factors ω are made dependent not only on $\hat{\chi}^2$, but also on a temperature value, which is very high during the first iterations, when the vertex estimate is supposed to be very bad, and is lowered gradually following the fit iterations. The dependence of ω on the temperature is then defined in such a way that, with high temperature values, only a light preference is given to tracks compatible with the actual estimated vertex position, while, with lower temperature values, the track down-weighting gets stricter and stricter, corresponding nearly to completely dropping all incompatible tracks and keeping all remaining tracks with weight $\omega = 1$.

Adaptive Vertex Fitter

The *Adaptive Vertex Fitter* (AVF), which was implemented for the first time in the CMS software framework [67], is based on a weight function $\omega(\hat{\chi}^2, T)$ of the following kind:

$$\omega(\hat{\chi}^2) = \frac{1}{1 + \exp\left(\frac{\hat{\chi}^2 - \chi_{cutoff}^2}{2T}\right)}, \quad (5.45)$$

which corresponds to a Fermi function, where the value of transition between region of high and low weight is given by the cutoff value χ_{cutoff}^2 and the transition can be made arbitrarily from smooth to sharp by changing the temperature value T from very high to very low numbers. The weight factor is shown as a function of $\hat{\chi}^2$ for few different choices of the value of the temperature in Fig. 5.3. In the limit of $T \rightarrow 0$, the weight factor becomes a simple step function, with weight equal to one if $\hat{\chi}^2 < \chi_{cutoff}^2$ and 0 otherwise; in the vertex fit, this would correspond to dropping all tracks with $\hat{\chi}^2 > \chi_{cutoff}^2$.

The procedure used to decrease the temperature value T with increasing fit iterations is the chosen *annealing scheme*. The simplest possible choice is the implementation of a *deterministic annealing scheme*, where a set of predefined temperature values in decreasing order is chosen, and, starting from the first one, after every new fit iteration, the next in the set is used, and the fit iterations stop after the last temperature value in the set is used. In this annealing

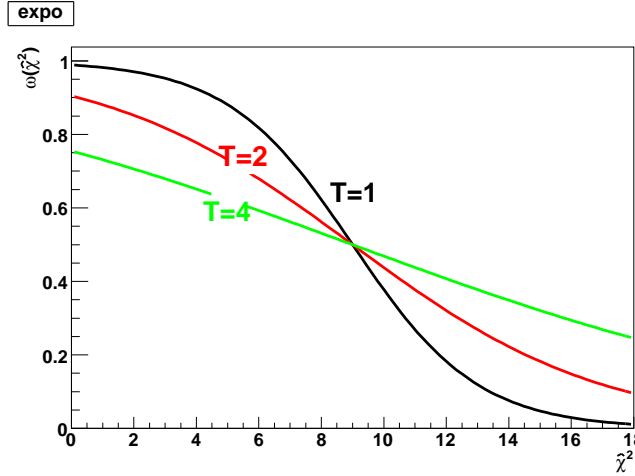


Figure 5.3: Weight factor, as a function of $\hat{\chi}^2$, and for few different values of the temperature T .

scheme both the set of temperatures and the value of $\hat{\chi}^2$ used in the weight factor ω can be tuned to give the best possible performance. The result of this tuning is illustrated in the appendix (Section A.3).

Once the weight factors in Eq. 5.44 are defined, the minimisation of the vertex fit χ^2 can be brought back to the form of Eq. 5.43 by multiplying the weight matrices of the single track parameters by the weight factors associated to the respective tracks. This makes the adoption of the Kalman filter approach of Section 5.3.2 straightforward also for being applied to the Adaptive Vertex Fitter. The value of $\hat{\chi}^2$ used to compute the weight factor ω one fit iteration later can be obtained either by *smoothing* the track according to the Kalman filter formalism, or by alternative faster but slightly more approximate methods.

One problematic issue with the iterative fitting procedure as outlined so far is how to get a good initial estimate of the vertex position, in order to avoid the need both of starting with extremely high temperature values and of considering a very large set of temperature values, which would make the fit extremely slow and inefficient. In order to achieve this, a vertex *seed finder* algorithm needs to be defined, which already includes a certain *robustness* against trajectories far away from the core space zone intersected by most of the tracks. This is part of the *finding* strategy and will be described in some detail in the next section.

Adaptive Multi-Vortex Fitter

An extension of the Adaptive Vertex Fitter allows the simultaneous fit of N tracks to M vertices, the Adaptive Multi-Vortex Fitter (AMVF), which was proposed in Ref. [68]. This vertex fitting algorithm is particularly useful when the experimental resolution is comparable with the distances among the vertices to be determined in the fit, which is a frequent scenario in proton-proton collisions at the LHC.

When performing a single vertex fit, as already mentioned in the present section, a significant degradation of the vertex resolution and a non-Gaussian behaviour of the fit result is expected due to outlying measurements, which can be of two kinds, either mismeasured tracks or real outliers. In the case of a single vertex fit, the influence of both kind of outliers can be reduced only by down-weighting the influence in the fit of tracks which are incompatible with a previous estimate of the fitted vertex position; this is what the Adaptive Vertex Fitter

is designed to do. However, outlying measurements of the second kind can be rejected more efficiently if the fit is able to classify these tracks as belonging to a different vertex, which is possible, as long as this additional vertex is sufficiently displaced from the first vertex and is intersected by at least two tracks. As a consequence, the optimal solution to the problem of vertex fitting with a significant amount of outliers originating in different but closely positioned displaced vertices is the simultaneous fit of such vertices, dynamically assigning the tracks to the vertices they are most compatible with.

The simplest solution to the problem of simultaneously fitting N tracks to M vertices is provided by the Adaptive Multi-Vertex Fitter, which consists in M simultaneous fits (equivalent to M simultaneous instances of the Adaptive Vertex Fitter), where every single vertex fit is based again on the modified χ^2 function of Eq. 5.44, but with a weight factor defined as:

$$\omega_{\text{vtx}}(\hat{\chi}_i^2) = \frac{\exp\left(-\frac{\hat{\chi}^2(i \rightarrow \text{vtx})}{2T}\right)}{\sum_{j=1}^M \exp\left(-\frac{\hat{\chi}^2(i \rightarrow j)}{2T}\right) + \exp\left(-\frac{\hat{\chi}_{\text{cutoff}}^2}{2T}\right)}, \quad (5.46)$$

where $\hat{\chi}^2(i \rightarrow j)$ represents the compatibility of track i with vertex j . This choice of the weight factor provides a simple extension of Eq. 5.45 and can be intuitively seen as a measure of the probability of the track i to be compatible with the considered vertex, divided by the sum of probabilities to be compatible with any of the M vertices plus the prior probability of the track to be an outlying measurement. During a single multi-vertex fit iteration, the M simultaneous fits are completely independent from each other, but the weight factors which are based on the previous fit iteration do correlate all vertex fits together.

While the Adaptive Multi-Vertex fit provides a very interesting way to try to optimally determine the position of M vertices out of N tracks, by dynamically re-weighting their assignment to the M vertices, it does not provide a solution of the underlying *pattern recognition* problem. In fact, it does not provide a solution to the problem of determining the number of vertices to fit and a first estimate of their position, which is absolutely necessary to start a first iteration of the multi-vertex fit. A possible solution to this problem will be discussed as part of the vertex finding algorithms proposed in the course of this thesis in the next section.

5.4 Vertex finding algorithms

The vertex finding algorithms are responsible for the overall primary vertex reconstruction process: they take all selected tracks as input, cluster them into a set of separate vertices and then make use of specific vertex fitting algorithms to determine their precise vertex positions. While the mathematical core of the vertex reconstruction problem is certainly represented by the vertex fitting algorithms, several vertex finding algorithms are implemented to solve the underlying *pattern recognition* problem and they will be shortly described in the following. Since the *adaptive* vertex finding methods heavily rely on the determination of a good vertex seed, the seed finding algorithms and their performance are described before introducing the adaptive vertex finders.

5.4.1 Default vertex finder

The default vertex finder (known in ATLAS under the name of *InDetPriVxFinderTool*) was already implemented in the ATLAS common vertexing framework at the time this thesis work started and it will be therefore used as a reference for all other vertex finders. It implements an approach where the vertex finding and fitting problems are completely decoupled, denoted

as *fitting after finding*, and it is usually based on the Billoir method for the vertex fitting part (which was described in Section 5.3.1), but it can make use equally well of the Kalman filter method, thanks to the modularity of the vertexing framework.

As a first step, which is common to all vertex finding algorithms, a track selection is applied to the input tracks: the aim is both to reduce the amount of *fake* or badly measured tracks and to ensure that tracks which pass sufficiently far away from the interaction region are not included in the fit, in order not to slow down the vertex fitting process. More details on the track selection step will be given in Section 5.5.1.

The tracks are clustered into separate vertices before any vertex fit is started. For each input track passing the selection a z position is defined: since the beam spot centre provides already an excellent approximation for the transverse position of any signal or pile-up primary vertex, the z position is defined for each track at its point of closest approach to the beam spot centre in the transverse plane (this corresponds to the z_0 track parameter in the perigee frame). Once the z position of every track is defined, all these positions are passed to a specific vertex clustering tool which clusters the tracks into separate vertices. Many different implementations of this vertex clustering tool exist: only the default approach will be mentioned here and used for the results presented in this thesis, which makes use of a simple *sliding window* algorithm, where a fixed window size in the z direction of 5 mm is used to cluster all tracks in the same window into a common vertex, making the window slide from low to high values of the z coordinate.

After the tracks are clustered into independent set of tracks, they are processed independently and in consecutive order: each set of tracks, with related initial vertex position estimate, is passed to the vertex fitting tool, so that each independent set of tracks results in a single vertex.

In order to reduce the influence of outliers due to badly measured tracks or secondary tracks or tracks belonging to a different primary interaction vertex, after a first vertex fit on a specific set of tracks was performed, tracks with a value of χ^2_{cut} below a certain value (as a default a cut on $\chi^2_{cut} < 6$ is used, corresponding to $\text{Prob}(\chi^2_{cut}) < 5\%$ ⁽⁷⁾) are removed from the fit and the fit is repeated iteratively, until no such tracks are present anymore.

In addition, usually the beam spot constraint is used during the fitting procedure, which means that an additional spatial three-dimensional Gaussian measurement centred around the beam spot center and with the beam spot size as width is applied during the vertex fit. Due to the fact that the beam spot constraint is integrated directly into the vertex fit, secondary tracks, which are typically incompatible with the beam spot on the transverse plane, will get a very bad χ^2 value and will thus be removed from the fit by the iterative fitting procedure itself. As a side remark, the use of the beam spot constraint permits to define a vertex position with a single track.

The main difference between this approach and the vertex finding procedures which will be introduced in the following sections is that, once the tracks are removed from the vertex they were associated to, they are not considered for any other vertex candidate: these tracks are lost. Beyond the z position of the tracks, no information about the compatibility of the tracks to their candidate vertices is used to improve the assignment of tracks to vertices. However, in particular at the beginning of data taking, when the error estimate available on the track parameters might be very poor, using such a clustering method may a good choice⁽⁸⁾.

⁽⁷⁾The $\text{Prob}(\chi^2)$ is defined as $\left(1 - \int_0^{\chi^2} f(\tilde{\chi}^2) d\tilde{\chi}^2\right)$, where $f(\tilde{\chi}^2, n_{\text{dof}})$ is the probability density function of the χ^2 distribution with appropriate number of degrees of freedom dof, which in the present case is two.

⁽⁸⁾after proper tuning the clustering size and the χ^2_{cut} parameters on data.

5.4.2 Seed finders

The seed finder is responsible for finding a first estimate of the vertex position, given a set of tracks. While this is desirable, but not necessary for a conventional vertex finder, it is fundamental when an *adaptive* vertex finding algorithm is used. Two algorithms have been implemented: a full *3d* and a simplified longitudinal cluster finding algorithm. While for primary vertex finding applications, due to the size of the beam spot in the transverse plane, clustering along the z axis yields a good first estimate of the vertex positions, the full *3d* seeding is potentially an important alternative strategy, which might be needed if the beam spot size is significantly larger or if its position is not sufficiently well known.

3d Seed Finder

A *3d seed finder* is designed to be able to find a point in the three-dimensional space where the probability of the tracks to intersect is highest. This is a non-trivial task, since the distance between tracks does not obey the triangular dis-equality: for example, if the distance between two pair of tracks $\text{dist}(a, b) < \epsilon_1$ and $\text{dist}(b, c) < \epsilon_2$, it is not possible to define any upper limit for $\text{dist}(a, c)$. This makes it particular difficult to find the region of space where the tracks are more likely to intersect.

The following procedure is adopted, which is very similar to what was implemented for the first time in CMS [67] and which reduces the spaces of tracks to a simpler euclidean space of points:

- all distances between all combinations of two tracks out of N are calculated;
- the middle points at half distance between the tracks are considered;
- a weight w is assigned to each of these points, according to their distance d ($w = \frac{1}{d^2}$).

To obtain the minimum distance between two tracks, the following distance needs to be minimised as a function of the position ϕ_1 and ϕ_2 along the first and second track:

$$d = \text{Inf}_{\phi_1, \phi_2} |\vec{r}_1(\phi_1) - \vec{r}_2(\phi_2)|. \quad (5.47)$$

Since no analytical minimisation is possible, a simple Newton iterative method in two dimensions has been used, re-deriving it for the ATLAS track parametrisation based on Eq. 5.32 and assuming the magnetic field to be directed along the z axis and constant in the vicinity of the interaction region:

$$\begin{pmatrix} \phi_{1,n+1} \\ \phi_{2,n+1} \end{pmatrix} = \begin{pmatrix} \phi_{1,n} \\ \phi_{2,n} \end{pmatrix} - \mathcal{J}^{-1}(\phi_1, \phi_2) \begin{pmatrix} \frac{\partial d(\phi_1, \phi_2)}{\partial(\phi_1)} \\ \frac{\partial d(\phi_1, \phi_2)}{\partial(\phi_2)} \end{pmatrix} \quad (5.48)$$

where

$$\mathcal{J} = \begin{pmatrix} \frac{\partial^2 d(\phi_1, \phi_2)}{\partial^2 \phi_1} & \frac{\partial^2 d(\phi_1, \phi_2)}{\partial \phi_1 \partial \phi_2} \\ \frac{\partial^2 d(\phi_1, \phi_2)}{\partial \phi_1 \partial \phi_2} & \frac{\partial^2 d(\phi_1, \phi_2)}{\partial^2 \phi_2} \end{pmatrix}, \quad (5.49)$$

where ϕ_1 and ϕ_2 are typically defined as the polar angles of the track momenta in the transverse plane in correspondence of the two defined positions along the two tracks.

The three-dimensional distribution of points at half-distance between all considered tracks is then split into three one-dimensional distributions. These distributions are supposed to

peak at the positions where most of the tracks intersect each other, which will yield a first estimate of the vertex position: in order to find the *mode* of these distributions in a scale invariant way, i.e. without depending on any specific choice for the binning, a *Half Sample Mode* (HSM) algorithm is used, as illustrated in Fig. 5.4:

- All intervals which cover at least 50% of the points are considered.
- The interval with smallest length is chosen.
- The procedure is started again from 5.4.2. with the remaining points, until less than 3 points are left.
- The mean of the remaining 2 or 3 points is returned as estimation of the peak.

The weights, defined as the inverse squared distance between the two considered tracks, can also be taken into account multiplying the interval lengths by the inverse of the sum of all the weights of the points contained in that interval: in this way the algorithm complexity grows from $O(N)$ to $O(N^2)$.

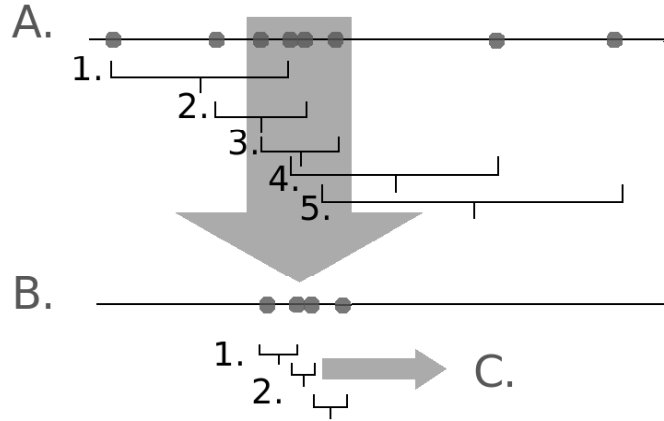


Figure 5.4: Schematic view of how the *Half Sample Mode* algorithm works.

Even if the final algorithm complexity of the seed finder, considering that the number of all possible pair of tracks is $\frac{N(N-1)}{2}$, is $O(N^2)$ for the minimum finding algorithm and $O(N^3)$ for the following mode finder algorithm, the whole procedure is still significantly faster than the vertex fitting step, so that no particular effort was dedicated to improve the speed of the seed finding algorithm.

***z*-Scan Seed Finder**

The *z*-scan seed finder returns as a seed for the primary vertex the *x* and *y* coordinate of the position of the beam spot centre, while it finds an estimate for the *z* coordinate applying the HSM algorithm to the *z* position of tracks at the *PCA* (point of closest approach) to the beam spot centre.

In order to avoid including secondary tracks during the seeding procedure the contribution of the tracks to the *z* distribution can be weighted by their respective compatibility to the

beam spot:

$$w_i(z_0, d_0, \phi) = \exp \left[- \left(\frac{\vec{r}_i(z_0, d_0, \phi) - \vec{r}_{BeamSpot}}{\sigma(\vec{r}_i)} \right)^2 \right]. \quad (5.50)$$

Performance of the seed finders

The performance of the 3d seed finder is analysed by looking at the resolution achievable by such an algorithm, both in the longitudinal and transverse direction, as shown in Fig. 5.5. The longitudinal resolution provided by the seed finder is around a factor of 2 worse than after a full vertex fit, while the transverse resolution is worse by a factor of around 2.5. This can be considered as a very reasonable performance, since it must be considered that the seed finder does not try to constrain the tracks to pass through a single position in space, as the vertex fit does, and does not explicitly rely on the track parameters' errors.

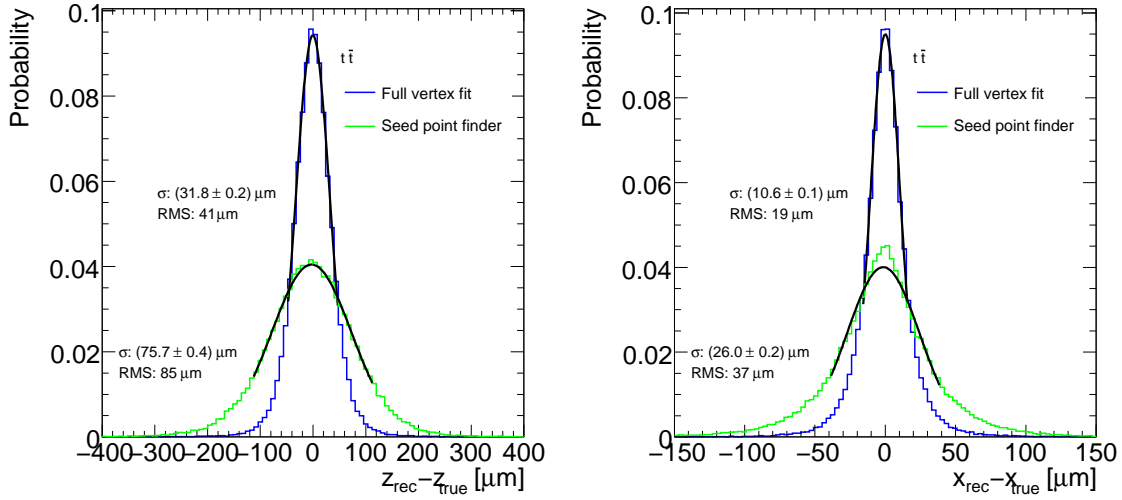


Figure 5.5: Comparison of resolution achieved during initial estimate of vertex position by the 3d seed finding algorithm and after the final vertex fit result, both in the z (left column) and x (right column) directions, based on simulated events based of the $t\bar{t}$ process.

5.4.3 Adaptive vertex finder

The adaptive vertex finder (known in ATLAS under the name of *InDetAdaptivePriVxFinder-Tool*) differs from the default vertex finder. It implements a *finding through fitting* approach, which is based on the following method:

1. A seed finder is used to find a first seed for a vertex candidate.
2. A vertex fitter (typically the Adaptive Vertex Fitter described in Section 5.3.5) is used to fit the new vertex candidate.
3. All incompatible tracks from the previous fits are used to find a new seed and start again from 1, until no tracks are left.
4. The output is a complete set of fitted vertices.

The compatibility of tracks to a certain vertex is evaluated by considering the same $\hat{\chi}^2$ used in the weight factors of Eq. 5.45. As a default $\text{Prob}(\hat{\chi}^2) < 0.1\%$ is used, which corresponds to slightly more than 4 standard deviations, to reduce the amount of vertices which split into two due to the presence of tracks whose error is underestimated.

5.4.4 Adaptive Multi-Vertex Finder

The adaptive multi-vertex finder (known in ATLAS under the name of *InDetAdaptiveMulti-PriVxFinderTool*) is an extension of the adaptive vertex finder, which deals with the simultaneous fit of all primary vertices in the event (from signal and pile-up) and is therefore intended to be used with the adaptive multi-vertex fitter (AMVF) introduced in Section 5.3.5. A novel strategy is proposed to solve the pattern recognition problem, i.e. to find a first estimate of the number of initial primary vertices and of their position, perform the adaptive multi-vertex fit and execute a splitting and/or merging procedure based on the multi-vertex fit result, based on the following steps:

1. A new vertex is seeded (using a certain seed finder).
2. The new vertex is fitted simultaneously with all previous vertices, in case they already exist, using the adaptive multi-vertex fitter.
3. Tracks compatible with the new vertex ($\hat{\chi}^2 < \chi_{cut}^2$) are removed from the list of tracks used for seeding.
4. A new iteration is performed starting again from 1, until no more vertices are found.
5. Vertices which are not significantly separated are merged and a new fit iteration is performed.

While this procedure is relatively simple, it has several drawbacks: the main problem is that it is extremely slow with increasing instantaneous luminosities (and thus increasing number of pile-up vertices), since at every new iteration the new vertex is fitted simultaneously with all previous ones while all their tracks compete to be attached to any of the considered vertices. While this permits to obtain the optimal separation between close vertices, a feature which is particularly useful with high instantaneous luminosities, it is not particularly useful in the case of well separated vertices, where it only slows down the whole process.

This problem is cured by modifying the two-dimensional matrix of weights $\omega_{i,j}$ (considering all weights introduced in Eq. 5.46, where each element represents a measure for the probability for track i to belong to vertex j), in such a way that tracks compete for certain vertices only if they have a reasonable minimum probability to contribute to their fit (tracks have to be within a Δz window of typically 8 mm to a vertex). Then the adaptive multi-vertex fit is called: if some of these tracks were already sufficiently compatible with any other previously fitted vertex, these vertices are added to the multi-vertex fit with all their remaining associated tracks. All vertices sufficiently far away from the new vertex of interest are not touched by this procedure, making the multi-vertex fit significantly faster. After the new vertex is found, the procedure is iterated again, until no new vertices are left.

In addition, some special situations which can arise during a multi-vertex fit need to be addressed with some care. One of these situations is given by cases where a primary vertex is split into two vertices. This can happen for example either because of badly measured tracks or because of secondaries very close to the primary interaction region. When this occurs, the fit of the first vertex will reject some outliers, which are incompatible with the fitted vertex according to their $\text{Prob}(\hat{\chi}^2)$. According to the multi-vertex finding procedure, these will be

used to try to seed a new vertex, and then a simultaneous multi-vertex fit will be performed both on the old and on the new vertex. There are two possible outcomes of the vertex fit in this case: if everything goes well, a new vertex will be found, eventually with few tracks in common with the first vertex but with a well separate core of tracks. However, in other cases the second vertex position will converge to the first vertex position, they will both acquire the same common tracks with weight close to $\approx \frac{1}{2}$ and the tracks which were used to seed the second vertex will become outlying measurements for both vertices. In order to avoid this to happen the following procedure was implemented:

- After a new vertex i is fit, a test is performed, in order to make sure that the new vertex is not closer to any other vertex j in space by less than 2σ , applying a cut on the vertex compatibility formula of Eq. 5.51 in case the 3d seeding procedure is used or on the formula of Eq. 5.52 in the case only the seeding procedure along the z direction is used.

$$\text{comp}(\text{vtx}_1, \text{vtx}_2)_{3d} = \exp \left[-\frac{(\vec{r}_i - \vec{r}_j)^T [\text{COV}(\vec{r}_i) + \text{COV}(\vec{r}_j)]^{-1} (\vec{r}_i - \vec{r}_j)}{2} \right] \quad (5.51)$$

$$\text{comp}(\text{vtx}_1, \text{vtx}_2)_z = \exp \left[-\frac{(z_i - z_j)^2}{2(\sigma(z_i)^2 + \sigma(z_j)^2)} \right] \quad (5.52)$$

- If the new vertex fails to fulfil this condition, i.e. if the two vertices have been merged into a single one, the second one is removed and a new multi-vertex fit is performed on vertex j , simultaneously with all other vertices which share tracks with it.

5.4.5 Signal selection in the presence of pile-up

After the primary vertex finder has been run successfully, a set of well defined primary vertices (position, vertex quality and corresponding tracks) is returned. It is of crucial importance for later stages of the ATLAS event reconstruction software to isolate among these vertices the one with the highest probability to contain the hard scatter event of interest. Many object identification algorithms require their candidates to point to the identified signal primary vertex; this is the case for example when impact parameter cuts are applied to tracks used as candidates for b -jets or τ -jets.

In order to isolate the signal vertex among all found primary vertices, some vertex properties of the charged track particles associated to the fitted vertices are used, in particular the following quantities have been considered for every single vertex:

- the transverse momentum of the associated tracks;
- the number of associated tracks.

In the case an *adaptive* primary vertex fitting algorithm is used, only tracks with a weight (expressing a measure for the probability to be associated to a certain vertex) $w > w_{cut}$ are considered for the vertex selection procedure, where w_{cut} is usually chosen to be 10^{-3} . Based on these input quantities, different methods have been implemented:

- Discriminator based on $\sum_{i=1}^{N_{trk}} p_{T,i}^2$.
- PROB method: discriminator expressing the probability, based on the p_T distribution for the tracks, for a vertex to originate from a minimum-bias event, neglecting correlations between different tracks associated to the same vertex.
- Neural network discriminator based on the p_T of all tracks.

In this thesis only the first method will be considered, since it is simple, robust and it performs very well. The interested reader can find a detailed description of the remaining mentioned methods in Ref. [69].

5.5 Performance

The way the vertex finding and fitting algorithms are implemented in the ATLAS software framework ATHENA is described in detail in the appendix (Section A.2), with emphasis on the vertexing framework and on the work-flows of the various different algorithms. In this section only the performance of the various primary vertex reconstruction algorithms described in this thesis is discussed.

In the absence of recorded collision events, the ATLAS offline reconstruction chain and different primary vertex finding algorithms are run on various kinds of simulated physics events, generated by using a Monte Carlo event generator (PYTHIA or MC@NLO) and passed through the full simulation of the ATLAS detector. The samples used for the study are listed in Table 5.1. They include samples with no contamination from pile-up events, which are useful to assess the nominal algorithmic performance, and with the level of contamination expected during very early ($\mathcal{L} \approx 10^{31} \text{ cm}^{-2}\text{s}^{-1}$) and nominal low luminosity ($\mathcal{L} \approx 10^{33} \text{ cm}^{-2}\text{s}^{-1}$) data taking periods. Both *in-time* and *out-of-time* pile-up contributions are considered. It is interesting to note that, as explained in Section 3.1, due to the higher bunch crossing spacing expected in the very early data taking runs, even at very low instantaneous luminosities ATLAS may have to deal with a significant number of overlaid interactions from pile-up events. It is therefore very important to make sure the algorithms can cope efficiently with such a scenario.

Process	Generator	Pile-up		
		\mathcal{L}	Δt_{BC}	$\langle N_{pu} \rangle$
$WH(120) \rightarrow \ell\nu b\bar{b}$	PYTHIA	-	-	0
$WH(120) \rightarrow \ell\nu u\bar{u}$	PYTHIA	-	-	0
$t\bar{t} (\rightarrow b\ell\nu bqq \text{ or } \rightarrow b\ell\nu b\ell\nu)$	MC@NLO	-	-	0
$H(120) \rightarrow \gamma\gamma$	PYTHIA	-	-	0
$WH(120) \rightarrow \ell\nu b\bar{b}$ with $p_T(W,H) > 150\text{GeV}$	HERWIG	-	-	0
minimum bias	PYTHIA	-	-	-
$WH(120) \rightarrow \ell\nu b\bar{b}$	PYTHIA	$10^{33} \text{ cm}^{-2}\text{s}^{-1}$	75ns	6.9
$WH(120) \rightarrow \ell\nu u\bar{u}$	PYTHIA	$10^{33} \text{ cm}^{-2}\text{s}^{-1}$	75ns	6.9
$t\bar{t} (\rightarrow b\ell\nu bqq \text{ or } \rightarrow b\ell\nu b\ell\nu)$	MC@NLO	$10^{33} \text{ cm}^{-2}\text{s}^{-1}$	75ns	6.9
$H(120) \rightarrow \gamma\gamma$	PYTHIA	$2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$	25ns	4.6
$WH(120) \rightarrow \ell\nu b\bar{b}$	PYTHIA	$9.7 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$	450ns	4.0
$WH(120) \rightarrow \ell\nu u\bar{u}$	PYTHIA	$9.7 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$	450ns	4.0
$t\bar{t} (\rightarrow b\ell\nu bqq \text{ or } \rightarrow b\ell\nu b\ell\nu)$	MC@NLO	$9.7 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$	450ns	4.0
$H(120) \rightarrow \gamma\gamma$	PYTHIA	$9.7 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$	450ns	4.0

Table 5.1: Monte Carlo samples used to assess the performance of primary vertex finding in ATLAS. Samples including different levels of pile-up contamination are considered. The information about pile-up is condensed in the instantaneous luminosity (\mathcal{L}), the interval in time between two subsequent bunch crossings (Δt_{BC}) and the average number of additional in-time pile-up events ($\langle N_{pu} \rangle$). The label $H(120)$ indicates a Higgs boson with a mass of 120 GeV.

The combination of vertex finding and vertex fitting algorithms considered in the present study are listed in Table 5.2. There is one more vertex finding algorithm available in ATLAS, based on an alternative implementation of the Billoir method (the VKalVrt vertexing package [70]): this is however not considered in the course of the present study.

Name	Vertex finder	Vertex fitter	Deals with pile-up
Default vertex finder	Default vertex finder	Billoir vertex fitter	yes
Sequential vertex finder	Default vertex finder	Kalman vertex fitter	yes
Adaptive vertex finder	Adaptive vertex finder	Adaptive vertex fitter	no
Adaptive multi-vertex finder	Adaptive multi-vertex finder	Adaptive multi-vertex fitter	yes

Table 5.2: Combinations of primary vertex finders and vertex fitters used in this study.

5.5.1 Track selection

The tracks used as input by the vertex finding algorithms are first processed by a track selection tool, which filters bad quality or secondary tracks. The track selection, due to the modularity of the vertex reconstruction framework, is common to all vertex finder tools considered in this study: this makes it easier to compare the pure algorithmic performance of the various algorithms. The cuts adopted during the track selection stage are listed in Table 5.3. The quality of the tracks is mainly ensured by requiring a minimum amount of hits to be found in the various layers of the inner detector and the track fit to be of a certain quality. Secondary tracks are mainly rejected by requiring a sufficiently small transverse impact parameter, but this is only possible if the beam spot position is known with sufficient good precision. If this is not the case, a looser track selection needs to be applied.

Quantity	Cut
Transverse momentum	$p_T > 500 \text{ MeV}$
Transverse impact parameter	$ d_0 < 4 \text{ mm}$
Error on transverse impact parameter	$\sigma(d_0) < 350 \text{ } \mu\text{m}$
Error on longitudinal impact parameter	$\sigma(z_0) < 2.5 \text{ mm}$
Track fit quality	$\chi^2/\text{dof} < 3.5$
Minimum pixel detector hits	$n_{pix} \geq 1$
Minimum SCT detector hits	$n_{SCT} \geq 5$
Minimum pixel+SCT detector hits	$m_{pix} + n_{SCT} \geq 7$
<i>if beam spot position is known:</i>	
Transverse impact parameter	$ d_0 < 1 \text{ mm}$
Transverse impact parameter significance	$\frac{d_0}{\sigma(d_0)} < 4$

Table 5.3: Selection cuts applied on tracks to be used as input for primary vertex finding. The transverse impact parameter is computed with respect to the beam spot position.

The effect of the track selection on the vertex reconstruction track multiplicity is shown in Fig. 5.6, where the number of tracks before and after track selection is plotted for all considered physics processes, in the absence of contamination from pile-up events. For completeness also the effective number of tracks used in the fit by the adaptive multi-vertex finder is shown (corresponding to the sum of the track weights ω after the last fit iteration): this gives a measure of the number of residual outlying measurements after track selection is applied. In general, a bit less than half of the original tracks are selected to be used in the primary vertex fit: the highest track multiplicity is seen in $t\bar{t}$ events, while the lowest one is seen in minimum bias events. In the case of the WH process, it is interesting to notice the difference between $H \rightarrow b\bar{b}$ events and $H \rightarrow u\bar{u}$ events. At the beginning the track multiplicity is higher in the first process, due to the higher number of tracks in a b -quark jet with respect to a light quark jet. After selection and, even more, after the intrinsic outlier removal performed by the

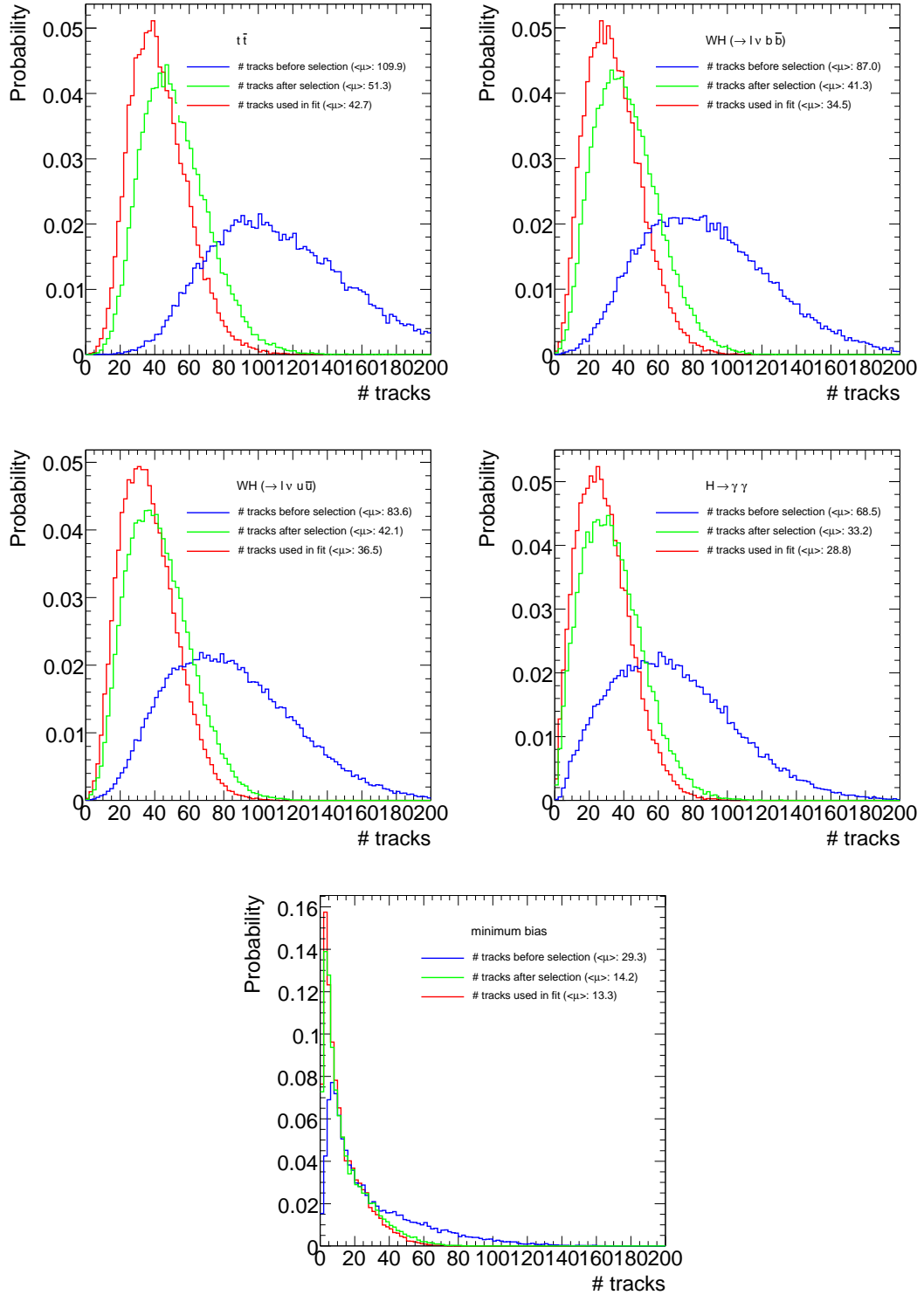


Figure 5.6: Number of tracks in the event before and after track selection, for all considered physics processes and in the absence of contamination from pile-up events. As a reference, also the effective number of tracks used in the adaptive multi-vertex fit is shown.

adaptive vertex fit, the amount of tracks surviving the selection is higher in the WH process with $H \rightarrow u\bar{u}$, due to the fact that many of the tracks in a b -quark jet stem from b -hadron decays and are therefore significantly displaced with respect to the primary vertex.

The final transverse momentum and pseudo-rapidity distributions for the tracks which feed into the vertex fit, together with the final track multiplicity at the vertex, are shown in Fig. 5.7. The resolution with which the vertex can be determined in the fit depends critically on these three variables. The track resolution depends strongly on the kinematic variables p_T and η (as shown in Fig. 4.2 of Chapter 4), while, under the hypothesis of a fixed distribution in p_T and η , the vertex uncertainty σ decreases approximately linearly with the inverse of the square root of the number of tracks⁹. While most of the tracks corresponding to charged particles produced in the hadronisation of quarks have relatively low transverse momenta, a clear tail to higher values can be seen in the p_T distributions for the $t\bar{t}$ and WH processes, due to isolated leptons from the W boson decays in these two processes.

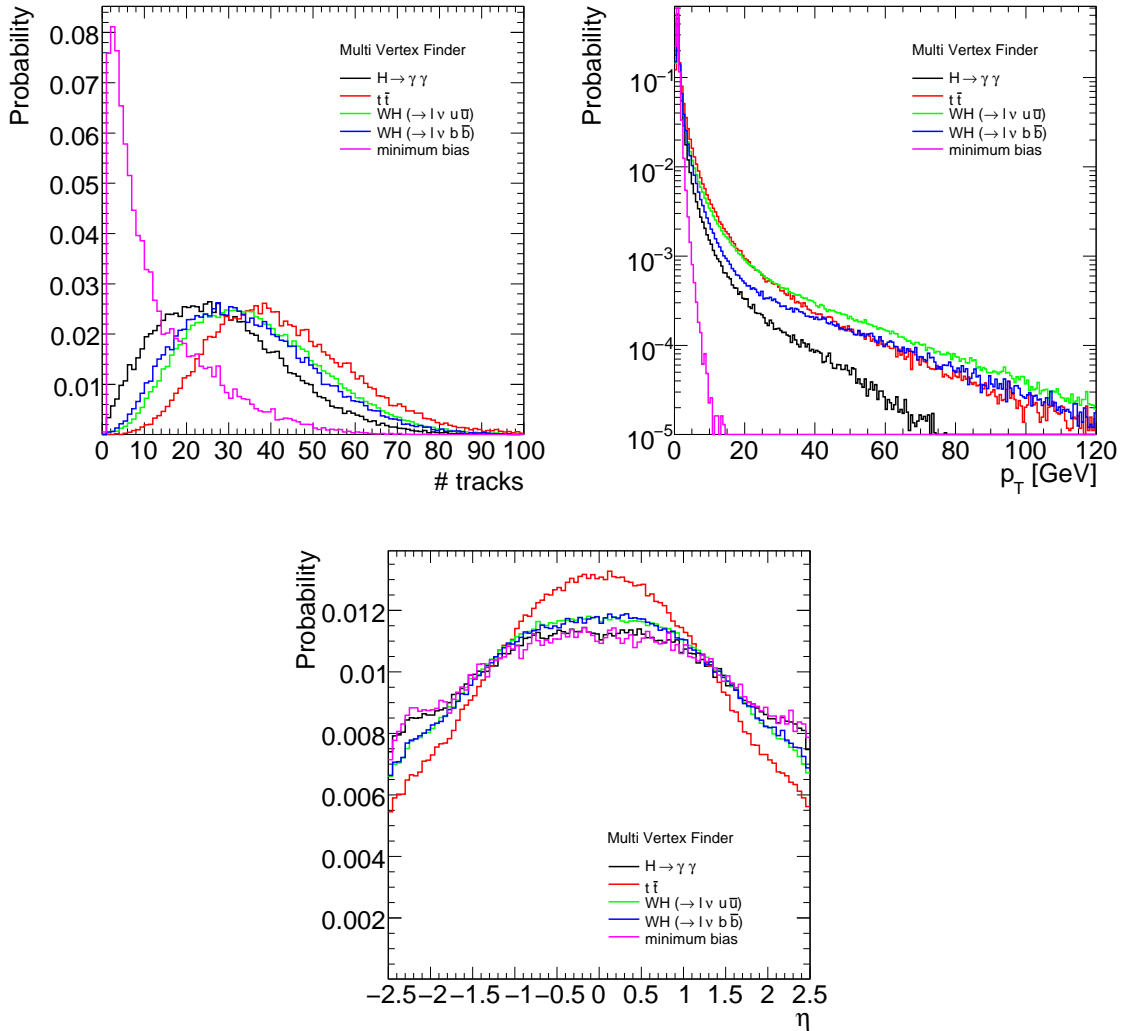


Figure 5.7: Track multiplicity, transverse momentum and pseudo-rapidity distribution of the tracks used in the vertex fit, after track selection and after the outlier removal of the adaptive multi-vertex finder.

⁹(if the constraint provided by the beam spot in the transverse plane is not considered)

5.5.2 Vertex reconstruction performance without pile-up

The vertex resolution depends not only on the precision achievable by constraining all tracks to intersect a single point in space, but also on the size of the beam spot, which is used as a constraint in the vertex fit. The resolution predicted by Monte Carlo simulations when running primary vertex finding on various physics processes is shown in Fig. 5.8, in the form of residuals of the fitted vertex position with respect to the simulated primary vertex position, separately on the x and z axes.

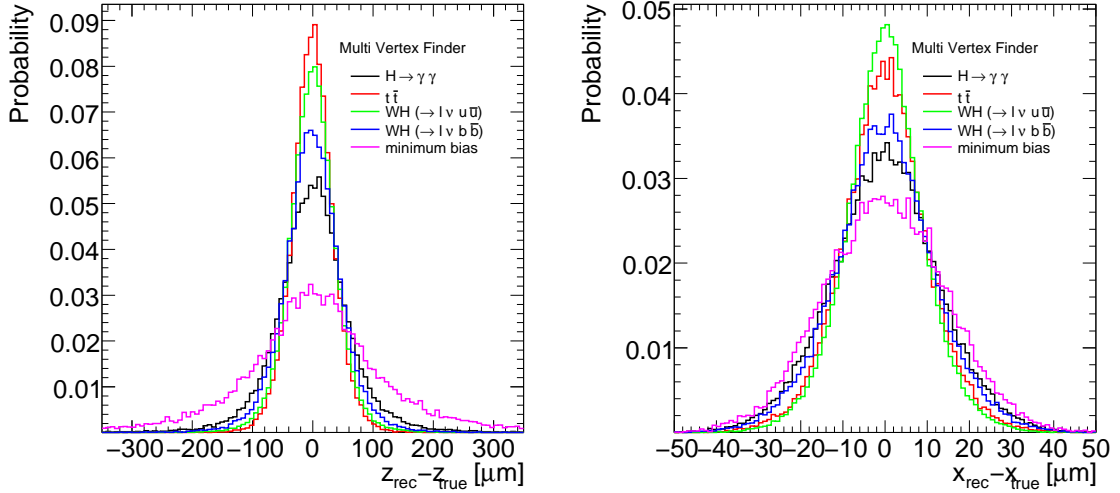


Figure 5.8: Residuals of the reconstructed vertex position with respect to the simulated vertex position along the z (left) and y (right) axes, obtained by applying the adaptive multi-vertex finding algorithm on various different physics events.

The performance of the adaptive vertex finder is compared to the other vertex finders in Table 5.4, for all considered physics samples. The core resolution in the x and z directions is defined as the standard deviation σ of a Gaussian fit on the respective residual distribution, where the fit is performed iteratively in a $\pm 1.5\sigma$ window until a stable result is obtained. For completeness the squared mean residual (RMS) of the whole distribution is also quoted, which gives an idea of the extension of the non-Gaussian tails. The percentage of outliers is defined as the number of events where the reconstructed vertex was either not found or was outside a window of 0.5 mm in the z direction with respect to the simulated vertex position.

As can be seen both in the plots and in the table, the best resolution can be achieved on $t\bar{t}$ events, where, using the adaptive finders, a core resolution of $\sigma_z \approx 30 \mu\text{m}$ and $\sigma_{x,y} \approx 9 \mu\text{m}$ can be obtained along the longitudinal and transverse directions, respectively, while the worse resolution is found on low p_T minimum bias events, with a core resolution of $\sigma_z \approx 80 \mu\text{m}$ and $\sigma_{x,y} \approx 15 \mu\text{m}$. In the case of minimum bias events the constraint provided by the tracks in the transverse plane is nearly unable to improve the knowledge available due to the beam spot constraint, which in the simulation is assumed to be $\sigma_{x,y} = 15 \mu\text{m}$.

The performance of the adaptive multi-vertex finder and of the adaptive vertex finder is essentially identical for all considered samples. This is not surprising: infact, in the case of no pile-up, only one vertex is present in the event and, thus, the multi-vertex fit never deals with the simultaneous fit of more vertices and performs the very same operations of the (single-vertex) adaptive vertex finder. The same consideration holds for the sequential and the default vertex finders, which also show nearly the same performance. As already shown

$H \rightarrow \gamma\gamma$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.4%	47.3 μm	11.8 μm	67.6 μm	12.9 μm
Adaptive Vertex Finder	0.4%	47.0 μm	11.7 μm	67.3 μm	12.9 μm
Sequential Vertex Finder	1.1%	52.0 μm	12.0 μm	83.4 μm	13.0 μm
Default vertex finder	1.1%	51.5 μm	12.0 μm	82.2 μm	13.0 μm
$t\bar{t}$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.0%	30.0 μm	8.9 μm	37.9 μm	10.4 μm
Adaptive Vertex Finder	0.0%	30.0 μm	8.9 μm	35.8 μm	10.3 μm
Sequential Vertex Finder	0.4%	34.6 μm	9.0 μm	55.3 μm	10.5 μm
Default vertex finder	0.4%	34.5 μm	9.0 μm	54.4 μm	10.5 μm
$WH(\rightarrow l\nu u\bar{u})$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.0%	33.8 μm	8.0 μm	41.8 μm	9.5 μm
Adaptive Vertex Finder	0.0%	33.6 μm	8.0 μm	41.5 μm	9.5 μm
Sequential Vertex Finder	0.6%	36.6 μm	8.4 μm	57.1 μm	10.2 μm
Default vertex finder	0.6%	36.4 μm	8.3 μm	56.4 μm	10.0 μm
$WH(\rightarrow l\nu b\bar{b})$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.1%	39.8 μm	10.4 μm	51.6 μm	12.0 μm
Adaptive Vertex Finder	0.1%	39.8 μm	10.4 μm	51.6 μm	12.0 μm
Sequential Vertex Finder	0.8%	43.8 μm	10.9 μm	70.8 μm	12.2 μm
Default vertex finder	0.8%	43.2 μm	10.8 μm	69.4 μm	12.2 μm
minimum bias					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	9.2%	80.8 μm	14.6 μm	128.8 μm	14.4 μm
Adaptive Vertex Finder	8.5%	81.0 μm	14.6 μm	130.4 μm	14.4 μm
Sequential Vertex Finder	6.8%	88.7 μm	14.7 μm	143.3 μm	14.5 μm
Default vertex finder	6.6%	88.3 μm	14.7 μm	142.2 μm	14.5 μm
$WH \rightarrow b\bar{b}$ at high p_T					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.0%	33.5 μm	9.9 μm	40.9 μm	11.1 μm
Adaptive Vertex Finder	0.0%	33.5 μm	10.0 μm	40.9 μm	11.1 μm
Sequential Vertex Finder	0.4%	39.5 μm	9.8 μm	64.6 μm	11.4 μm
Default vertex finder	0.4%	38.0 μm	9.6 μm	64.4 μm	11.4 μm

Table 5.4: Performance achievable by the different vertex finding algorithms on various physics events samples, without pile-up: the percentage of outliers, the core resolution (σ) and average squared mean (RMS) of the residuals distributions are listed separately for the x and z directions. The residuals in the y direction are not quoted, since they are identical to the corresponding ones in the x direction. The statistical error on all resolutions quoted is 0.1 or smaller.

in Table 5.2, these two finders use the same finding algorithm, but use two different vertex fitting approaches: the sequential vertex finder relies on the Kalman filter approach, while the default vertex finder relies on the Billoir fit approach. As demonstrated at the end of Section 5.3.2, these two approaches are mathematically identical and thus provide very close results.

While the differences in resolution in the transverse plane are diluted by the strong constraint provided by the beam spot, the longitudinal resolution undergoes a clear improvement when moving from more conventional algorithms to adaptive methods. As shown in Fig. 5.8 and in Table 5.4, the longitudinal core resolution achievable on $t\bar{t}$ events improves by 13% from $34.6 \mu m$ to $30.0 \mu m$, but the most visible effect is in the tail of the z vertex position residual distribution, where the mean squared residual improves by 34% from $54 \mu m$ to $36 \mu m$. This difference in performance corresponds to the two different strategies implemented to reduce the influence of outlying measurements in the fit: the default and sequential vertex finders try to accomplish this scope by iteratively removing incompatible tracks from the vertex fit, while the adaptive fitters smoothly down-weight outlying tracks during the fit itself, following the thermodynamic annealing procedure described in Section 5.3.5.

The distribution for the estimated errors on the longitudinal and transverse vertex position are shown in Fig. 5.9 and their average is listed in Table 5.5. These follow closely the effective resolutions achievable on the various different samples. It is worth noting the effect of the beam spot constraint in the transverse plane; the expected uncertainty in the transverse position of the primary vertex is approximately:

$$\sigma_{x,y} \approx \frac{1}{\sqrt{\frac{1}{\sigma(\text{BS})_{x,y}^2} + \frac{1}{\sigma(\text{PV})_{x,y}^2}}},$$

where BS refers to the beam spot and PV to the primary vertex. If, as in the case of minimum bias events, $\sigma(\text{PV}) \gg \sigma(\text{BS})$, the overall $\sigma_{x,y}$ will approach $\sigma(\text{BS}) \approx 15 \mu m$, as the plot can confirm. Even in the case of $t\bar{t}$ events, the constraint provided by the beam spot constraint still improves the knowledge of the vertex position: using the above formula it is easy to estimate that the $\sigma_{x,y}$ for these physics events would blow up from in average $8.7 \mu m$ to $\approx 10.7 \mu m$ if no beam spot constraint is applied, as will be seen in Section 5.5.3, where the beam spot constraint will be explicitly removed.

Whether the various vertex finders determine correctly the vertex position uncertainty can be checked by looking at the pull distributions, defined as the vertex position residuals divided by the corresponding estimated uncertainty, which are shown in Fig. 5.9 and listed in Table 5.5. The impact of outlying measurement is reduced for all vertex finders and fitters at least at a level that all pull distributions are fairly well represented by a simple Gaussian distribution centred around 0, with a well defined width σ . If the error is estimated correctly on an event per event basis, this width has to be very close to one. In the case of the $H \rightarrow \gamma\gamma$, $WH(l\nu \rightarrow u\bar{u})$ and minimum bias samples no tracks from b -decays are expected in the vicinity of the primary vertex and the impact of outlying measurements is reduced: here both adaptive and more conventional finders perform fairly well, with widths very close to one. On the remaining samples, $t\bar{t}$ and $WH(l\nu \rightarrow b\bar{b})$, the width is a bit above one, which means the errors, in particular in the z direction, are underestimated due to the presence of outliers, with the adaptive finders being more robust and underestimating them by up to 15% and the remaining finders underestimating them by up to 35%.

Before introducing the algorithmic performance in the presence of contamination from pile-up interactions, it is useful to look at the number of reconstructed vertices in the absence of it, where only one reconstructed vertex is expected. What can happen in some cases is, that

$H \rightarrow \gamma\gamma$				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	60.5 μm	12.8 μm	1.03	1.01
Adaptive Vertex Finder	60.6 μm	12.8 μm	1.03	1.01
Sequential Vertex Finder	68.3 μm	12.6 μm	1.15	1.03
Default vertex finder	67.0 μm	12.5 μm	1.15	1.04
$t\bar{t}$				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	30.3 μm	8.7 μm	1.14	1.17
Adaptive Vertex Finder	30.3 μm	8.7 μm	1.14	1.17
Sequential Vertex Finder	35.4 μm	9.2 μm	1.29	1.13
Default vertex finder	34.6 μm	9.0 μm	1.30	1.16
$WH(\rightarrow l\nu u\bar{u})$				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	37.8 μm	9.3 μm	1.05	0.99
Adaptive Vertex Finder	37.8 μm	9.3 μm	1.05	0.99
Sequential Vertex Finder	42.8 μm	9.5 μm	1.17	1.03
Default vertex finder	41.9 μm	9.3 μm	1.18	1.03
$WH(\rightarrow l\nu b\bar{b})$				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	42.9 μm	10.5 μm	1.13	1.14
Adaptive Vertex Finder	42.9 μm	10.5 μm	1.13	1.14
Sequential Vertex Finder	51.2 μm	10.9 μm	1.23	1.11
Default vertex finder	49.5 μm	10.7 μm	1.25	1.13
minimum bias				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	135.4 μm	14.8 μm	0.99	0.97
Adaptive Vertex Finder	138.4 μm	14.9 μm	1.00	0.97
Sequential Vertex Finder	154.9 μm	14.4 μm	1.06	1.01
Default vertex finder	152.0 μm	14.4 μm	1.07	1.01
$WH \rightarrow b\bar{b}$ at high p_T				
	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,\text{PULL}}$	$\sigma_{x,\text{PULL}}$
Multi-Vertex Finder	32.6 μm	10.1 μm	1.16	1.09
Adaptive Vertex Finder	32.6 μm	10.1 μm	1.16	1.09
Sequential Vertex Finder	40.6 μm	10.4 μm	1.33	1.08
Default vertex finder	39.7 μm	10.3 μm	1.35	1.10

Table 5.5: Performance achievable by the different vertex finders on various physics events samples, without pile-up: the average estimated vertex error ($\langle \sigma \rangle$) and the Gaussian width of the pull distributions (σ_{PULL}) are listed, separately for the x and z directions. The errors and pulls in the y direction are not quoted, since they are identical to the corresponding ones in the x direction.

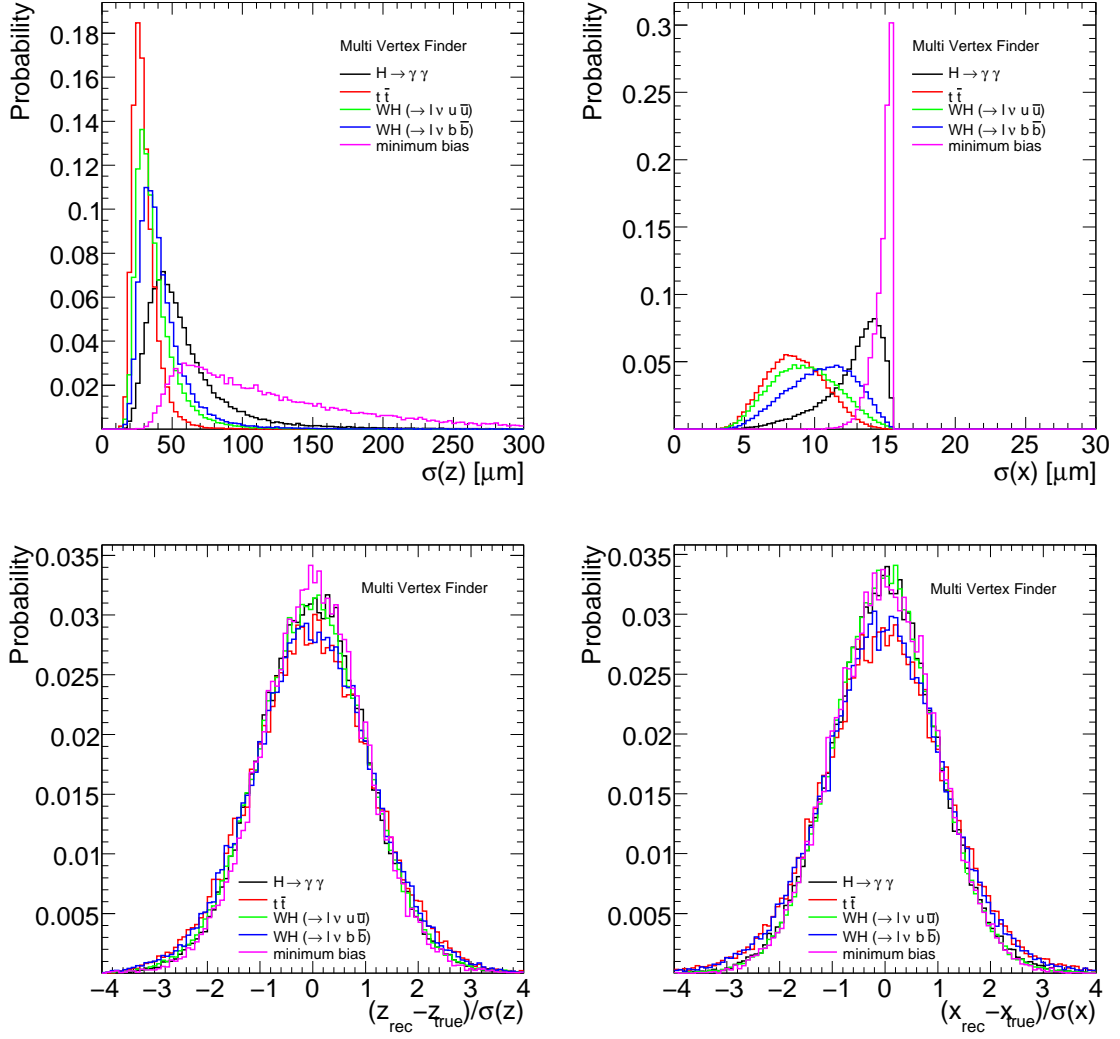


Figure 5.9: (Top row) Estimated error on the vertex position along the z (left) and x (right) axis. (Bottom row) Residuals of the reconstructed vertex position with respect to the simulated vertex position, divided by the fit error, along the z (left) and x (right) axes. All distributions are obtained by applying the adaptive multi-vertex finding algorithm on various different physics events.

either no vertex is reconstructed or, that more than one vertex is found, where the additional vertex (vertices) has (have) usually very few tracks. This last case can be considered as an undesired splitting of the signal vertex into more than one vertex. The distributions with the number of reconstructed vertices for the adaptive multi-vertex and for the default vertex finder are shown in Fig. 5.10. It can be clearly seen that the two finding approaches give quite different results here. The adaptive finder uses all incompatible tracks to the previously fitted vertices in order to seed a new vertex, which means that, in the presence of badly reconstructed tracks or of tracks coming from secondary interactions, these can result in an additional vertex; however, the figure shows that this happens only in around 1% of the cases, slightly more in the presence of outlying tracks from b -decays ($t\bar{t}$ and $WH \rightarrow b\bar{b}$). The default finding algorithm, as described in detail in Section 5.4.1, adopts a *fitting after finding* approach, so the distinction of the vertices relies exclusively on the finding algorithm, which

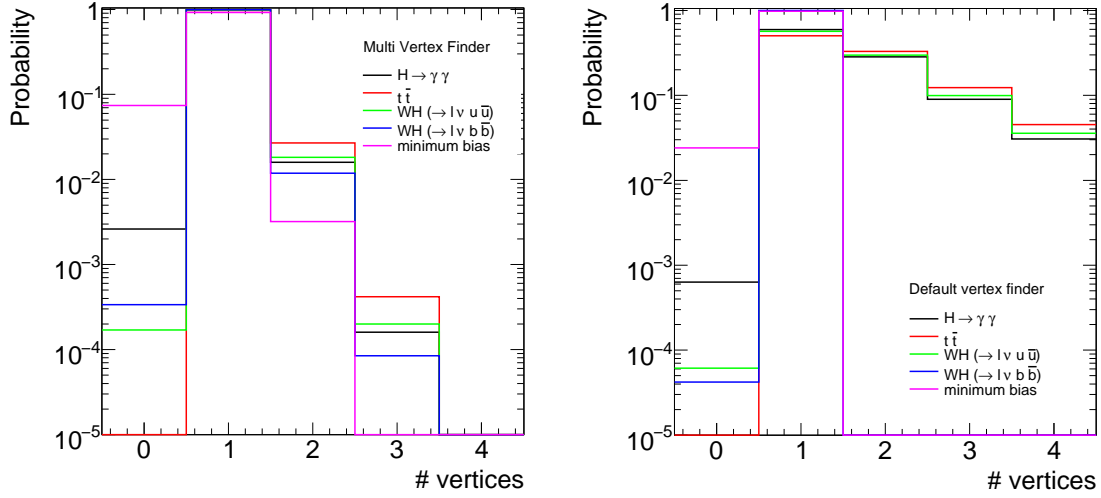


Figure 5.10: Number of reconstructed vertices per events, running the adaptive multi-vertex finder (left) or the default vertex finder (right) on various physics events, without any contamination from pile-up events.

clusters tracks into vertices using a *sliding window* algorithm with fixed size of $\Delta z = 5$ mm before the vertex fit is started. As a result, this approach reconstructs in around 40-50% of cases more than one vertex. By increasing the window size of the clustering algorithm this fraction could be reduced, but, as a drawback, it will become easier that, in the presence of pile-up, two vertices are found in a single cluster and are therefore reconstructed as a single vertex. On the other side, Fig. 5.10 shows also that in a higher fraction of cases the adaptive finder finds no vertex at all; this is due to the fact that in the default vertex finder single track vertices are allowed, whereas they are forbidden (in the default configuration) by the adaptive finder. The fraction of outliers, listed in the first column of Table 5.4, demonstrates however that many of the single-track vertices found by the default finder are indeed not the real primary vertex (they are displaced with respect to the simulated vertex by more than 0.5 mm), so that the number of outliers is still lower for the adaptive finder for all samples, except for minimum bias events. This means that the case of minimum bias events is the only one where the choice of considering single-track vertices does pay off.

The vertex reconstruction efficiency of the primary vertex on minimum bias events is around 90 - 95%, as shown in Table 5.4, depending on the vertex finding algorithm considered. Apart from using single-track vertices, this inefficiency can be further reduced by using dedicated algorithms to reconstruct very low p_T tracks (tracks with p_T below ≈ 400 MeV tend to curl within the inner detector) and then by including in the fit tracks with transverse momenta as low as 150 MeV. However, this was not considered in the course of the present thesis.

5.5.3 Performance with no beam spot constraint

During the early phases of data taking the knowledge of the beam spot size and position may be poor, so that it is important to cross-check the performance of the primary vertex reconstruction algorithms when the beam spot constraint is not applied. This does not only affect the primary vertex fitting algorithms, where the beam spot constraint needs just not to be applied, but also the seeding phase during the vertex finding stage, which needs to cope with a much higher uncertainty on the knowledge of the primary vertex position in the

transverse plane. The adaptive multi-vertex finder copes with this by switching from the z -scan based seeding to the full 3d seeding (both described in Section 5.4.2).

The results predicted by Monte Carlo simulations in this scenario are presented in Table 5.6, which can be compared directly to Table 5.4, where the beam spot constraint was applied.

The vertex reconstruction efficiency of the sequential and default vertex finder is significantly degraded. The reason is that a simple sliding window algorithm in z is used, which works well only if the z position of the tracks can be defined close enough to the real vertex position: however even a very approximated knowledge of the beam spot position would easily solve this problem.

This problem does not affect the adaptive multi-vertex finder, due to the use of the full 3d seeding. However, as a general drawback of not having a reliable estimate of the beam spot position, the primary vertex finding procedure will find all primary plus eventually several additional secondary vertices in the event, since secondary tracks will not be excluded from the fit due to the incompatibility with the beam spot constraint. This means that in the case e.g. of the $t\bar{t}$ and $WH(\rightarrow l\nu b\bar{b})$ events, additional vertices will be found which correspond to secondary vertices from b -hadron decays: in these cases, if the charged particles associated to such vertices are the highest in p_T in the event, it can happen that the signal vertex selection algorithm wrongly identifies these secondary vertices as signal vertices. Either using the track multiplicity of these vertices (which is very low for such secondary vertices) or calculating an approximate interaction region position by averaging the primary vertex positions over many events and removing iteratively outlying measurements from secondary vertices, it should be straightforward to select the real signal vertices. In addition it will be easier for badly measured tracks to form a vertex, since the space of possible vertices is not anymore restricted by the beam spot constraint. For simplicity, when more vertices, close in space, are present in the event, in this study the vertex with highest track multiplicity was selected. Thanks to the 3d seeding procedure implemented in the adaptive multi-vertex finder, the fraction of outliers (vertex candidates outside a window both in z and in $r\phi$ of 0.5 mm from the simulated vertex position) is only very slightly increased when removing the beam spot constraint, and stays in nearly all cases below 0.5%, apart for $WH(\rightarrow b\bar{b})$, where the degradation is a bit more significant. An exception is also seen in the case of minimum bias events, where actually removing the beam spot constraint reduces the vertex reconstruction inefficiency: this is most probably due to the fact that badly reconstructed tracks, which are particularly frequent in the low p_T region which characterises minimum bias events, are not anymore discarded if they are incompatible with the beam spot constraint.

In the case no beam spot constraint is applied, the resolution is only slightly degraded in the longitudinal direction, but it is significantly worse in the transverse plane, where the width of the residual distribution both in the x and y directions now accounts only for the intrinsic vertex resolution. This is particularly evident in $H \rightarrow \gamma\gamma$ and in minimum bias events, where the resulting transverse core resolution in the case of the adaptive multi-vertex finder is $\approx 20 \mu\text{m}$ and $\approx 45 \mu\text{m}$, respectively, significantly worse than the expected nominal beam spot transverse size of $\approx 15 \mu\text{m}$. The distributions of uncertainties on the vertex positions as estimated during the fit, both in the longitudinal and transverse directions, are shown in Fig. 5.11, for the adaptive multi-vertex finder and all considered physics samples.

5.5.4 Vertex reconstruction performance in the presence of pile-up

In the presence of pile-up the performance of the primary vertex reconstruction algorithms is expected to be degraded, due to mainly two reasons, as was already introduced in Section 5.2:

- The signal vertex and the additional pile-up vertices, once their typical vertex recon-

$H \rightarrow \gamma\gamma$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.5%	49.2 μm	21.8 μm	74.5 μm	22.1 μm
Sequential Vertex Finder	7.6%	62.2 μm	27.6 μm	122.8 μm	23.8 μm
Default vertex finder	7.2%	61.4 μm	27.0 μm	120.7 μm	23.6 μm
$t\bar{t}$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.2%	31.8 μm	10.6 μm	41.7 μm	15.1 μm
Sequential Vertex Finder	3.8%	46.0 μm	17.1 μm	101.5 μm	20.8 μm
Default vertex finder	3.7%	45.2 μm	16.4 μm	100.3 μm	20.5 μm
$WH(\rightarrow l\nu u\bar{u})$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.1%	33.8 μm	9.6 μm	44.7 μm	14.0 μm
Sequential Vertex Finder	4.3%	43.0 μm	12.6 μm	99.0 μm	18.1 μm
Default vertex finder	4.2%	42.3 μm	11.6 μm	97.2 μm	17.4 μm
$WH(\rightarrow l\nu b\bar{b})$					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.5%	41.6 μm	15.1 μm	58.1 μm	19.0 μm
Adaptive Vertex Finder	0.1%	40.1 μm	10.7 μm	51.7 μm	12.0 μm
Sequential Vertex Finder	5.1%	57.7 μm	23.1 μm	113.6 μm	22.7 μm
Default vertex finder	4.9%	56.5 μm	22.1 μm	111.6 μm	22.4 μm
minimum bias					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	4.8%	80.7 μm	42.4 μm	137.1 μm	26.3 μm
Sequential Vertex Finder	12.3%	100.5 μm	51.9 μm	161.5 μm	27.1 μm
Default vertex finder	10.9%	98.5 μm	49.5 μm	158.2 μm	27.0 μm
$WH \rightarrow b\bar{b}$ at high p_T					
	<i>Outliers</i>	σ_z	σ_x	RMS _z	RMS _x
Multi-Vertex Finder	0.7%	35.7 μm	14.1 μm	46.9 μm	17.1 μm
Sequential Vertex Finder	3.9%	52.5 μm	20.0 μm	100.2 μm	22.6 μm
Default vertex finder	3.7%	52.9 μm	20.8 μm	100.1 μm	21.8 μm

Table 5.6: Algorithmic performance of the various different vertex finders on different physics events samples, in the case of not applying the beam spot constraint and without contamination from pile-up events: the fraction of outliers, the core resolution (σ) and average squared mean (RMS) of the residuals distributions are listed separately for the x and z directions.

struction resolutions and the constraint provided by the beam spot are considered, are essentially overlaid on top of each other in the transverse plane and spread across the longitudinal beam spot size. The primary vertex reconstruction algorithm needs to separate them efficiently by distinguishing the single vertex positions along the z axis. Primary vertices from signal and pile-up interactions can occasionally overlap, making the distinction difficult. As a result, the signal vertex can be contaminated by tracks from pile-up vertices or, much more rarely, if the signal vertex has very few tracks with low transverse momenta, the signal vertex may be lost completely, its tracks being attached to a nearby lying pile-up vertex.

- Once the primary vertices both from signal and pile-up events have been reconstructed, the signal vertex needs to be identified among these (vertex selection). This can be again problematic for a signal vertex with very few tracks with low transverse momenta.

In the following the selection of the signal vertex candidate out of the reconstructed primary vertices is based on the $\sum_{i=1}^{N_{trk}} p_{T,i}^2$ criterion defined in Section 5.4.5.

Using the pile-up samples listed in Table 5.1, the primary vertex reconstruction performance was evaluated and the results are summarised in Table 5.7. The first one presents the

$\mathcal{L} = 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 75 \text{ ns}$, $\langle N_{pu} \rangle = 6.9$							
$t\bar{t}$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.2%	0.5%	0.3%	0.7	32.2 μm	8.8 μm	38.7 μm
Sequential Vertex Finder	0.5%	3.0%	3.1%	0.6	35.1 μm	9.5 μm	60.7 μm
Default vertex finder	0.5%	2.8%	2.8%	0.6	35.3 μm	9.4 μm	60.7 μm
$WH(\rightarrow l\nu u\bar{u})$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.5%	1.6%	1.4%	0.7	33.5 μm	8.0 μm	45.0 μm
Sequential Vertex Finder	0.9%	5.4%	5.4%	0.6	37.0 μm	8.5 μm	68.2 μm
Default vertex finder	0.9%	5.0%	4.9%	0.6	36.5 μm	8.4 μm	67.4 μm
$WH(\rightarrow l\nu b\bar{b})$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.7%	4.0%	3.8%	0.6	41.8 μm	10.4 μm	55.3 μm
Sequential Vertex Finder	1.0%	11.0%	11.2%	0.5	43.6 μm	10.7 μm	79.3 μm
Default vertex finder	1.0%	9.5%	9.7%	0.5	44.0 μm	10.5 μm	78.1 μm
$\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 25 \text{ ns}$, $\langle N_{pu} \rangle = 4.6$							
$H \rightarrow \gamma\gamma$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	1.4%	24.0%	24.0%	0.5	47.0 μm	11.6 μm	70.8 μm
Sequential Vertex Finder	1.7%	28.1%	28.4%	0.4	50.5 μm	11.9 μm	94.1 μm
Default vertex finder	1.7%	27.4%	27.7%	0.4	50.3 μm	11.9 μm	93.2 μm
$\mathcal{L} = 9.7 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 450 \text{ ns}$, $\langle N_{pu} \rangle = 4.0$							
$t\bar{t}$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.1%	0.3%	0.2%	0.3	31.0 μm	8.8 μm	37.8 μm
Sequential Vertex Finder	0.2%	1.6%	1.7%	0.2	34.1 μm	9.3 μm	54.8 μm
Default vertex finder	0.2%	1.4%	1.5%	0.2	33.9 μm	9.2 μm	54.4 μm
$WH(\rightarrow l\nu u\bar{u})$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.2%	0.9%	0.9%	0.2	33.5 μm	8.1 μm	44.2 μm
Sequential Vertex Finder	0.4%	3.0%	3.2%	0.2	36.5 μm	8.4 μm	61.5 μm
Default vertex finder	0.4%	2.7%	2.9%	0.2	36.5 μm	8.2 μm	60.5 μm
$WH(\rightarrow l\nu b\bar{b})$							
	<i>Lost</i>	<i>Misid.</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.3%	2.6%	2.6%	0.2	40.9 μm	10.5 μm	55.3 μm
Sequential Vertex Finder	0.4%	7.2%	7.7%	0.2	43.7 μm	10.5 μm	75.4 μm
Default vertex finder	0.4%	5.8%	6.4%	0.2	43.5 μm	10.5 μm	73.9 μm
$H \rightarrow \gamma\gamma$							
	<i>Misid.</i>	<i>Lost</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
Multi-Vertex Finder	0.8%	21.5%	22.1%	0.2	47.5 μm	11.4 μm	70.2 μm
Sequential Vertex Finder	0.9%	24.6%	25.3%	0.2	49.8 μm	11.7 μm	89.8 μm
Default vertex finder	0.9%	24.0%	24.7%	0.2	49.9 μm	11.7 μm	88.7 μm

Table 5.7: Performance of the primary vertex finders in the presence of different levels of pile-up contamination. Shown are the fraction of misidentified vertices (*Misid.*), the fraction of not correctly reconstructed vertices (*Lost*), the fraction of outliers (*Outliers*), the average number of tracks from pile-up ($\langle n_{trk} \rangle_{PU}$) and the core resolutions (σ) and the RMS of the vertex residual distributions along the z and x axes. These quantities are explained in the text.

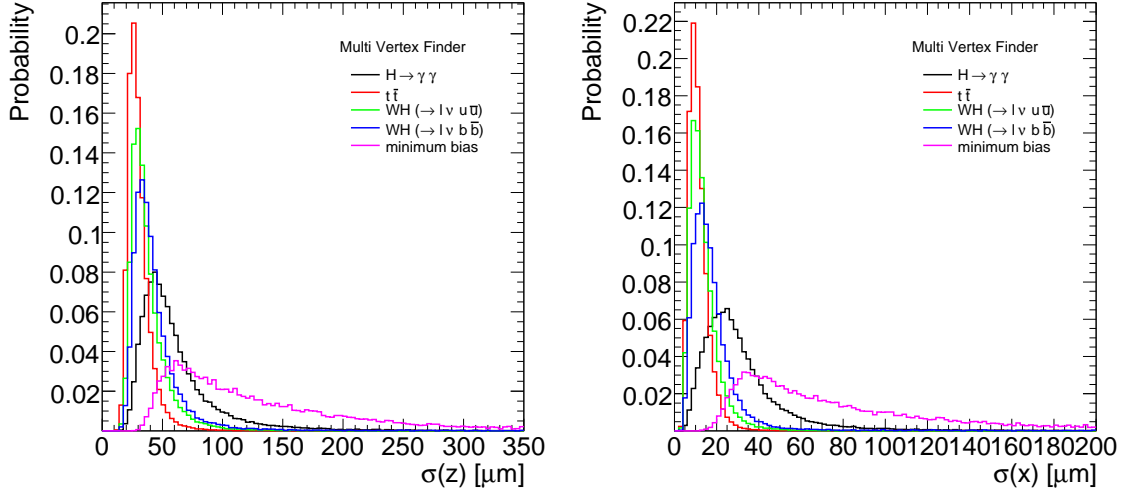


Figure 5.11: Estimated error on the vertex position along the z (left) and x (right) axis, obtained running the adaptive multi-vertex finder without beam spot constraint on various kinds of physics events.

performance expected at around the luminosity for the first physics run, after end of detector commissioning ($\mathcal{L} \approx 10^{33} \text{ cm}^{-2}\text{s}^{-1}$), while the second presents the pile-up contamination which might be expected relatively early during the commissioning of the LHC machine and of the ATLAS detector ($\mathcal{L} \approx 10^{32} \text{ cm}^{-2}\text{s}^{-1}$). However, due to the different bunch crossing spacing, all the cases considered result in a similar number of in-time additional interactions from pile-up events (4 – 6.9) and thus in very similar degradations of the primary vertex performance.

Two criteria are defined in order to evaluate if a reconstructed primary vertex stems from the signal process of interest:

- *purity based*: at least 50% of the reconstructed tracks associated with the primary vertex are required to originate, at truth level, from tracks produced in the hard scattering process;
- *geometrical*: the reconstructed vertex is required to be inside a window of 0.5 mm in z direction with respect to the simulated vertex position.

Based on the *purity based* definition, two quantities are defined to estimate the vertex reconstruction and identification performance:

- The fraction of misidentified signal vertices (*Misid.*): the fraction of events where the signal vertex was either not correctly reconstructed or not correctly selected.
- The fraction of lost vertices (*Lost*): the fraction of events where the signal vertex was not correctly reconstructed, so that it does not appear in the list of reconstructed primary vertices.

Based on the *geometrical* definition, an additional quantity is defined:

- Fraction of outliers (*Outliers*): fraction of events where the signal vertex candidate is not inside a window of 0.5 mm in the z direction with respect to the simulated vertex position.

This quantity is analogous to the fraction of misidentified vertices (*Misid.*), but is based on the *geometrical*, rather than on the *purity based*, definition for associating the reconstructed vertex to the true signal vertex¹⁰.

Finally, the average number of tracks stemming from pile-up events contaminating the signal vertex ($\langle n_{trk} \rangle$) is also considered.

Concentrating on the results for $\mathcal{L} \approx 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, it can be seen that the fraction of lost vertices is lowest on $t\bar{t}$ event (0.2-0.5%), a bit higher on WH events (0.5-1%) and highest on $H \rightarrow \gamma\gamma$ events (1.4-1.7%). Comparing Table 5.4 and Table 5.7, it can be seen that the resolutions (both core resolutions and RMS of residual distributions) are only slightly degraded by the considered level of pile-up. This is confirmed by the average number of tracks from pile-up events contaminating the signal vertex ($\langle n_{trk} \rangle_{PU}$), which is around 0.5-0.7. The performance numbers shown in Table 5.7 for the case of $\mathcal{L} \approx 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ can be interpreted along the same lines, with a more moderate misidentification rate due to the lower number of in-time additional interactions from pile-up.

As a first conclusion, levels of pile-up corresponding to at most 6.9 additional in-time interactions, as considered in this study, result in very small degradations of the vertexing resolution and in an intrinsic vertex inefficiency (corresponding to lost vertices) of 0.2-1.4%, depending on the sample and instantaneous luminosity. Apart from this intrinsic loss in performance, the most important degradation depends on the eventual inability to select the signal vertex out of the reconstructed vertices, which depends very strongly on the kind of physics event considered.

5.5.5 Selection of the signal primary vertex

The result of applying the default $\sum p_T^2$ vertex selection criterion introduced in Section 5.4.5 on the vertex candidates reconstructed by the adaptive multi-vertex finder in various kinds of physics events, in the presence of low luminosity pile-up, is summarised in Table 5.8. The two variables used to estimate the signal identification performance are the signal misidentification rate (*Misid.*) and the fraction of outliers (*Outliers*), as already introduced in Section 5.5.4. The misidentification rate with 6.9 additional in-time interactions from pile-up is around 0.5% on $t\bar{t}$ events, below 2% on $WH(\rightarrow l\nu u\bar{u})$ and around 4% for $WH(\rightarrow l\nu b\bar{b})$ events. In the case of $H \rightarrow \gamma\gamma$, with 4.6 additional in-time interactions from pile-up, quite a high misidentification rate of $\approx 24\%$ is seen.

The identification of the correct primary vertex is crucial in many studies and the selection of the wrong one will result - in many cases - in completely losing the corresponding fraction of events, thus typically reducing the signal selection efficiency for the physics process of interest, which is clearly highly undesirable. In the case of the $t\bar{t}$ process under consideration, where the top quarks do not both decay hadronically, and in the case of the $WH(\rightarrow l\nu b\bar{b})$ process, these events are expected to be collected by the on-line trigger acquisition system by requiring a lepton to be present in the event, above a certain p_T threshold (for illustration a value of 10 GeV is chosen). The influence of requiring a lepton with $p_T > 10$ GeV to be present in the event and to be associated with any of the reconstructed primary vertices on the signal identification inefficiency is shown in Table 5.9: the misidentification rate is essentially reduced to the intrinsic loss in vertexing reconstruction efficiency, which is as low as 0.2-0.3%. The eventual inefficiency due to lepton candidates not being associated to any of the reconstructed primary vertices is not considered here: however, these are mainly due to

¹⁰The geometrical definition does not consider the effect of pile-up vertices which by chance happen to be close to the signal vertex and therefore it might slightly underestimate the real fraction of misidentified signal vertices.

$\mathcal{L} = 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 75 \text{ ns}$, $\langle N_{pu} \rangle = 6.9$		
Sample	<i>Misid.</i>	<i>Outliers</i>
$t\bar{t}$	$(0.5 \pm 0.0)\%$	$(0.3 \pm 0.0)\%$
$WH(\rightarrow l\nu u\bar{u})$	$(1.7 \pm 0.0)\%$	$(1.4 \pm 0.0)\%$
$WH(\rightarrow l\nu b\bar{b})$	$(4.1 \pm 0.1)\%$	$(3.9 \pm 0.1)\%$

$\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 25 \text{ ns}$, $\langle N_{pu} \rangle = 4.6$		
Sample	<i>Misid.</i>	<i>Outliers</i>
$H \rightarrow \gamma\gamma$	$(24.0 \pm 0.1)\%$	$(24.0 \pm 0.1)\%$

Table 5.8: Result of applying the $\sum p_T^2$ selection criterion to identify the signal vertex in the presence of low luminosity pile-up, on various different physics samples, and based on the vertices reconstructed by the adaptive multi-vertex finding algorithm. Listed are the fraction of misidentified vertices (*Misid.*) and the fraction of outliers (*Outliers*), according to the definitions introduced in Section 5.5.4.

badly reconstructed lepton tracks (mostly electrons) and their association to one of the found primary vertices can be addressed by a dedicated procedure.

$\mathcal{L} = 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 75 \text{ ns}$, $\langle N_{pu} \rangle = 6.9$		
$t\bar{t}$		
Sample	<i>Misid.</i>	<i>Outliers</i>
$t\bar{t}$	$(0.2 \pm 0.0)\%$	$(0.1 \pm 0.0)\%$
$WH(\rightarrow l\nu u\bar{u})$	$(0.3 \pm 0.0)\%$	$(0.1 \pm 0.0)\%$
$WH(\rightarrow l\nu b\bar{b})$	$(0.3 \pm 0.0)\%$	$(0.1 \pm 0.0)\%$

Table 5.9: Same as Table 5.8, but the signal identification performance is evaluated after requiring one lepton with $p_T > 10 \text{ GeV}$ to be associated with any of the reconstructed primary vertices.

It can be therefore concluded that the loss in signal efficiency due to pile-up with up to 6.9 in-time overlaid interactions from pile-up is expected to be as low as 0.2-0.3% in events with a lepton with transverse momentum above 10 GeV. This is an important result for the study of $WH(\rightarrow l\nu b\bar{b})$ which will be presented later in the course of the present thesis.

The case of $H \rightarrow \gamma\gamma$ is a bit more complex. If the primary vertex is misidentified, the event is not lost, but the mass resolution of the di-photon pair is significantly degraded. The primary vertex position is used to correct the direction of the photon momenta as reconstructed in the calorimeter, which are otherwise computed with the assumption that they originate in the centre of the ATLAS reference frame (0,0,0): however, due to the spread in z of the interaction region, the real primary vertex position can be significantly displaced with respect to $z = 0$ and this results in a significant bias on the polar direction of the photon momenta. The present study shows that the wrong primary vertex will be picked up in 24% of the cases in the nominal low luminosity running scenario. Even if this is outside the scope of a general vertexing algorithm, since, as quoted in Table 5.7, most of the signal vertices are correctly reconstructed but not correctly selected (the fraction of lost vertices is below 1.5%), there are two more handles which can be used to select the correct signal vertex:

- In above 50% of the events a photon converts into an e^+e^- . Conversion candidates can be extrapolated to the beam line and used to select the correct primary vertex candidate.
- The position of the energy deposits of the photons in the various layers of the ATLAS calorimeter can be combined and used to constrain the z position of the signal primary

vertex. The achievable resolution is only ≈ 17 mm, varying significantly according to the photon pseudo-rapidity, but is usually sufficient to select the correct vertex candidate out of the two or three most signal like ones.

Combining all these selection criteria, a significant decrease of the signal misidentification rate can be obtained, corresponding to an improvement of the Higgs invariant mass resolution and thus of the discovery potential of a Higgs boson in the low mass region in the $H \rightarrow \gamma\gamma$ channel (see Ref. [44] for more details).

5.5.6 Number of vertices as a measure of instantaneous luminosity

Another important point is how well the number of primary vertices reconstructed by the vertex finding algorithms reproduces the total number of signal plus pile-up vertices. Once the bunch crossing spacing is defined, the instantaneous luminosity is directly proportional to the average number of in-time pile-up vertices, which is expected to be distributed according to a Poisson distribution, and the average number of reconstructed vertices could thus provide a measurement of the instantaneous luminosity. Even if at the beginning the cross section for minimum-bias events will only be poorly known, so that this measurement will be unsuited to get a handle on the absolute instantaneous luminosity, it will still be useful to get a relative measure of it, with a high resolution in time, in such a way to be able to follow in detail the decreasing instantaneous luminosity of the beams of the LHC machine in the course of a single fill. The distribution for the number of reconstructed vertices is shown in Fig. 5.12. The default vertex finding algorithm significantly overestimates the number of vertices. A significant amount of pile-up vertices is being split into more than one vertex, due to the fixed window size in the z direction (5 mm) used during the finding when separating the tracks into first vertex candidates: further tuning of the window size and a refined clustering algorithm will certainly improve the situation. The adaptive multi-vertex finder, on the other side, can more accurately reproduce the distribution of the number of simulated vertices, but loses some of the pile-up vertices. This inefficiency is however expected and could be accounted for when counting the number of vertices.

In the presence of no other vertices, a pile-up event (minimum bias), as shown in Table 5.5, has a reconstruction inefficiency of slightly less than $\approx 10\%$. However, the probability of two vertices to overlay and thus the reconstruction inefficiency of pile-up vertices is expected to further increase with the number of simulated pile-up vertices. The expected distribution of number of reconstructed vertices, starting from the true distribution and considering an average inefficiency of 15% on pile-up events, is also shown in Fig. 5.12 in the left column, and is quite close to the distribution of the adaptive finder. The agreement could be further improved by studying the pile-up vertex reconstruction inefficiency as a function of the number of simulated vertices and by getting a pile-up vertex inefficiency correction factor as a function of the number of reconstructed vertices. An eventual measurement of the relative evolution of the instantaneous luminosity of this kind would take into account several sources of systematic errors, the most important ones being:

- the vertex reconstruction inefficiency in minimum bias events (as a function of vertex multiplicity);
- the fraction of signal vertices which split into more than one vertex and its dependence on the specific physics process considered;
- the influence of tracks from out-of-time pile-up events.

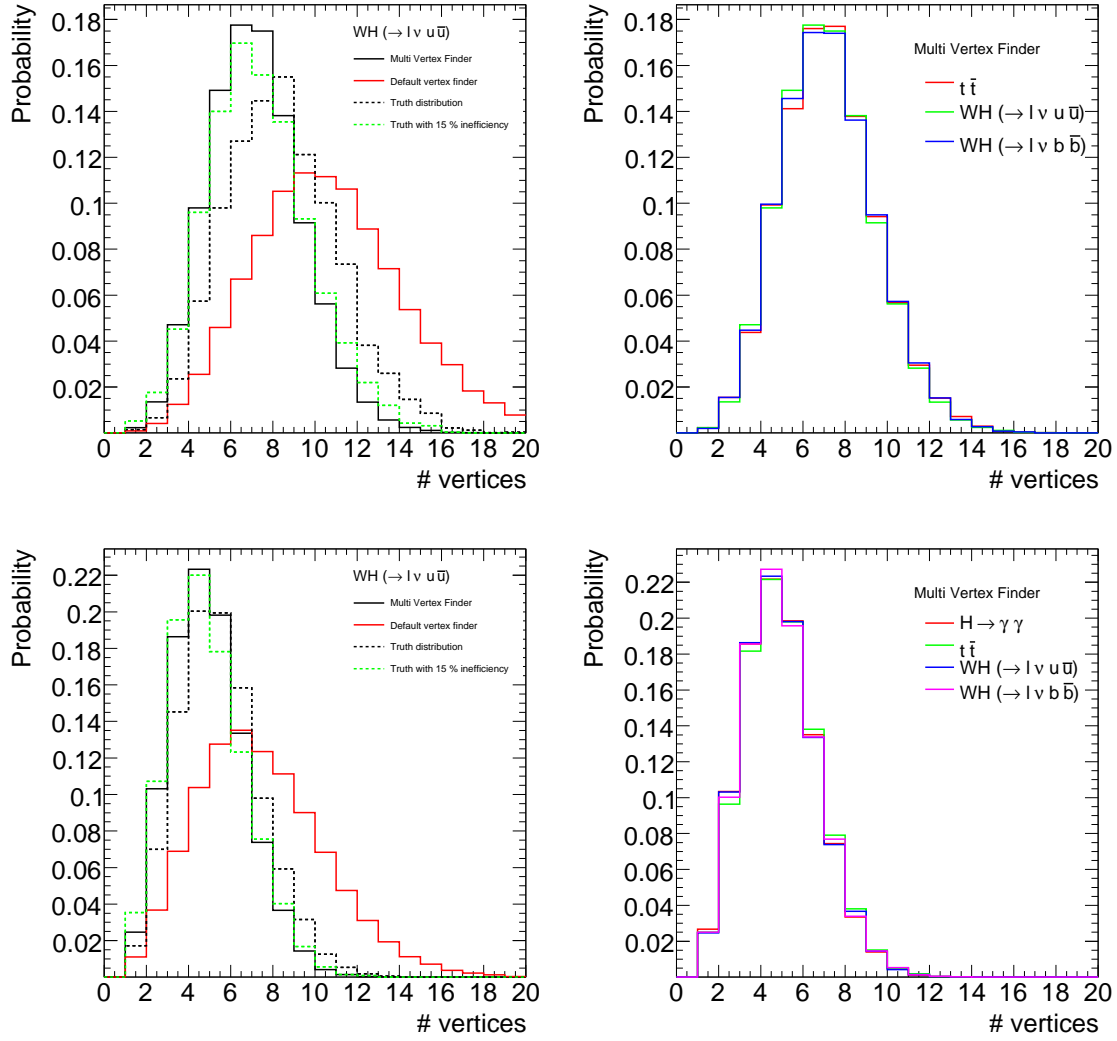


Figure 5.12: Number of reconstructed primary vertices per events in the presence of in average 6.9 in-time additional interactions from pile-up events (top row) and 4 average in-time additional interactions from pile-up events (bottom row). In the left column this is shown for $t\bar{t}$ events, applying various different vertex finders, while in the right column it is shown for the the multi-vertex finder only, applied to different kind of physics events. In the left column the number of reconstructed vertices is also compared with the number of real simulated vertices, with or without applying a single pile-up vertex reconstruction inefficiency of 15%.

While a method will hopefully be found to measure the vertex reconstruction efficiency in minimum bias events, a detailed estimation of the connected uncertainty will be crucial to get an idea of the precision of such a luminosity measurement. The second point seems to be fairly well under control, since, as shown in Fig. 5.10, a single hard interaction vertex seems to split into two in at most 2% of the cases, while the fraction of such splittings is negligible for minimum bias vertices. This is confirmed by Fig. 5.12 (right column), which shows that the dependence on the number of reconstructed vertices on the kind of hard physics event considered is small. The use of single-track vertices and of dedicated tracking algorithms to reconstruct tracks with transverse momenta as low as 150 MeV will further increase the pile-up vertex reconstruction efficiency, however a detailed study is needed to understand if also the amount of wrongly reconstructed or split vertices would increase in this case.

5.5.7 Improvement of track momenta resolution using vertex constraint

An interesting feature of a vertex fit is the possibility to improve the estimation of the helix parameters of a charged particle track associated to a vertex under the hypothesis that the track has its real geometrical origin in that vertex. In the case of a Kalman filter (Section 5.3.2) this is achieved during the *smoothing* phase. The improvement achievable by using the vertex constraint is illustrated for $t\bar{t}$ physics events in Fig. 5.13, where the direction of the track at the vertex is compared to its true direction by using Monte Carlo information, before and after *smoothing* is applied.

The improvement is very significant for low p_T tracks (e.g. for p_T below 5 GeV the resolution on the track direction can be improved by a factor of 3) and tends to become much smaller for high p_T tracks (e.g. for tracks with p_T above 20 GeV the improvement in the core resolution is around 30% for the azimuthal direction and very roughly a factor of 2 in the polar direction, and the improvement in the tails is even smaller). The achievable relative improvement depends on how significantly the track contributes to the vertex fit with respect to the remaining tracks associated to the same vertex. In case a track with $p_T > 20$ GeV is considered (which, in $t\bar{t}$ events, is most likely a prompt lepton from the leptonic decay of one of the two W bosons), this track is already expected to nearly dominate the primary vertex fit, so the knowledge coming from the other tracks can more hardly further improve the result.

While the application of the vertex constraint is also useful for low p_T physics, one of the most interesting applications is the improvement of the invariant mass resolution of lepton pairs originating from weak boson decays. This is the case for example for the decay of $Z \rightarrow \ell\ell$ or, even more important, the decay of $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, where a good invariant mass resolution is crucial to extract the Higgs signal out of the 4ℓ continuum $ZZ^{(*)}$ background. In this case the primary vertex fit is expected to be dominated by four high p_T leptons: when applying the vertex constraint on each of the four tracks, the additional knowledge coming from the remaining high p_T tracks is used and an improvement in the resolution of track momenta can be obtained. For this reason the adaptive vertex fitter has been recently applied also to the $H \rightarrow 4\ell$ analysis [44].

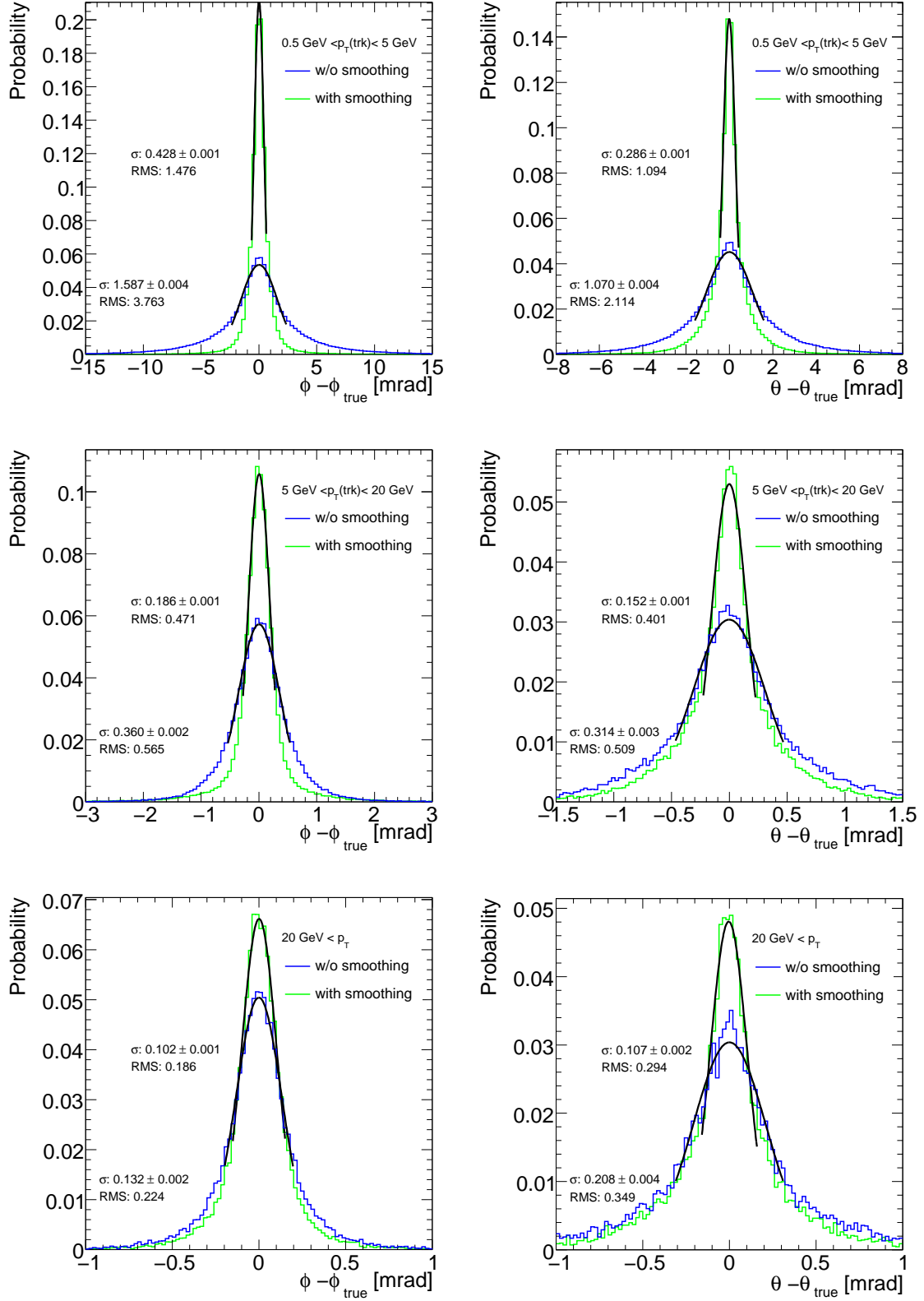


Figure 5.13: Resolution on the azimuthal (left column) and polar (right column) direction of tracks, before and after *smoothing*, for various ranges of transverse track momenta, evaluated in $t\bar{t}$ events with pile-up corresponding to $\mathcal{L} = 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ instantaneous luminosity and $\Delta t_{BC} = 75 \text{ ns}$.

6 Identification of b -quark jets

6.1 Motivation

The identification of jets originating from *bottom*-quarks (denoted as b -tagging in the following) is an important ingredient for the high p_T physics program planned for the ATLAS experiment, since many interesting physics processes contain *bottom*-quarks in the final state, while the most abundant backgrounds contain mostly up, down and strange quark or gluon jets or, in a smaller fraction of cases, charm quark jets. The reason for this interest is that b -quarks are the heaviest quarks in the Standard Model which still form hadrons before undergoing a weak decay, contrary to what happens for *top* quarks, the heaviest quarks in the Standard Model, which can be detected only indirectly by analysing their decay products (in most of the cases, since $|V_{tb}| \approx 1$, again a bottom quark and an additional W boson).

The aim of b -tagging is therefore to extract the b -quark jets with high efficiency, while rejecting most of the background contamination from jets originating from the fragmentation of light (u , d and s) quarks, gluons and c -quarks.

From the physics point of view the hadronization of b -quarks has several unique properties, which can be exploited by b -tagging. A b -quark, once produced, fragments necessarily into a b -flavoured hadron, since strong interactions are flavour blind. This is, in most of the cases ($\approx 87\%$), first an excited b -hadron, like a B^* or a B^{**} , which decays immediately, strongly or electromagnetically, into a ground state b -hadron plus one or more further particles, while in the remaining cases, a ground state b -hadron is produced directly. From an experimentalist perspective, one is however only interested in the transition from a b -quark into the final ground state b -hadron, since the timescale typical for electromagnetic and strong interactions is so small that the B^* , B^{**} decay vertices are not significantly displaced with respect to the primary vertex. The various different fractions of ground state b -hadrons produced out of the fragmentation of an original b -quark are presented in Table 6.1. In most of the cases ($\approx 91\%$) a B-meson is produced.

b -hadron	Branching fraction (Γ_i/Γ)
B^+	$(40.0 \pm 1.2)\%$
B^0	$(40.0 \pm 1.2)\%$
B_s^0	$(11.4 \pm 2.1)\%$
b -baryon	$(8.6 \pm 2.1)\%$

Table 6.1: Branching fraction of b -hadrons produced out of the fragmentation of b -quarks. Taken from Ref. [71].

The fragmentation function describes the distribution of the fraction of energy of the original b -quark which is kept by the b -hadron. Due to the b -quark fragmentation function being very hard, most of the original b -quark energy is transmitted to the final b -hadron: this fraction is for example on average $\approx 70\%$ for b -quarks with a momentum of ≈ 45 GeV. This property can be exploited during b -tagging, since the fragmentation function for light-quarks into light hadrons or c -quarks into c -hadrons is softer.

Any of the finally produced b -hadrons decay through weak interactions and therefore have a

significant lifetime, which is on average, for all b -hadrons considered, $(1.568 \pm 0.009) \times 10^{-12} \text{ s}$. The effective distance travelled in the detector by the b -hadron before decaying depends on the b -hadron momentum, which enters the relativistic boost factor $\beta\gamma$. This means that a b -quark with a momentum of 50 GeV will travel around 3 mm, which is a visible flight length in the detector. Due to the combination of the b -hadron lifetime and relatively high mass ($m_B \approx 5.28 \text{ GeV}$), which results in a non-negligible decay angle of the b -hadron decay products with respect to the b -hadron flight direction, the charged particles produced at the decay vertex will be on average significantly displaced with respect to the primary vertex position.

This is the main signature which is exploited by the *lifetime* based b -tagging algorithms, which are based either on the presence of significantly displaced tracks, as in impact parameter based b -tagging algorithms, or on the explicit reconstruction of the b -hadron decay vertex, as in secondary vertex based b -tagging algorithms.

Weak decays are governed by the CKM matrix mechanism [72, 73]: since $|V_{cb}|^2 \gg |V_{ub}|^2$, b -hadrons decay preferably into a c -hadron plus additional particles. These c -hadrons can be again excited states, like D^* or D^{**} , but these again decay with negligible lifetime to weakly decaying c -hadrons. The fractions of b -decays where various different kinds of c -hadrons are produced are shown in Table 6.2: they clearly show that in most of the cases at least one c -hadron is produced out of the decay of the b -hadron. The lifetime of a c -hadron, as shown in the same table, is not much lower than for b -hadrons, but in general the momentum of the c -hadron will be lower than the original b -hadron momentum. However, the c -hadron can still travel for a significant path in the detector and form with its decay products a visible *tertiary* vertex, displaced with respect to both secondary and primary vertices.

c -hadron	Branching fraction (Γ_i/Γ)	D-meson lifetime (τ)
$B \rightarrow D^\pm$ anything	$(23.3 \pm 1.3)\%$	$(1.040 \pm 0.007) \times 10^{-12} \text{ s}$
$B \rightarrow D^0/\bar{D}^0$ anything	$(62.2 \pm 2.9)\%$	$(0.4101 \pm 0.0015) \times 10^{-12} \text{ s}$
$B \rightarrow D_s^\pm$ anything	$(22.5 \pm 1.5)\%$	$(0.500 \pm 0.007) \times 10^{-12} \text{ s}$

Table 6.2: Main inclusive B-meson decay modes with charm mesons in the final state. Taken from Ref. [71].

There are several possible strategies to deal with such tertiary vertices from c -hadron decays. In the case of b -tagging algorithms based purely on impact parameter information from the charged particle tracks, the presence of tracks with a more significant displacement does not require an explicit strategy: these tracks will in general just improve the b -tagging performance. Their presence is taken into account when the algorithm is calibrated using Monte Carlo simulated events or, later, as soon as the amount of collected data will allow it, using directly data. In the case of a secondary vertex based b -tagging algorithm, the more conventional approach is to try to find a single inclusive secondary vertex, which will result in a weighted average of the b - and one or more c -hadron vertex positions: this approach is reasonable, since in most of the cases either the resolution of the tracks or the decay vertex multiplicities at such vertices do not allow to separate them efficiently. An alternative approach which instead tries to resolve the two b - and c -hadron decay vertices, but at the same time uses a kinematic constraint to try to overcome both the problem of the limited decay multiplicities and of the limited track resolutions, was developed in the course of this thesis and will be described in detail later.

Another property which is usually exploited by b -tagging is the fraction of b - and c -hadron decays into leptons: a lepton from the semi-leptonic decay of a b -hadron ($b \rightarrow \ell$) or from the subsequent c -hadron decay ($b \rightarrow c \rightarrow \ell$) turns out to be produced in a b -quark in $\approx 21\%$ of

the cases: the fraction of $b \rightarrow \ell$ decays is $\approx 11\%$, while the fraction of $b \rightarrow c \rightarrow \ell$ decays is $\approx 10\%$. This is valid both in case the lepton is an electron ($\ell = e$) or a muon ($\ell = \mu$), which brings the overall fraction of b -quarks ending up into a lepton to $\approx 42\%$. Due to the b - or c -hadron mass, the lepton will be emitted with an average transverse momentum comparable with $m_{b\text{-had}}$ or $m_{c\text{-had}}$. By identifying either an electron or a muon originating from a jet and by requiring it to have a sufficiently high transverse momentum with respect to the jet axis (which usually yields a reasonable approximation of the b - and c -hadron directions), it is possible to identify b -jets, rejecting light jets, where leptons are expected mainly from the in-flight decay of charged pions and kaons, from Dalitz decays of neutral pions, from γ -conversions or from misidentified leptons.

6.2 Selection and association of tracks to jets

In ATLAS the b -tagging algorithms rely on the jets as reconstructed by a jet algorithm. The reconstruction of jets, as introduced in Section 4.3, is intended to match the final state colourless hadrons reconstructed as energy deposits in the calorimeters to the colour-connected parton level description of the physics process of interest. Independently on the chosen jet algorithm, the jet reconstruction returns a well defined set of four momenta, which give a certain representation of the set of partons in the hard scattering event.

The b -tagging algorithm then takes as input the three-momenta of these jets and tries to address the question whether they originated or not from a b -quark. In particular the jet direction is used to associate the charged particles reconstructed as tracks in the inner detector to the jet. Since the 2 Tesla solenoidal magnetic field of the ATLAS inner detector bends charged particles in the transverse plane, in particular in the case of low p_T tracks, the tracks are best matched to the jet by using their momenta at the point of closest approach to the interaction region. If \vec{p}_{jet} is the jet momentum and \vec{p}_{trk} is the track momentum, the criterion for associating charged particle tracks to jets is simply:

$$\Delta R(\vec{p}_{jet}, \vec{p}_{trk}) < \Delta R_{cut},$$

where usually a value of $\Delta R_{cut} = 0.4$ is used, independently on the parameters used for jet clustering. Some studies have been performed in ATLAS, where it has been demonstrated that making the ΔR_{cut} dependent on $p_{T,jet}$, the performance can be improved, but typically only in the case of significant contamination of light quark jets in the presence of close lying b -quark jets [44].

After the tracks are associated to the jets, they are filtered in order to remove tracks with bad quality or which can easily erroneously be identified as secondary tracks from b -decays. These include tracks originating from decays of even longer lived particles, like K_S^0 ($c\tau \approx 2.69$ cm) and Λ baryons ($c\tau \approx 7.89$ cm), from electromagnetic interactions in the detector material like conversions in electron-positron pairs ($\gamma \rightarrow e^+e^-$) or from hadronic interactions with the detector material, which result in two or more tracks with high impact parameter.

The track selection needs therefore to be different compared to the one defined for primary vertex finding in Section 5.5.1. It needs to be particularly tight in the case of the impact parameter based b -tagging algorithms, since in that case the explicit presence of a vertex is not required, so that the influence of badly reconstructed tracks or tracks from long lived particles or material interactions does directly limit the intrinsically achievable performance.

First, to reject badly reconstructed tracks, quality cuts are applied. They are listed in Table 6.3 for both impact parameter and secondary vertex based b -tagging algorithms.

In addition, for the impact parameter based b -tagging algorithms, the p_T cut on the tracks is increased to $p_T > 1$ GeV. The reason for this is that, as already explained, in a b -jet most

Quantity	Tight selection	Loose selection
Transverse momentum	$p_T > 1 \text{ GeV}$	$p_T > 500 \text{ MeV}$
Error on transverse impact parameter	$\sigma(d_0) < 350 \mu\text{m}$	
Error on longitudinal impact parameter	$\sigma(z_0) < 2.5 \text{ mm}$	
Track fit quality	-	$\chi^2/\text{dof} < 3.5$
Minimum number of b-layer hits	$n_{b\text{-lay}} \geq 1$	-
Minimum number of pixel detector hits	$n_{pix} \geq 2$	$n_{pix} \geq 1$
Minimum number of SCT detector hits	-	$n_{SCT} \geq 4$
Minimum number of pixel+SCT detector hits	$m_{pix} + n_{SCT} \geq 7$	$m_{pix} + n_{SCT} \geq 7$

Table 6.3: Selection cuts applied on tracks to be used as input for the impact parameter based (Tight selection) or the secondary vertex based (Loose selection) b -tagging algorithms.

of the momentum is kept by the b -hadron, which has a decay multiplicity of around five charged particles, so that on average their p_T is higher than for the remaining tracks from the fragmentation, as shown in Fig. 6.1. It is interesting to note that, even if the fragmentation function for c -jets is on average softer than for b -jets ($\approx 55\%$ of the original c -quark energy is kept by the c -hadron, against $\approx 70\%$ of a b -jet), the average charged decay multiplicity for a c -hadron decay is ≈ 3 against ≈ 5 of a b -jet, so that the transverse momentum per track is larger in c -hadrons produced in c -jets than in b -hadrons produced in b -jets.

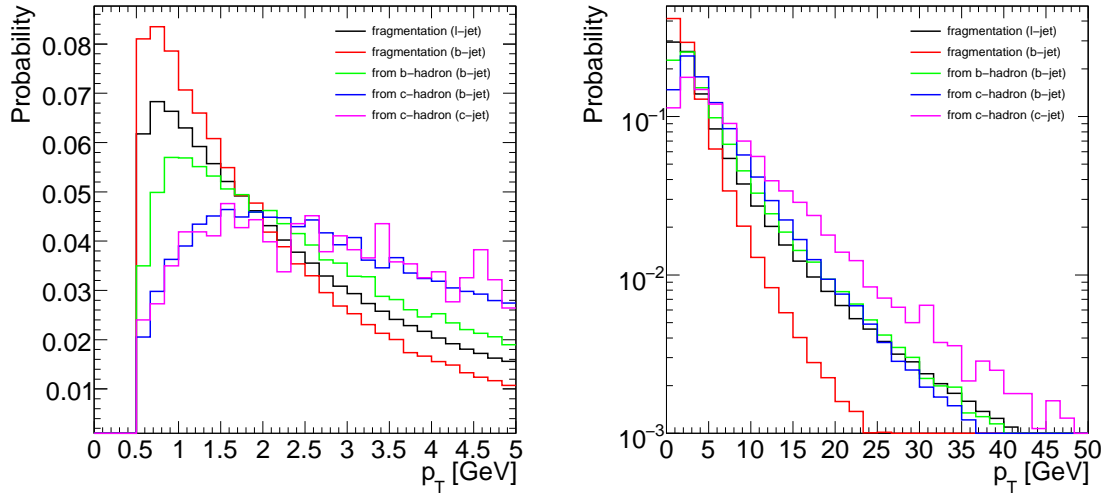


Figure 6.1: Transverse momentum of tracks associated to light, c - and b -jets, subdivided into various different categories according to their physical origin, in linear (left) and logarithmic (right) scale.

Selection cuts on the transverse and longitudinal impact parameters are crucial for b -tagging and are worth a more detailed discussion. These impact parameters are computed with respect to the primary vertex, where all primary particles are supposed to originate from. The primary vertex is reconstructed by using the adaptive multi-vertex finding algorithm, as described in detail in Chapter 5. In practise, the tracks are extrapolated to the point of closest approach to the primary vertex in the transverse plane; their five helix parameters, at this point, are defined as described in Section 5.3.3. Then, the impact parameters, $IP_{r\phi}$ and IP_z are computed in the following way:

$$IP_{r\phi} = d_0 \quad (6.1)$$

$$IP_z = z_0 \sin(\theta) \quad (6.2)$$

As can be most easily seen in Fig. 5.1, the longitudinal impact parameter, which in the figure is the distance of closest approach of the track to the origin (primary vertex position) in the x - z plane, is not directly z_0 , but, approximating the track around the point of closest approach by a straight line, it can be obtained by multiplying z_0 by $\sin(\theta)$.

When computing impact parameter significances, the errors on $IP_{r\phi}$ and on IP_z need to take into account both the uncertainty on the track trajectory and the uncertainty on the primary vertex. Both effects are considered, but in general the uncertainty on the primary vertex is small compared to the trajectory uncertainty.

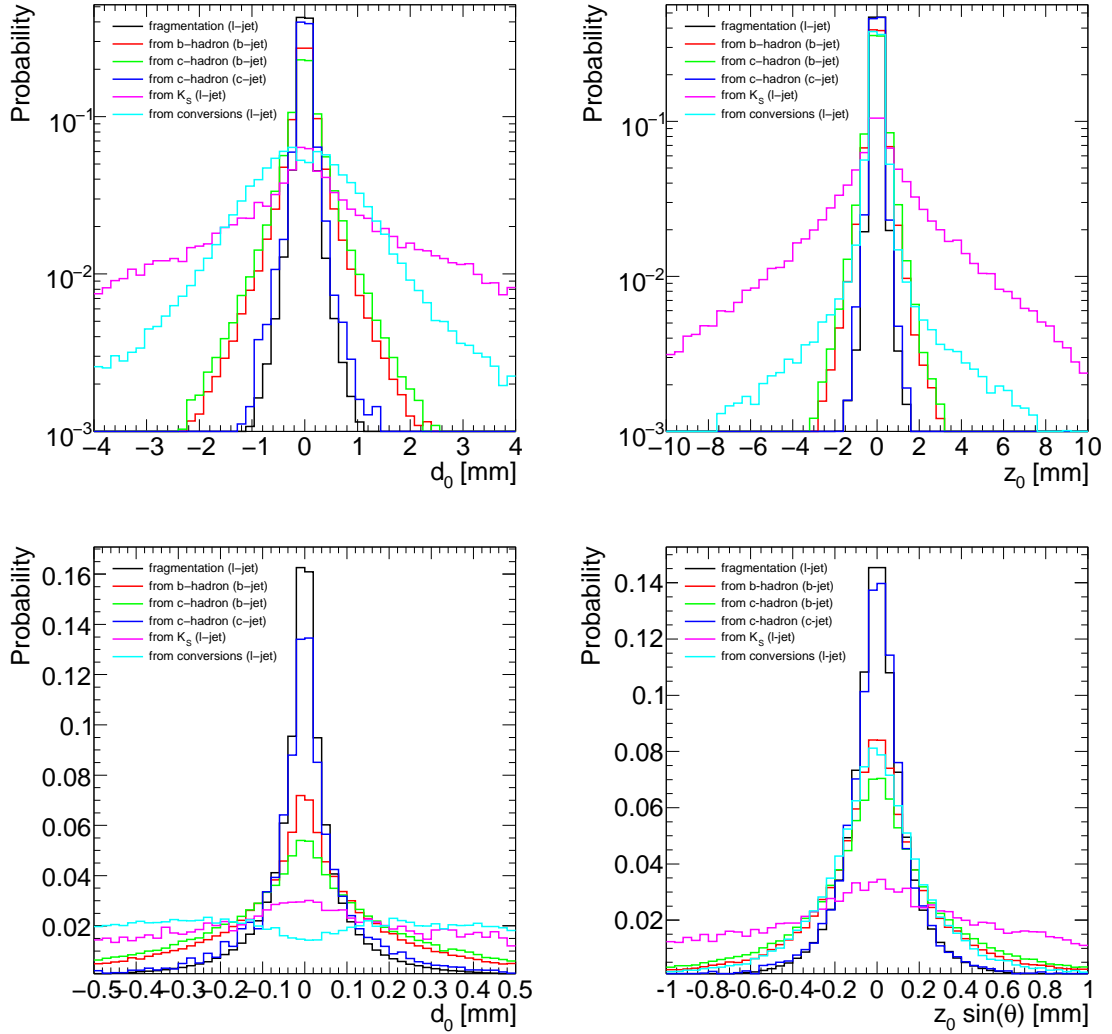


Figure 6.2: Impact parameters of tracks associated to light, c - and b -jets in the transverse (left column) and longitudinal (right column) direction, subdivided into various categories according to their physical origin, in logarithmic scale (top row) and linear scale (bottom row). These distributions are obtained from a sample of $pp \rightarrow WH$ events (with $H \rightarrow b\bar{b}, u\bar{u}$ and $c\bar{c}$).

The distributions for $IP_{r\phi}$ and for IP_z are shown in Fig. 6.2, for tracks reconstructed in b -, c - or light jets, subdivided according to the physics process they originated from, as derived from Monte Carlo information. In addition, Table 6.4 shows the relative composition of tracks

in light, c - and b -jets in terms of the various involved physics processes, after applying three different selection criteria on the impact parameters. The impact parameter distributions shows that the tracks originating from weakly decaying b -hadrons have significantly higher transverse and longitudinal impact parameters compared to prompt tracks originating directly from fragmentation. Tracks originating from c -hadrons in c -jets have impact parameters only a bit larger than prompt tracks and tracks originating from c -hadrons in b -jets, i.e. from the $b \rightarrow c$ -hadron cascade, have impact parameters only slightly larger than tracks arising directly from the weakly decaying b -hadrons. If only prompt tracks would be present in light-jets, no track selection cuts would be needed in the transverse and longitudinal impact parameters and the impact parameter significance, in both directions, could be used directly in the b -tagging algorithm, since it yields the best separation between primary and secondary tracks. Fig. 6.2 shows clearly that the impact parameters of conversions and K_S^0 mesons extend to significantly higher values than for tracks from heavy flavour decays. The same holds for Λ baryons and material interactions, which in the table are not listed explicitly, but are included in the remaining 3% of tracks in light quark jets. The contamination of such displaced tracks in light-jets, before any track selection is applied, is above 5%: these correspond to tracks with high impact parameters and would therefore severely reduce the b -tagging performance.

Selection	
initial	$ IP_{r\phi} < 7 \text{ mm}, IP_z < 10 \text{ mm}$
loose	$ IP_{r\phi} < 3.5 \text{ mm}, IP_z < 5 \text{ mm}$
tight	$ IP_{r\phi} < 1 \text{ mm}, IP_z < 1.5 \text{ mm}$

	Selection	light-jet	c -jet	b -jet
Average track multiplicity	initial	5.4	6.1	6.9
	tight	5.3	5.9	6.8
	loose	5.0	5.7	6.2

Track	Selection	light-jet	c -jet	b -jet
from fragmentation	initial	91.5 %	61.2 %	36.3 %
	loose	92.9 %	62.0 %	36.9 %
	tight	95.6 %	64.0 %	39.5 %
from b -hadron	initial	0.0 %	0.0 %	26.4 %
	loose	0.0 %	0.0 %	26.8 %
	tight	0.0 %	0.0 %	27.1 %
from c -hadron	initial	0.1 %	30.3 %	28.6 %
	loose	0.1 %	30.8 %	29.0 %
	tight	0.1 %	31.7 %	28.6 %
from K_S	initial	1.7 %	2.0 %	1.9 %
	loose	1.2 %	1.5 %	1.4 %
	tight	0.5 %	0.6 %	0.6 %
from conversion	initial	3.9 %	4.0 %	3.7 %
	loose	3.7 %	3.8 %	3.5 %
	tight	2.5 %	2.5 %	2.2 %
other	initial	2.8 %	2.5 %	3.1 %
	loose	2.1 %	1.9 %	2.6 %
	tight	1.3 %	1.2 %	2.0 %

Table 6.4: Average track multiplicity and relative contribution of tracks originating from different physics processes in light, c - and b -quark jets, after either initial, loose or tight selection cuts on the impact parameters d_0 and $z_0 \cdot \sin(\theta)$ have been applied. The values are normalised respectively to the overall number of tracks in light, c - and b -jets. The quality track selection criteria in the *Loose* version have already been applied in all cases. These numbers are obtained from a sample of $pp \rightarrow WH$ events (with $H \rightarrow b\bar{b}, u\bar{u}$ and $c\bar{c}$).

The solution adopted is to select only tracks with transverse impact parameter $|d_0| < 1 \text{ mm}$ and longitudinal impact parameter $|z_0 \cdot \sin(\theta)| < 1.5 \text{ mm}$: the physics motivation for this cut

is that the average impact parameter depends nearly only on the lifetime of the decaying particle, since it is - up to a good approximation - independent on its Lorentz boost. A higher flight length (proportional to $\beta\gamma$) is compensated by a smaller decay angle (proportional to the inverse of $\beta\gamma$), so that the two effects nearly cancel each other. The effect of applying this cut is illustrated in Table 6.4 (tight selection): the relative amount of tracks from long lived particles, conversions and material interactions is significantly decreased, while around $\approx 91\%$ of the tracks from weakly decaying b - and c -hadron vertices are kept. In the case of the secondary vertex based b -tagging algorithms a much looser selection is applied, cutting at $|d_0| < 3.5$ mm and $|z_0 \cdot \sin(\theta)| < 5$ mm: the efficiency and resulting composition is shown in Table 6.4 (loose selection). Nearly all tracks from b - and c -hadron decays are kept ($\approx 99.9\%$), at the cost of keeping also most of the long lived particles, conversions and hadronic interactions whose decay vertex are sufficiently close to the primary vertex. This is however desired, since the vertices corresponding to conversions, long lived particles and hadronic interactions can, in many cases, be explicitly reconstructed.

An interesting feature to note in Fig. 6.2 is the different behaviour of tracks from conversions and from K_S^0 decays for the transverse and longitudinal impact parameters. Due to the converting photons being massless, the tracks from conversions are emitted at zero opening angle with respect to the original photon direction: while in the transverse plane the electron tracks are bent by the magnetic field and result in a high impact parameter, in the longitudinal plane the tracks are nearly compatible with originating directly from the primary vertex, with some tails due to late conversions and electron bremsstrahlung. Opposite to that, charged particle tracks from K_S^0 meson decays, due to their mass of around 500 MeV, are emitted with non-zero opening angle and are therefore characterised by a significant impact parameter both in the longitudinal and transverse directions. In the case of conversions, this feature can be used in the impact parameter based b -tagging algorithms as an additional handle to discriminate against conversions.

6.3 Impact parameter based b -tagging algorithms

If long lived particles, conversions and hadronic interactions are not considered, the best discrimination between prompt tracks and displaced tracks from b - and c -hadron decays can be obtained by using the impact parameter significance both in the transverse plane and longitudinal direction ($IP_{r\phi}$ and IP_z divided by their respective errors).

In order to increase the discriminating power of the impact parameter significance, a lifetime sign is assigned to these variables (replacing the sign of the geometrical definition of the impact parameter): the lifetime sign is positive if the track is more likely to intersect the flight axis in front of the primary vertex (i.e. is more compatible with having its origin in a secondary decay vertex in the direction of flight expected for the b -hadron) or negative if the track is more likely to intersect the flight axis behind the primary vertex, opposite to the jet direction. Both cases are illustrated in Fig. 6.3, together with the variables needed to define the lifetime sign. The vector $\vec{r}_{IP} - \vec{r}_{PV}$ defines the three-dimensional impact parameter of the track with respect to the primary vertex, \vec{p}_{jet} the jet momentum and \vec{p}_{trk} the track momentum defined at the point of closest approach to the primary vertex.

The lifetime sign can then be defined in three-dimensions, according to the variables \vec{p}_{jet} , \vec{p}_{trk} and $\Delta\vec{r}_{IP} = \vec{r}_{IP} - \vec{r}_{PV}$:

$$\text{sign}_{3D} = \text{sign}([\vec{p}_{trk} \times \vec{p}_{jet}] \cdot [\vec{p}_{trk} \times \Delta\vec{r}_{IP}]), \quad (6.3)$$

where the symbol \cdot denotes the scalar product and the symbol \times the cross product of two vectors.

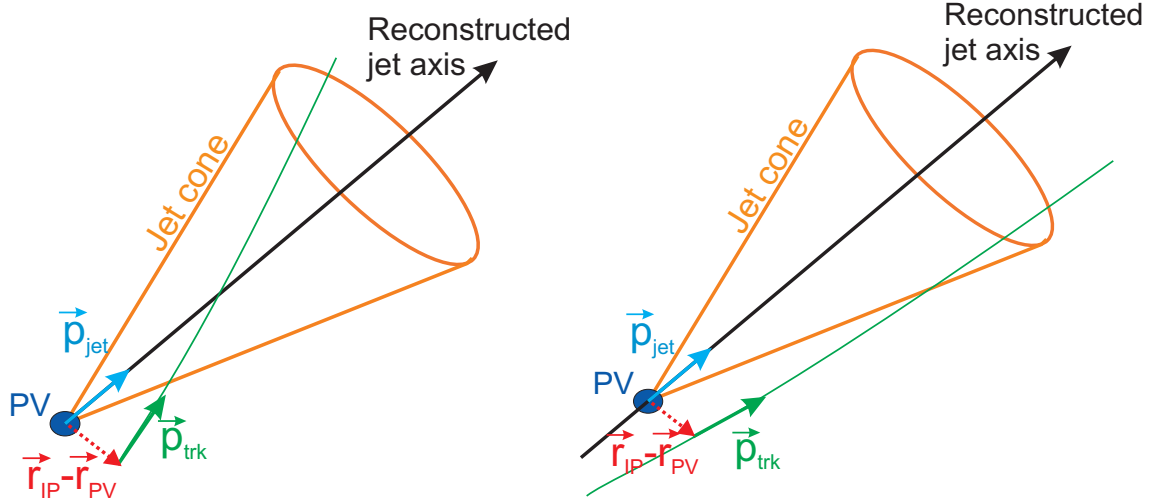


Figure 6.3: Definition of variables needed to compute the lifetime sign of a track in the three-dimensional space. In addition a positive (left) and negative (right) lifetime signed track is shown.

This formula can be more easily understood by first considering the hypothesis that all three vectors \vec{p}_{jet} , \vec{p}_{trk} and $\Delta\vec{r}_{IP}$ lie in the same plane. In this case the jet axis and the track, approximated with a straight line starting from the impact parameter point, will necessarily intersect each other. The lifetime sign is defined in such a way that, if this intersection occurs in the direction of flight of the b -jet, it is positive, if the intersection occurs on the other side, it is negative.

While the formula of the three-dimensional lifetime sign (Eq. 6.3) relies on the three-dimensional impact parameter $\vec{r}_{IP} - \vec{r}_{PV}$, the lifetime sign can be also defined on the transverse plane (x - y) or on the longitudinal plane ($r\phi$ - z) by considering the projections of the three-dimensional impact parameter respectively on the x - y and $r\phi$ - z planes, which is equivalent to considering respectively the transverse and longitudinal impact parameters. In these cases the three-dimensional formula simplifies and, considering the definition of the helix parameters d_0 and z_0 as illustrated at the beginning of Section 5.3.3, their lifetime sign can be expressed as:

$$\text{sign}_{r\phi} = \text{sign}(\sin(\phi_{jet} - \phi_{trk}) \cdot d_{0,trk}) \quad (6.4)$$

$$\text{sign}_z = \text{sign}((\eta_{jet} - \eta_{trk}) \cdot z_{0,trk}) \quad (6.5)$$

When moving from Eq. 6.3 to Eq. 6.4, the geometrical definition of the sign of the helix parameter d_0 needs to be taken into account, which is such that the vector $\vec{p}_{trk} \times \Delta\vec{r}_{IP}$, once \vec{p}_{trk} and $\Delta\vec{r}_{IP}$ are projected on the transverse plane, is simply $\text{sign}(d_0)$ times the unit vector pointing along the z axis in the $z > 0$ direction.

The computation of the lifetime sign assumes that the jet direction reproduces, up to a good approximation, the b -hadron direction: from the physical point of view this assumption can only be fulfilled in an approximate way, since the jet momentum is supposed to reproduce the momentum of the initial b -quark, which is given by the sum of the momentum of the b -hadron and of the remaining tracks arising directly from fragmentation. Under this assumption and up to resolution effects both on the jet direction and on the impact parameters and momentum of the track, the lifetime sign for tracks originating from b -hadron decays is positive.

The distributions for the transverse ($IP_{r\phi}$) and longitudinal (IP_z) impact parameter significance are shown in Fig. 6.4: the sign of the transverse impact parameter is defined by

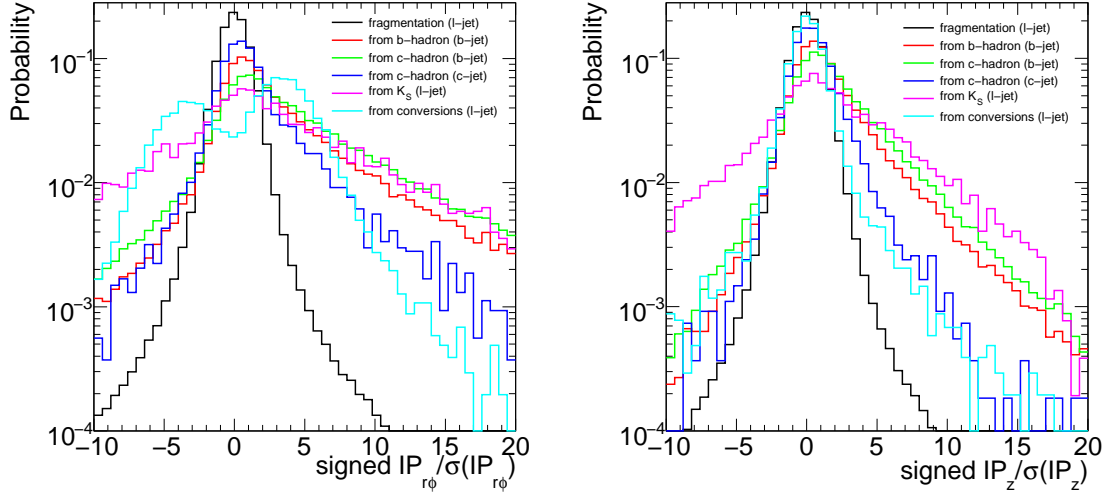


Figure 6.4: Impact parameter significance of tracks associated to light, c - and b -jets in the transverse (left column) and longitudinal (right column) direction, subdivided into various different categories according to their physical origin.

“ $\text{sign}_{r\phi}$ ”, while the longitudinal impact parameters significance uses “ sign_z ”. It can be noticed that tracks from fragmentation in light-jets¹ have a signed impact parameter distribution which is perfectly symmetric around 0, since they have no correlation with the jet direction. Tracks from b - and c -hadron decays, as expected, have an asymmetric distribution, with the most significant contribution at positive significances; however a negative tail extending beyond the pure fragmentation contribution is also seen, corresponding to resolution effects and to an eventual mismatch between the b -jet and b -hadron directions. The signed impact parameter significance distribution in light-jets would be perfectly symmetric if no contribution from long lived particles, conversions and hadronic interactions would be present: among these, the distribution for conversions and K_S^0 mesons are shown and, since they are also emitted at some secondary vertex which is preferably in the direction of the jet, they also show a significant contribution at high positive impact parameter significances. However, since the jet direction can provide a good approximation only for the b -hadron direction in b -jets (since only the b -jet fragmentation function is sufficiently hard), both distributions also extend significantly to negative values of signed impact parameter significances.

A comparison between the transverse impact parameter distribution with the lifetime sign defined in the transverse plane by “ $\text{sign}_{r\phi}$ ” and by “ sign_{3D} ” is shown in Fig. 6.5. The result is nearly identical for prompt tracks, while for tracks from b -hadron decays the first definition seems to yield a more positive shifted distribution and thus potentially a better performance. This result is a bit surprising and seems to indicate that, since the lifetime sign is not taking into account the different errors on the longitudinal and transverse impact parameters, that the longitudinal component of the lifetime sign is diluting the information from its transverse component.

The extension of the impact parameter significance distribution to high values for K_S^0 decays and conversions are limited by the applied impact parameter cuts.

Summing up all components characterising b -, c - and light jets, the signed impact parameter distributions of Fig. 6.6 are obtained. The component with non-prompt tracks in light jets

¹The same happens in b -jets, as shown later in Fig. 6.5.

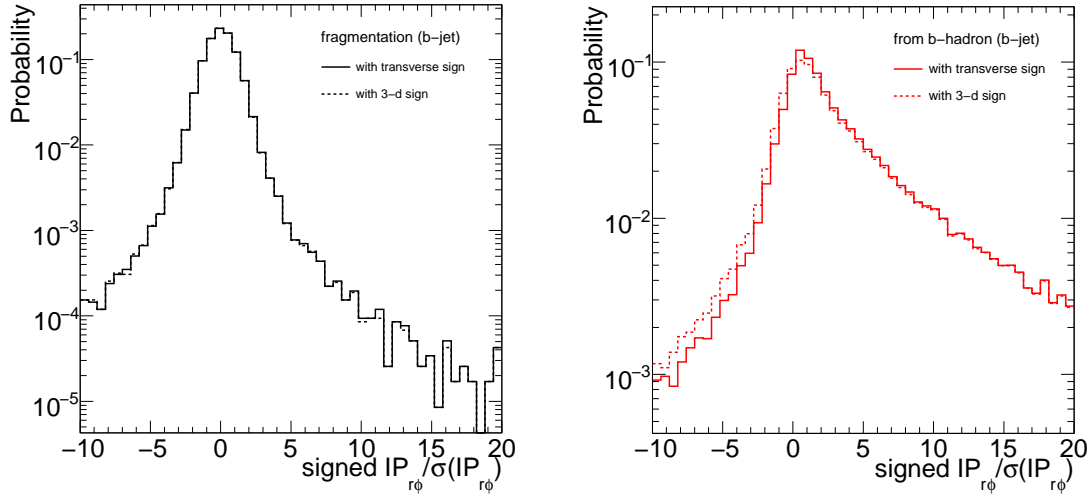


Figure 6.5: Comparison between the transverse impact parameter distribution where the lifetime sign is defined according to the transverse definition (full lines) and to the three-dimensional definition (dashed lines), on the left for prompt tracks and on the right for tracks from the decay of b -hadrons, in both cases in b -jets.

is clearly visible. The component with prompt tracks is characterised by vanishing intrinsic impact parameter, so the distribution of impact parameter significances gives directly access to the resolution function. The result of a simple fit with a Gaussian function in the interval $[-3, 3]$ is shown in Fig. 6.7: the standard deviation is very close to unity, which means that the track parameter errors are correctly estimated by the tracking reconstruction algorithm. A small non-Gaussian tail is also present. An eventual derivation of the resolution function from data is complicated by the presence of non-prompt components, which are present even in light-jets; in principle the resolution function could be obtained from the negative lifetime side of the impact parameter distribution, however this is again not straightforward, since not all long-lived particles, conversions and hadronic interactions with high impact parameter significances in light jets land at positive lifetime values, so that they also contribute with large negative impact parameter significances, contaminating the non-Gaussian tails of the resolution function.

The impact parameter significances of all N tracks associated to the jet to tag need to be combined into a single discriminating variable. It is assumed that the tracks are uncorrelated, so that their probability density functions (PDF) (defined based on the transverse and/or longitudinal impact parameter significance shown in Fig. 6.6) are uniquely defined as a function of the jet flavour. Using a likelihood function defined according to the product of these PDFs, under the hypothesis of uncorrelated tracks, the following likelihood ratio provides the optimal separation, according to the Neyman-Pearson lemma:

$$\text{LR}(IP_1, IP_2, \dots, IP_N) = \frac{\prod_{i=1}^N \text{PDF}_b(IP_i)}{\prod_{i=1}^N \text{PDF}_l(IP_i)} \quad (6.6)$$

For convention the discriminating variable used for b -tagging is then defined as:

$$\text{weight}(IP_1, IP_2, \dots, IP_N) = \log(\text{LR}(IP_1, IP_2, \dots, IP_N)) \quad (6.7)$$

Since the natural logarithm is a continuous monotonic increasing function of its real argument, using the logarithm or the likelihood ratio is equivalent in terms of separation power.

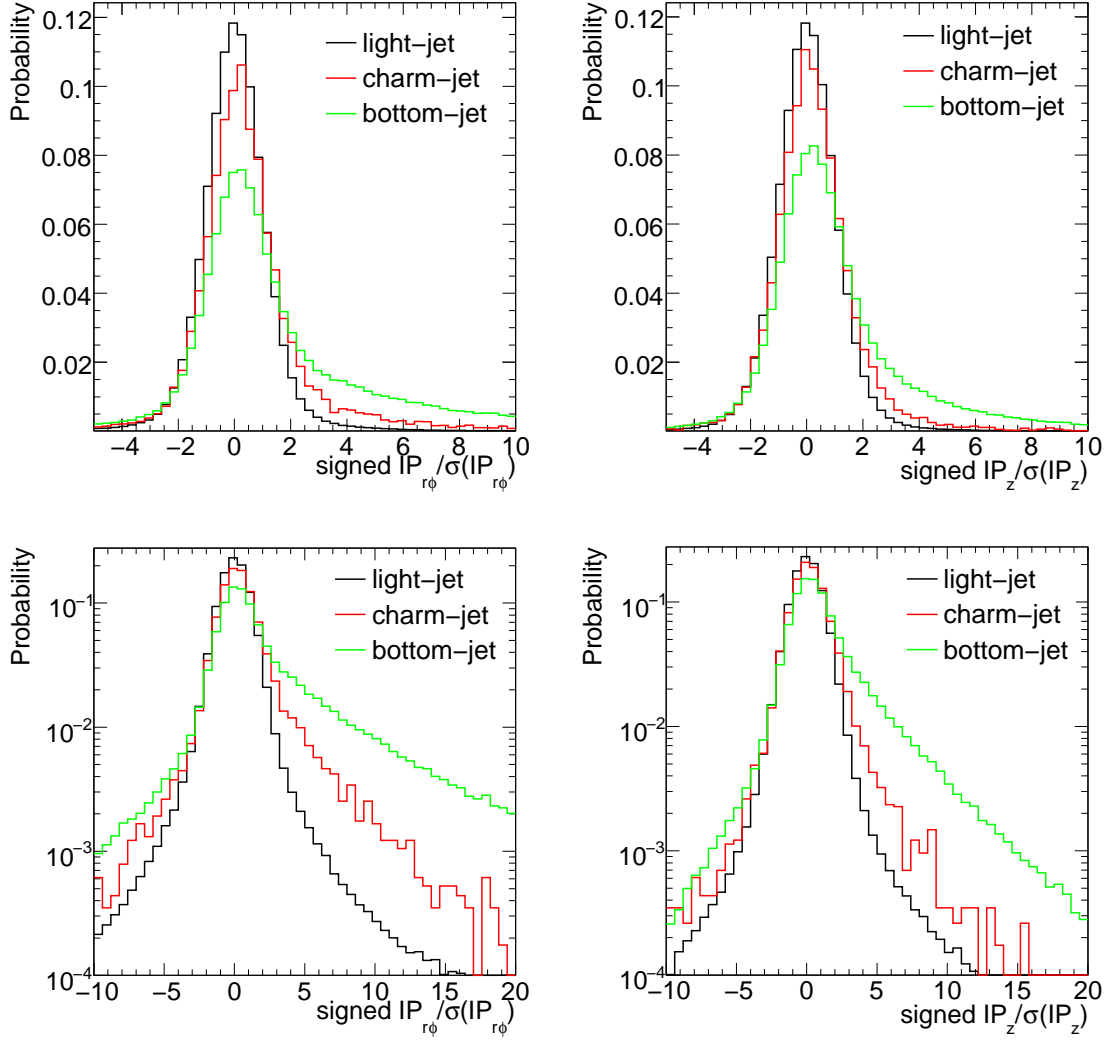


Figure 6.6: Transverse (left) and longitudinal (right) signed impact parameter significance distribution, for tracks selected in b -, c - and light jets, both in linear (top) and logarithmic (bottom) scale.

Using such a formalism, two impact parameter based b -tagging algorithms are constructed, based on the definition of $\text{PDF}(IP_i)$:

1. IP2D: $\text{PDF}(IP_i) = \text{PDF}(IP_{i,r\phi})$
2. IP3D: $\text{PDF}(IP_i) = \text{PDF}(IP_{i,r\phi}, IP_{i,z})$

In the first case the track PDF is one-dimensional, based on the transverse impact parameter significance. In the second case it is a two-dimensional PDF, based on both the transverse and longitudinal impact parameters significance. The distributions defining the two-dimensional PDFs are shown in Fig. 6.8 for b -, c - and light-jets. The non-prompt component in light jets is barely visible, but it nevertheless has an important effect on the b -tagging performance.

The use of a two-dimensional PDF ensures a better performance than using directly the three-dimensional impact parameter significance, which is condensed in the χ^2 function expressing the displacement of the track with respect to the primary vertex position properly

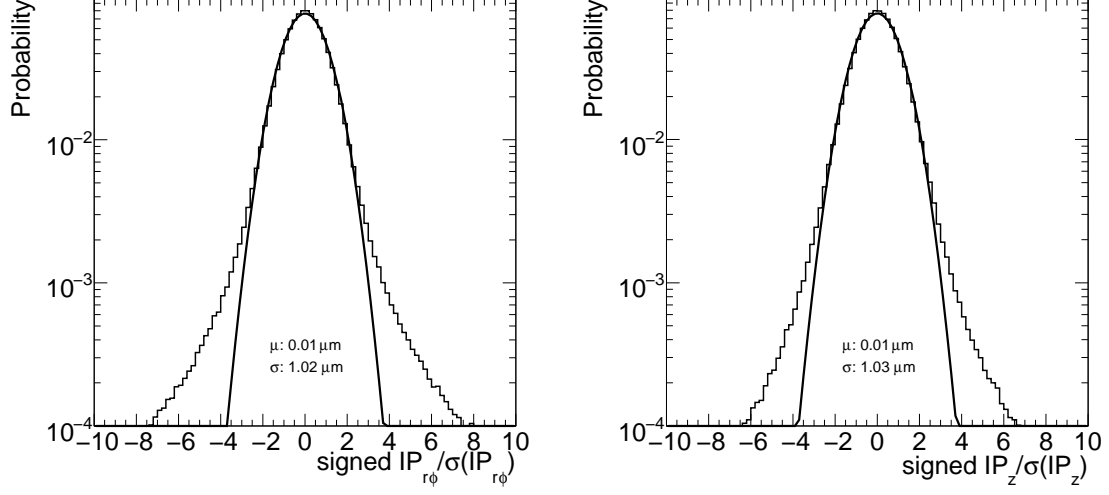


Figure 6.7: Gaussian fit to the transverse (left) and longitudinal (right) impact parameter significance distributions of prompt tracks in light-jets. The statistical errors on the mean and standard deviation of the Gaussian function are negligible.

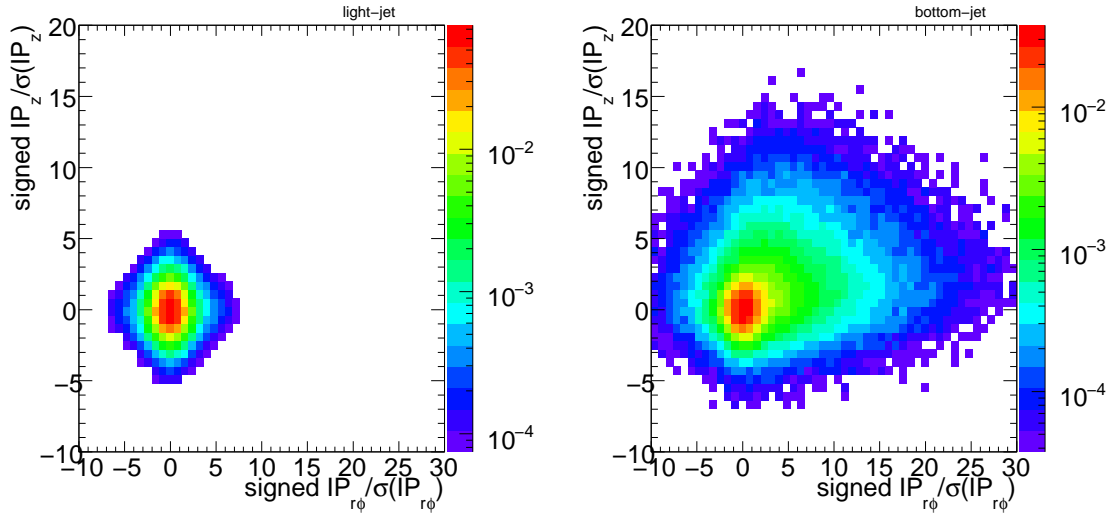


Figure 6.8: Two-dimensional distribution of transverse and longitudinal impact parameters significance in light-jets (left) and b -jets (right). All plots are normalised to unity (along the z axis) and in logarithmic scale.

divided by its error. This last quantity is equivalent to the contribution of a single track to the χ^2 function during a vertex fit, already seen in Eq. 5.5:

$$\chi_i^2(r_{PV}) = (\vec{q}_{pred,i}(\vec{r}_{PV}) - \vec{q}_{meas,i})^T \text{COV}_i^{-1} (\vec{q}_{pred,i}(\vec{r}_{PV}) - \vec{q}_{meas,i}), \quad (6.8)$$

where the helix parameters of the track \vec{q} can be limited, for simplicity, to the parameters d_0 and z_0 . The residual of the measured helix parameters d_0 and z_0 is considered with respect to the values which would correspond to the track passing exactly through the reconstructed primary vertex, weighted by the inverse of the track covariance matrix in the helix reference frame. The covariance matrix takes into account also the correlation between the d_0 and z_0 helix parameters of the track.

However, the direct use of the track to vertex χ^2 would yield a worse performance than the use of a two-dimensional PDF. A PDF based on a one-dimensional distribution can yield the same performance of a two-dimensional distribution only if the two variables of the two-dimensional PDF can be combined in such a way that no information is lost in the combination. The distribution of prompt tracks in light-jets is dominated by resolution effects and the transverse and longitudinal impact parameter significances can be thus optimally combined in a three-dimensional impact parameter, based on the χ^2 defined in Eq. 6.8². The same does not hold for the impact parameter significance distribution for tracks from b - and c -hadron decays, which is significantly different in the transverse and longitudinal planes, due mainly to the fact that the resolution is significantly better in the transverse direction, but also because of the effect of the magnetic field bending charged particles in the transverse plane. In addition, some of the non-prompt components in light-jets have a distinctive two-dimensional distribution, which would be diluted by using the three-dimensional impact parameter. This is for example clearly the case for conversions: in this case, using the two-dimensional PDFs, the likelihood function implicitly contains the information that conversions are more likely to be found at high transverse impact parameter significances and low longitudinal impact parameter significances and, as a consequence, the b over light-jet likelihood ratio will be decreased in these cases, slightly decreasing the fraction of light-jet misidentified as b -jets due to conversions.

The final discriminating variable (defined in ATLAS as jet-weight) for the two impact parameter based algorithms, IP2D and IP3D is shown in Fig. 6.9. It can be clearly seen that IP3D provides the most powerful discrimination of b -jets against light-jets. The discrimination against c -jet is poorer: this could be improved a bit by explicitly considering PDFs for c -jet when building up the likelihood ratio.

The performance of the impact parameter based b -tagging algorithms can be improved by using some information directly from the secondary vertex based vertexing algorithms in two aspects. In first instance the vertices corresponding to the decay of long lived particles (like K_S^0 mesons and Λ baryons) or conversions or hadronic interactions can be explicitly reconstructed, at least as long as both charged particles at the decay or conversion vertex (or at least two charged particle tracks from the interaction vertex in the case of hadronic interactions) have been reconstructed as tracks in the inner detector. Tracks associated to such vertices can be removed from the tracks considered for the impact parameter based algorithms. In second instance, the direction between the secondary and primary vertex positions can be used to improve the reliability of the lifetime sign defined in Eq. 6.3, substituting \vec{p}_{jet} with $\vec{r}_{SV} - \vec{r}_{PV}$. The latter provides a significantly improved estimation of the b -hadron direction. Both kinds of information improve slightly the performance of the impact parameter based b -tagging algorithms.

²This is only correct under the hypothesis that the non-Gaussian tails are the same in the transverse and longitudinal impact parameter significance distributions.

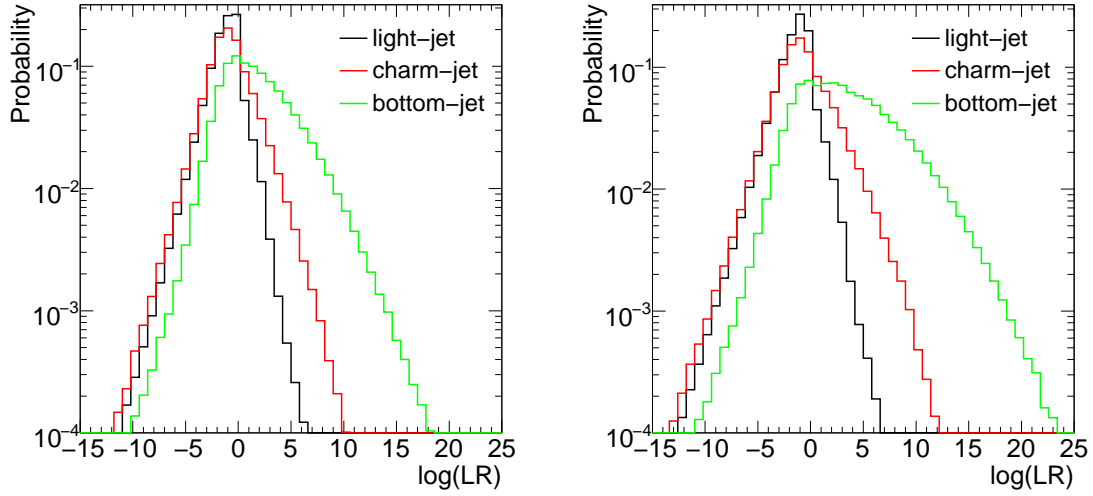


Figure 6.9: Distributions for the discriminating variable ($\log(\text{LR})$) for b -, c - and light-jets, on the left for IP2D and on the right for IP3D.

The impact parameter based algorithms permit to obtain a very good b -tagging performance, as will be analysed at the end of this chapter, and at the same time allow to keep the method fairly simple, with a single track-based PDF to be calibrated for each quark flavour. One of the reasons for the good performance is that the likelihood method based on the impact parameter significance contains implicitly, as a prior knowledge in the PDF for b -jets, the fraction of tracks expected to arise from fragmentation and the fraction expected to arise from b - and c -hadron decays. However, this method has also few drawbacks. One is that the fraction of tracks from fragmentation depends on the energy of the initial b -quark, which is at the moment not accounted for, but could be cured by introducing PDFs with a conditional dependence on the jet transverse momentum. The second is that, while the impact parameter significance of prompt tracks is independent on the jet p_T and η , the impact parameter significance of tracks from secondary decays is not: in fact, while the impact parameter as such is nearly invariant under Lorentz boosts of the b -hadron, the error decreases with increasing track p_T and smaller η (as shown in Fig. 4.2). The track-based PDF is thus not invariant neither under a boost of the b -quark nor under a boost of its emerging charged particles tracks. This can be cured both by making the track-PDF category dependent, where the categories corresponds to different interval either in the jet kinematics or in the track errors. Using a category dependent PDF (but only according to the track p_T) is an approach that was tried out in ATLAS in the last two years: it brings a significant improvement in performance, but has the disadvantage of increasing the number of free parameters of the likelihood model and making an eventual calibration on data more difficult (more information is contained in Ref. [44]). An alternative approach could be based on a likelihood model expressing the convolution of the expected, boost invariant, d_0 and z_0 distribution with the track resolution function based on σ_{d_0} and σ_{z_0} . This would have the advantage of separating the physics dependent variables (d_0 and z_0) from the experimental resolution effects, which would be condensed in the resolution function.

6.4 Secondary vertex reconstruction in b -jets

The explicit reconstruction of secondary vertices in b -jets can significantly improve the b -tagging performance of the impact parameter based algorithms.

As described in Section 6.1, the typical topology of particle decays in a b -jet is a decay chain with two vertices, one stemming from the b -hadron decay and at least one from c -hadron decays. An exclusive reconstruction of the huge number of different possible b -decay modes cannot be performed with high efficiency, for several reasons: many of these decay modes involve neutral particles, which are difficult to reconstruct cleanly in a hadron collider environment and cannot be used directly in the vertex reconstruction³, and the set of selection cuts needed to reconstruct all the different decay modes would severely limit the reconstruction efficiency. The reconstruction of secondary b - and c -hadron decay vertices in jets thus has to be done in an inclusive way, where the number of charged particle tracks originating from b - and c -hadron decays is not known a-priori.

Trying to resolve the b - and c -hadron vertices of the decay cascade is challenging for the following reasons:

- The probability to have at least two reconstructed charged particle tracks both from the b - and c -hadron decays is limited (precise numbers will be evaluated in Section 6.8). Several reasons contribute to this: the small charged particle multiplicities involved in these decays, the fraction of charged track particles from b - and c -hadron decays which happen to be compatible with the primary vertex and are therefore not considered for the reconstruction of secondary vertices and the limited track reconstruction efficiency (caused mainly by material interactions in the detector).
- The resolutions of the relevant track parameters, especially at low transverse momenta, are not sufficient to separate the two vertices efficiently.

Two strategies to detect a secondary decay vertex in b -jets are available in ATLAS. The first one is based on a the fit of a single geometrical vertex. Even if this hypothesis is not correct, given the reasons given above, this approximation works well for a large fraction of cases. The second algorithm is based on a kinematic approach, which assumes that the primary event vertex and the b - and c -hadron decay vertices lie approximately on the same line, the flight path of the b -hadron. While the first approach was already used in ATLAS at the time when this thesis work started and will be therefore only briefly described here, the second is a direct result of this thesis work and will be described in greater detail.

6.5 Inclusive secondary vertex reconstruction

The inclusive fit of a single displaced vertex in b -jets is based on the VKalVrt [70] vertex reconstruction package, which contains an implementation of the Billoir method for vertex fitting described in Section 5.3.1.

The main idea of this algorithm is to maximise the b/c -hadron vertex detection efficiency, keeping at the same time the probability to find a vertex inside light jets low. The algorithm starts with all tracks associated to the jet and passing the *loose* selection cuts introduced in Section 6.2. The vertex search starts with looking for all track pairs and trying to form a two-track vertex which satisfies the following conditions:

³Neutral particles can be used only to constraint the particle momenta at the decay vertex, which is useful only if all particles at the vertex are reconstructed, so that four momentum conservation can be exploited.

- the two tracks form a vertex with vertex probability $\text{Prob}(\chi_{vxt}^2)$ above 3.5%;
- each track of the pair must have a three-dimensional impact parameter significance with respect to the primary vertex (defined in Eq. 6.8) larger than 2σ and the sum of these two significances must be larger than 6σ ;
- to reduce the influence of badly measured tracks, the two-tracks vertices are required to be produced in the direction of flight of the b -quark, by requiring the scalar product $(\vec{r}_{2tr} - \vec{r}_{primary}) \cdot \vec{p}_{jet}$ to be positive.

Some of the reconstructed two-track vertices stem from K_s^0 and Λ^0 decays, $\gamma \rightarrow e^+e^-$ conversions and hadronic interactions in the detector material. Charged particle tracks coming from such vertices are not considered for the inclusive b -decay vertex fit. All tracks corresponding to the remaining accepted two-track vertices are used to determine a single secondary vertex. If the resulting vertex has a very small vertex probability (corresponding to a low $\text{Prob}(\chi_{vtx}^2)$), the track with the highest contribution to the vertex χ^2 is removed from the vertex fit and the vertex fit is repeated. This procedure is iterated until $\text{Prob}(\chi_{vtx}^2) > 0.1\%$.

The result of this procedure is the (eventual) presence of a vertex, its position and covariance matrix and the list of its associated charged particles, whose properties are then explored in a dedicated b -tagging algorithm.

More details about this vertex reconstruction algorithm can be found in Ref. [44].

6.6 The JetFitter vertex reconstruction algorithm

JetFitter is an inclusive secondary vertex reconstruction algorithm which exploits the topological structure of weak b - and c -hadron decays inside a jet.

6.6.1 Reconstruction of the decay chain

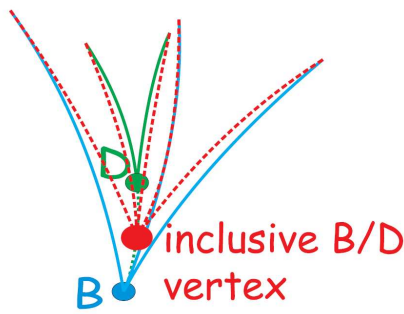


Figure 6.10: The inclusive secondary vertex reconstruction algorithm fits all displaced tracks to an inclusive vertex.

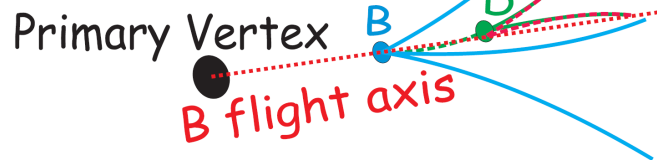


Figure 6.11: JetFitter performs a multi-vertex fit using the b -hadron flight direction as constraint.

As opposed to the algorithm described in the last section, in which displaced tracks are selected and an inclusive single vertex is obtained using a Billoir based χ^2 fit (as sketched in Figure 6.10), the algorithm described here, called *JetFitter*, is based on a different hypothesis. It assumes that the b - and c -hadron decay vertices lie on the same line defined through the b -hadron flight path. All charged particle tracks stemming from either the b - or c -hadron decay thus intersect this b -hadron flight axis. There are several advantages to this method:

- Incomplete topologies can also be reconstructed (in principle even the topology with a single track from the b -hadron decay and a single track from the c -hadron decay is accessible).
- The fit evaluates the compatibility of the given set of tracks with a b - c -hadron like cascade topology, increasing the discrimination power against light quark jets.
- Constraining the tracks to lie on the b -hadron flight axis reduces the degrees of freedom of the fit, increasing the chance to separate the b/c -hadron vertices.

From the physics point of view this hypothesis is justified through the kinematics of the particles involved as defined through the hard b -quark fragmentation function and the masses of b - and c -hadrons. The lateral displacement of the c -hadron decay vertex with respect to the b -hadron flight path is small enough not to violate significantly the basic assumption within the typical resolutions of the tracking detector.

This assumption was analysed in more detail on Monte Carlo simulated events of the WH process, with a Higgs boson ($m_H = 120$ GeV) decaying into a pair of b -quarks, looking for b - and c -hadrons in b -jets with transverse momenta above 15 GeV and within the acceptance of the tracking device. Considering the average b -quark transverse momentum of ≈ 60 GeV, the average transverse flight length of b -hadrons turns out to be 4.3 mm, while the subsequent transverse flight length of the c -hadrons produced at the b -vertex is on average 1.9 mm. As a result the overall average flight length is ≈ 6.2 mm. The average lateral displacement of the c -hadron with respect to the b -flight direction, projected onto the transverse plane is ≈ 30 μm : this displacement is certainly small with respect to the overall $PV \rightarrow b \rightarrow c$ -hadron flight length.

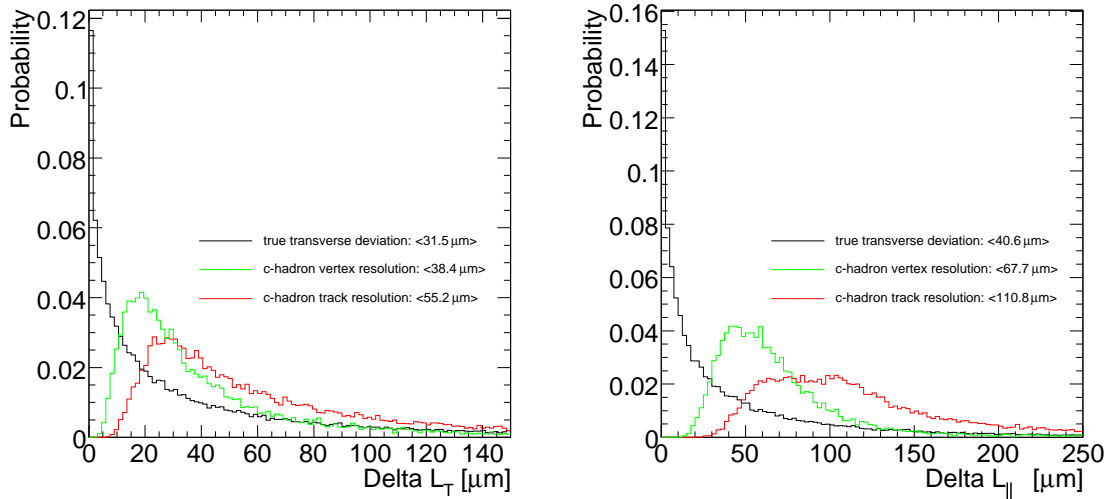


Figure 6.12: Average lateral displacement of the c -hadron decay vertex with respect to the b -hadron flight direction, in the transverse (left) and longitudinal (right) planes, compared to the respective experimental resolution either of a single track or of the vertex resulting from the vertex fit of the tracks originating from the d -hadron vertex.

However, in order for such an assumption not to cause any bias on the vertex fit χ^2 , this displacement must be small with respect to the involved experimental resolutions. This comparison is shown in Fig. 6.12, where the real displacement in the transverse and longitudinal planes is compared to the typical experimental resolutions achievable by considering the tracks

originating from the c -hadron decay vertex. This corresponds to the impact parameter resolution of a single track, if - after selection and removal of tracks compatible with the primary vertex - only one track from the c -hadron is left, or to the resolution in the direction of interest⁴ of a vertex fit applied only to the tracks originating from the c -hadron decay vertex. In the longitudinal plane a real displacement of on average $\approx 40 \mu\text{m}$ has to be compared with a single track resolution of $\approx 110 \mu\text{m}$ or, when more than one track is available, with a single vertex resolution of $\approx 70 \mu\text{m}$, so that the bias on the $\text{PV} \rightarrow b \rightarrow c$ -hadron decay chain fit can be considered small. In the transverse plane the situation is more delicate, with a true average displacement of $\approx 30 \mu\text{m}$, to be compared with a single track resolution of $\approx 55 \mu\text{m}$ and, when available, with a single vertex resolution of $\approx 40 \mu\text{m}$. In addition, the real displacement both in the transverse and longitudinal projections have a tail to very high values.

The assumption adopted in JetFitter is thus well justified by the small average displacement of the c -hadron with respect to the b -hadron momentum ($30 \mu\text{m} \ll 6 \text{ mm}$), however the track resolution on the transverse plane can occasionally be better than such a displacement and therefore can cause some bias on the overall fit χ^2 . This must be either corrected for by inflating artificially the width of the flight axis in correspondence of the points where the b - and c -hadron vertices are expected to lie during the fit, or, more simply, it must be taken into account when removing configurations which are incompatible with the expected decay chain topology, by making the requirement on the overall $\text{Prob}(\chi^2)$ of the decay chain vertex fit looser. The second solution was adopted in JetFitter.

The hypothesis of a negligible lateral displacement of the c -hadron decay vertex with respect to the b -hadron flight path was explored for the first time in the *ghost track algorithm* developed by the SLD Collaboration [74], where the already defined b -hadron flight axis is substituted by a *ghost track* and where a numerical global χ^2 minimisation procedure was used to perform the multi-vertex fit. The use of a dedicated Kalman filter has several advantages compared to such a conventional global χ^2 minimisation.

6.6.2 Vertex fit: formalism and implementation

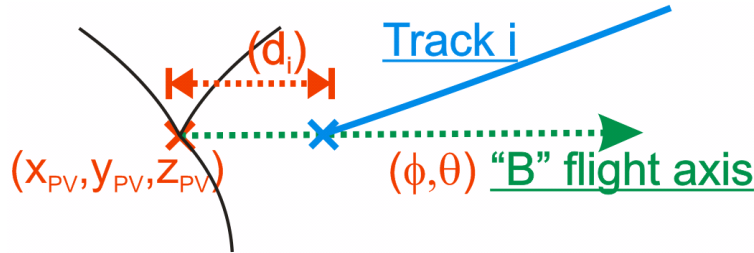


Figure 6.13: The JetFitter algorithm iteratively updates the primary vertex position, the b -hadron flight direction and the intersections of this axis with the fitted tracks.

In JetFitter the vertexing task is mathematically implemented as an extension of the Kalman Filter formalism for vertex reconstruction, which was introduced in Section 5.3.2. While in a conventional Kalman vertex fit the variables to be iteratively updated in the fit are the vertex position \vec{x} , in JetFitter the decay chain is described through the following variables (see Fig. 6.13):

$$\vec{d} = (x_{PV}, y_{PV}, z_{PV}, \phi, \theta, d_1, d_2, \dots, d_N), \quad (6.9)$$

⁴The direction of interest is defined to be perpendicular to the direction of flight of the b -hadron and its transverse and longitudinal components are separately considered.

with:

- (x_{PV}, y_{PV}, z_{PV}) : the primary vertex position.
- (ϕ, θ) : the azimuthal and polar directions of the b -hadron flight axis.
- (d_1, d_2, \dots, d_N) : the distances of the fitted vertices, defined as the intersections of one or more tracks and the b -hadron flight axis, to the primary vertex position along the flight axis (N representing the number of vertices).

Before starting the fit, the variables are initialised with their prior knowledge:

- The primary vertex position (with covariance matrix), as provided by the primary vertex finding algorithm.
- The b -hadron flight direction, approximated by the direction of the jet axis, the error being provided by the expected combined effect of the jet direction resolution and the average displacement of the jet direction relative to the b -hadron flight axis, as determined from Monte Carlo simulations and as described in the following.

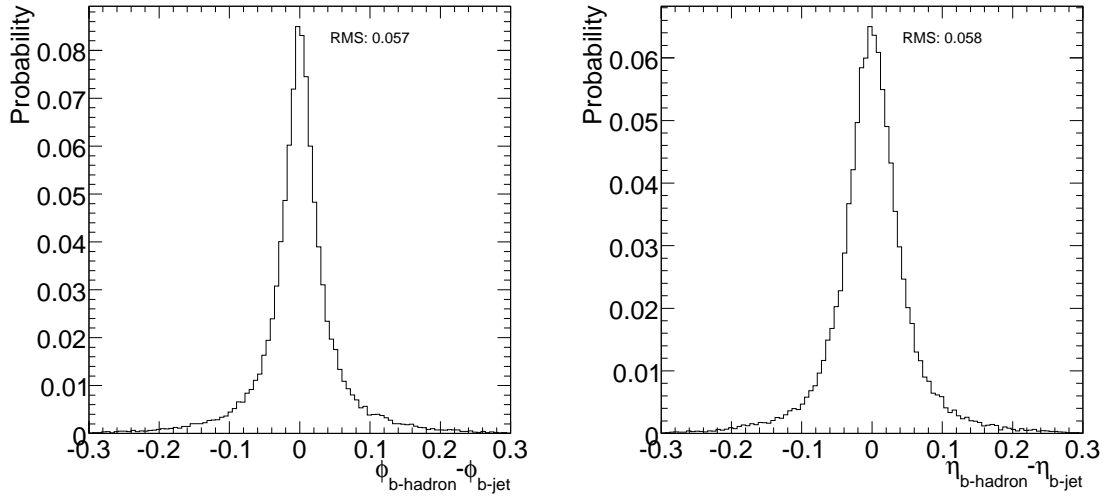


Figure 6.14: Residual distribution of the reconstructed jet direction with respect to the b -hadron direction in b -jets, separately in the azimuthal angle (left) and in the pseudo-rapidity variable (right).

The distribution of residuals of the reconstructed jet direction with respect to the real b -hadron direction in both transverse and longitudinal direction, as obtained from $WH \rightarrow \ell \nu b \bar{b}$ fully simulated events, is shown in Fig. 6.14. Instead of the polar angle, the pseudo-rapidity is chosen for determining the width of the polar angle distribution. Both the root mean squared residuals in $\Delta\phi$ and in $\Delta\eta$ are ≈ 0.06 . In order to provide only a loose constraint in the fit and not to cause any significant bias due to the non-Gaussian tails of the distributions, the errors on the b -flight axis direction during the fit are set respectively to 0.07 and 0.065. The error on η is translated into an error on θ by simple linear error propagation.

The fit is then performed by adding iteratively as input the track parameters of the individual tracks, specifying at each iteration the vertex number to be updated.

As opposed to the measurement equation of a single-vertex Kalman vertex fit of Eq. 5.7, which for simplicity will be repeated here:

$$\begin{aligned}
 \vec{q} &= \vec{q}_0 + A[\vec{x} - \vec{x}_0] + B[\vec{p} - \vec{p}_0] = \\
 &= \vec{q}(\vec{x}, \vec{p}) = \vec{q}(\vec{x}, \vec{p})_{\vec{x}=\vec{x}_0, \vec{p}=\vec{p}_0} + \left. \frac{\delta \vec{q}(\vec{x}, \vec{p})}{\delta \vec{x}} \right|_{\vec{x}=\vec{x}_0} (\vec{x} - \vec{x}_0) + \left. \frac{\delta \vec{q}(\vec{x}, \vec{p})}{\delta \vec{p}} \right|_{\vec{p}=\vec{p}_0} (\vec{p} - \vec{p}_0) \\
 &= \vec{C} + A\vec{x} + B\vec{p},
 \end{aligned} \tag{6.10}$$

the new measurement equation to be used to update the fit variables is:

$$\begin{aligned}
 \vec{q} &= \vec{q}(\vec{d}, \vec{p}) = \vec{q}(\vec{d}, \vec{p})_{\vec{d}=\vec{d}_0, \vec{p}=\vec{p}_0} + \left. \frac{\delta \vec{q}(\vec{d}, \vec{p})}{\delta \vec{d}} \right|_{\vec{d}=\vec{d}_0} (\vec{d} - \vec{d}_0) + \left. \frac{\delta \vec{q}(\vec{d}, \vec{p})}{\delta \vec{p}} \right|_{\vec{p}=\vec{p}_0} (\vec{p} - \vec{p}_0) \\
 &= \hat{\vec{C}} + \hat{A}\vec{d} + \hat{B}\vec{p}
 \end{aligned} \tag{6.11}$$

where the linearisation has now to be performed around some initial \vec{d}_0 .

Before the first sequence of Kalman iterations is performed, N is initialised to the number of tracks, since during the first stage only single-track vertices are considered. A first approximate estimation of \vec{d}_0 is provided by the points along the flight axis (as estimated from the jet direction and having as origin the primary vertex) corresponding to the minimum distance to the single tracks. Since this task cannot be solved analytically, this minimum distance is obtained through a *Newton* based iterative method, which minimises the distance between a straight line (the b -hadron flight axis) and a helix (the track). This is essentially analogous to the minimisation of the distance between two charged particle tracks⁰ presented in Section 5.4.2 and condensed in Eq. 5.49, with the difference that the equation of motion of the second of the two charged particle tracks is substituted with the equation of motion of a straight track.

Defining as $\vec{x} = (x, y, z)$ the position of the vertex on the b -hadron flight axis to be updated by the track added to the fit, the relation between \vec{x} and \vec{d} is:

$$\begin{aligned}
 x &= x_{PV} + d_i \cdot \sin(\theta) \cos(\phi) \\
 y &= y_{PV} + d_i \cdot \sin(\theta) \sin(\phi) \\
 z &= z_{PV} + d_i \cdot \cos(\theta),
 \end{aligned} \tag{6.12}$$

where d_i is the flight length corresponding to the vertex considered, so that the measurement equations (6.10) and (6.11) are identically equal, once the equality

$$\vec{x} - \vec{x}_0 = \vec{x}(\vec{d}) - \vec{x}(\vec{d}_0) = \left. \frac{d\vec{x}(\vec{d})}{d\vec{d}} \right|_{\vec{d}=\vec{d}_0} (\vec{d} - \vec{d}_0) \tag{6.13}$$

is used.

While $\hat{\vec{C}} = \vec{C}$ and $\hat{B} = B$, the position Jacobian \hat{A} to be used in the fit is different from the single-vertex fit one of Eq. 6.10, but can simply be obtained by computing $\hat{A} = A \left. \frac{d\vec{x}(\vec{d})}{d\vec{d}} \right|_{\vec{d}=\vec{d}_0}$, where the transformation matrix $\frac{d\vec{x}(\vec{d})}{d\vec{d}}$ is:

$$\frac{d\vec{x}(\vec{d})}{d\vec{d}} = \begin{pmatrix} 1 & 0 & 0 & -d_i \sin(\theta) \cos(\phi) & d_i \cos(\theta) \cos(\phi) & \dots & \sin(\theta) \cos(\phi) & \dots \\ 0 & 1 & 0 & d_i \sin(\theta) \sin(\phi) & d_i \cos(\theta) \sin(\phi) & \dots & \sin(\theta) \sin(\phi) & \dots \\ 0 & 0 & 1 & 0 & -d_i \sin(\theta) & \dots & \cos(\theta) & \dots \end{pmatrix} \tag{6.14}$$

and the columns number $(6, \dots, i+4, i+6, \dots, N+5)$ related to the vertices not considered in the current iteration $(d_1, d_2, \dots, d_{i-1}, d_{i+1}, \dots, d_N)$ are identically zero.

While the original Jacobian A is usually quite insensitive to small changes of the position of the vertex \vec{x} (the linear approximation for a track around a vertex holds quite well), the transformation matrix and thus the Jacobian \hat{A} has a highly non-linear dependence on the distance from the primary vertex d_i of the vertex formed by one or more intersecting tracks.

In order for the fit to converge, more cycles of iterations will be needed due to the non-linearity of the measurement equation and during each new iteration the position Jacobian \hat{A} needs to be recomputed. However, this operation can be considerably sped up by re-linearising the measurement equation (6.11) only in the part of \hat{A} which involves the transformation matrix around the new linearisation position \vec{d}_1 , while keeping the Jacobian A at the old linearisation position \vec{d}_0 :

$$\begin{aligned} \vec{q} = & \vec{q}(\vec{x}(\vec{d}), \vec{p})_{\vec{d}=\vec{d}_0, \vec{p}=\vec{p}_0} + \left. \frac{\delta \vec{q}(\vec{x}(\vec{d}), \vec{x})}{\delta \vec{x}} \right|_{\vec{x}=\vec{x}(\vec{d}_0)} [\vec{x}(\vec{d}) - \vec{x}(\vec{d}_1) + \vec{x}(\vec{d}_1) - \vec{x}(\vec{d}_0)] \\ & + \left. \frac{\delta d \vec{q}(\vec{d}, \vec{x})}{\delta \vec{p}} \right|_{\vec{p}=\vec{p}_0} (\vec{p} - \vec{p}_0) = \hat{\vec{C}} + \hat{A} \vec{d} + B \vec{p} \end{aligned} \quad (6.15)$$

where the transformation matrix can be applied directly at the new linearisation point \vec{d}_1 , to relate $\vec{x}(\vec{d}) - \vec{x}(\vec{d}_1)$ to $\vec{d} - \vec{d}_1$:

$$\vec{x}(\vec{d}) - \vec{x}(\vec{d}_1) = \left. \frac{d\vec{x}}{d\vec{d}} \right|_{\vec{d}=\vec{d}_1} (\vec{d} - \vec{d}_1) \quad (6.16)$$

Comparing Eq. (6.15) with Eq. 6.10, the coefficients $\hat{\vec{C}}$ and the Jacobian \hat{A} can be obtained from the coefficients \vec{C} and the Jacobian A :

$$\begin{aligned} \hat{\vec{C}} = \vec{C} + A \left[\vec{x}_1 - \left. \frac{d\vec{x}}{d\vec{d}} \right|_{\vec{d}=\vec{d}_1} \vec{d}_1 \right] \\ \hat{A} = A \left. \frac{d\vec{x}}{d\vec{d}} \right|_{\vec{d}=\vec{d}_1} \end{aligned} \quad (6.17)$$

The prescription for a fast linearisation is thus:

1. Linearise the measurement equation (Eq. (6.10)), computing \vec{C} , A and B .
2. At each new iteration typically only recompute the transformation matrix (Eq. (6.14)) and obtain the new $\hat{\vec{C}}$ and \hat{A} to be used in the Kalman update step (Eq. (6.17)).

Once the measurement equation for all tracks to be added iteratively to the fit is determined, the Kalman filter based vertex fit is fully defined. In a typical decay chain fit around 4-5 iterations are required before the fit converges: the convergence criterion is that, after a new iteration, $|\chi_{new}^2 - \chi_{old}^2|$ is required to be smaller than 0.001.

The fit procedure remains perfectly defined also if more than one track is fitted to the same vertex i (out of N) along the flight axis. In this case, when the tracks corresponding to this vertex will be added iteratively to the decay chain fit, they will both contribute the same flight length d_i corresponding to the vertex to be updated.

So, once the assignment of tracks to vertices is made, the vertex fitting procedure can be started and yields the optimal values for the b -hadron flight axis direction and positions of the fitted vertices along the flight axis.

Vertex probability estimation

The clustering of tracks into vertices, as in the case of primary vertex finding, is in first instance a separate problem, and is dealt with by the vertex finding algorithm, which will be described in Section 6.6.3. This interacts with the vertex fitting algorithm in order to try to converge to the correct topology. Two of the fundamental ingredients for the clustering, which the vertex fitting algorithm needs to provide, are:

- The probability of a certain vertex to be compatible with the fitted decay chain.
- The probability for two initially separated vertices along the flight axis to be compatible with a single vertex and with the overall new resulting decay chain.

The first quantity can be obtained by applying the *smoothing* procedure. While in the standard Kalman based fit of a single vertex the *smoothing* procedure provides the correct χ^2 contribution of a track to the fitted vertex, in *JetFitter* the corresponding quantity is the χ^2 contribution of a certain track on the b -hadron flight axis to the fit of the whole decay chain. However, when more than one track is fitted to a vertex along the b -flight axis, the quantity of interest is the overall χ^2 contribution of that vertex to the whole decay chain fit. In this case the *smoothing* procedure needs to be slightly modified, as described in the following steps:

- the final fitted decay chain is taken and the χ_{smooth}^2 value is set to 0;
- all tracks belonging to the chosen vertex on the b -hadron flight axis are removed iteratively from the decay chain and their χ^2 contributions are added up iteratively to χ_{smooth}^2 ;
- the new decay chain is compared with the final fitted one and the difference between the final and new configuration (in terms of $[\vec{d}_{\text{new}} - \vec{d}_{\text{fin}}]^T \text{COV}_{\vec{d}}^{-1} [\vec{d}_{\text{new}} - \vec{d}_{\text{fin}}]$, where $\text{COV}_{\vec{d}}^{-1}$ is the weight matrix of the final decay chain) is considered as an additional contribution to χ_{smooth}^2 .

The obtained χ_{smooth}^2 has then $2 \cdot N_t - 1$ degrees of freedom⁵, where N_t is the number of tracks involved in the vertex considered.

The second quantity, the probability of two separate vertices along the jet axis to be compatible with a single vertex and at the same time with the overall decay chain, is a bit more difficult to get, at least if the correlations of the positions of the vertices with the direction of the b -hadron flight axis are correctly accounted for. The simplest option is to perform a new complete decay chain fit with the two vertices merged together: the χ^2 probability of the merged vertex, computed according to the smoothing procedure already described, is then used as a measure of how likely the merging hypothesis is correct. This approach is simple, however, since this kind of probability estimation can be iterated many times during the vertex finding, repeating the full fit at each iteration is extremely slow.

In order to avoid to repeat the complete fit frequently, an alternative approach has been set up, in which the hypothesis of merging two vertices is tested adding on top of the previous complete fit a single Kalman update step, where the constraint $d_i = d_j$ (for merging two vertices on the b -hadron flight axis) or $d_i = 0$ (to merge vertex i with the primary vertex) is applied. This corresponds formally to adding a new measurement to the vertex fit, with zero

⁵The number of degrees of freedom of a vertex in the decay chain fit is $2 \cdot N_t - 1$, instead of $2 \cdot N_t - 3$ of a conventional single vertex fit, since only a one-dimensional variable is needed to describe the vertex position along the b -flight axis.

associated error. As discussed in Section 5.3.2, such a *degenerate* measurement can be added to a Kalman filter based fit by switching from the Kalman *weighted mean* formalism to the *gain* formalism.

Considering the case of two vertices i and j to be merged into a single one, the measurement equation corresponding to the constraint is trivial:

$$\begin{aligned} m &= f(\vec{d}) \\ &= d_i - d_j \end{aligned} \quad (6.18)$$

while the value of the measurement m is zero with zero associated variance $\text{VAR}(m)$. This measurement equation is already linear in the variables $\vec{d} = (x_{PV}, y_{PV}, z_{PV}, \phi, \theta, d_1, d_2, \dots, d_N)$, to be updated. Taking $\vec{d} = \vec{0}$ as the linearisation point, the Jacobian and the constant terms are then simply:

$$\begin{aligned} m &= f(\vec{d})_{\vec{d}=\vec{0}} + \left. \frac{\delta f(\vec{d})}{\delta \vec{d}} \right|_{\vec{d}=\vec{d}_0} (\vec{d} - \vec{d}_0) \\ &= C_{\text{constr}} + A_{\text{constr}} \cdot \vec{d} \\ &= 0 + \begin{pmatrix} 0 & 1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \dots \\ d_i \\ \dots \\ d_j \\ \dots \end{pmatrix} \end{aligned} \quad (6.19)$$

The measurement m , its variance $\text{VAR}(m)$, the constant term C_{constr} and the Jacobian matrix A_{constr} uniquely define the Kalman update step. After the constraint is applied to the fit, the vertex probability of the new joint vertex is estimated. Instead of applying the smoothing procedure on the new vertex, it is possible to deduce the $\chi^2_{\text{smooth}}[\text{new vtx}]$ contribution based on the vertex $\chi^2_{\text{smooth}}[\text{vtx } 1/2]$ of the two vertices that have been merged and the $\chi^2_{\text{fit}}[\text{before/after}]$ of the decay chain before and after merging:

$$\chi^2_{\text{smooth}}[\text{new vtx}] = \chi^2_{\text{smooth}}[\text{vtx } 1] + \chi^2_{\text{smooth}}[\text{vtx } 2] + \chi^2_{\text{fit}}[\text{after}] - \chi^2_{\text{fit}}[\text{before}], \quad (6.20)$$

where the number of degrees of freedom of the new merged vertex is the sum of the degrees of freedom of the two original vertices, plus one (corresponding to applying the one-dimensional constraint). This avoids the need of computing any new explicit smoothing step.

Negative flight lengths

When the fit is applied to tracks belonging to a real b -jet, it may sometimes happen that the b -hadron flight axis converges to a non physical direction pointing opposite to the real b -hadron flight direction, on which most or all of the fitted vertices show negative flight lengths.

The following solution has been implemented to avoid this problem:

1. A first fit is performed until convergence is reached.
2. The fit is iteratively repeated under the hypothesis that the re-linearisation of all tracks with negative flight length d_i is performed as if the tracks would have had opposite (positive) flight lengths.

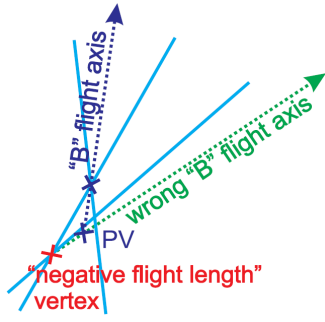


Figure 6.15: The fit can end up forming a vertex behind the primary interaction vertex, with a negative flight length, which represent a local minimum for the B flight axis direction

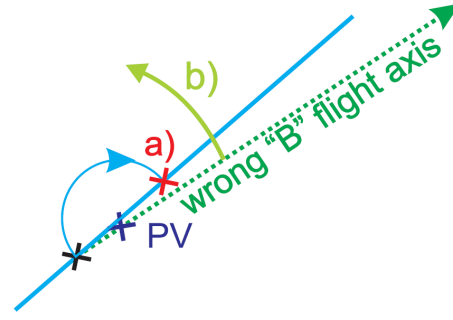


Figure 6.16: Re-linearising the *negative flight length* tracks with the opposite flight length (a) results in a global attempt to move the fit out of the wrong minimum (b).

3. Fit iterations are repeated until χ^2 convergence is reached (which also means that no more tracks have been swapped between positive and negative flight lengths).
4. The tracks with persisting negative flight lengths are removed and the default fit is performed until final convergence is reached (usually only one more iteration of the fit).

The re-linearisation of the tracks with negative flight length as to have positive flight length does not only change the sign of the derivative in the Jacobian \hat{A} of the b -hadron flight direction with respect to a shift of the fitted vertex position, but also considerably affects the fit because the constant term \hat{C} in Eq. (6.11) contains the residual (which the fit tries to minimise) between the track and the vertex position on the b -hadron flight axis being now *artificially* chosen to be opposite to the negative fitted one.

Tracks originating from the primary interaction vertex, even with their flight length swapped, will still have small residuals and continue to barely affect the fit of the b -hadron flight direction, while tracks coming from secondary vertices which randomly have formed negative flight length vertices will generally show consistent residuals and will try to move the fit towards succeeding in having their vertex on the right side, out of the wrong minimum, as is schematically illustrated in Fig. 6.16.

6.6.3 Vertex finding strategy

In this chapter the basic finding algorithm implemented in JetFitter will be described: it assumes that N input tracks to be used for the clustering of the $PV \rightarrow b \rightarrow c$ -hadron decay chain have been already selected. Such a selection is crucial to improve the vertex fit quality and will be described more in detail in Section 6.7.1.

After the primary vertex and the b -hadron flight axis have been initialised, a first fit is performed under the hypothesis that each track represents a single vertex along the b -hadron flight axis, until χ^2 convergence is reached, obtaining a first set of fitted $(\phi, \theta, d_1, d_2, \dots, d_N)$.

A clustering procedure is then performed, where all combinations of two vertices (picked up among the vertices lying on the b -hadron flight axis plus the primary vertex) are taken into consideration, using the vertex probability estimation techniques described in the previous section, filling a *table of probabilities*, of the kind of Table 6.5.

After the table of probabilities is filled, the vertices with the highest compatibility are merged, a new complete fit is performed with the new decay chain structure and a new table of probabilities is filled. This procedure is then iterated until no pairs of vertices with a probability above a certain threshold exist anymore. Avoiding the complete fit to be performed

Vertex	1	2	3	...	Primary vtx
1	X	P_{12}	P_{13}	...	P_{1P}
2	...	X	P_{23}	...	P_{2P}
3	X	...	P_{3P}
...	X	...
Primary vtx	X

Table 6.5: *Table of probabilities.* During the clustering a compatibility test is performed between all pairs of vertices to analyse if they form a common vertex on the B flight axis and the *Table of probabilities* is filled accordingly.

when estimating the compatibility of each of such pair of vertices (by using the Kalman update step described in the previous section) reduces the number of iterations needed for the computation from $\propto N_{tracks}^3$ to $\propto N_{tracks}$.

The result of this clustering procedure is a well defined decay chain topology with a certain association of tracks to vertices along the b -hadron flight axis, where each vertex possesses at least one track.

6.6.4 Optimisation of the algorithmic execution speed

There are various reasons why the JetFitter algorithm turns out to be slower than a more conventional Kalman or Billoir based inclusive fit of a single b/c -hadron displaced decay vertex. The first is that, due to the high non-linearity of the measurement equation, more iterations are required than in a normal single-vertex Kalman or Billoir based fit, where 2 or 3 iterations at most are sufficient to reach convergence. The second is that, at least in the Kalman filter *weighted-mean* formalism, the inversion of a matrix with the size of the state vector is required: this size increases from 3 in the case of a conventional single-vertex fit to $5+N$ in the case of JetFitter. This is the most expensive part of the fit, since the inversion of such a $5+N$ matrix is extremely time consuming (the number of single computations required for such an inversion is proportional to $O([N+5]^2)$).

While there is nothing to do to avoid the first issue, which is a consequence of the involved physics, the second issue is only a consequence of the adopted vertex fit formalism. The optimisation of the vertex formalism is discussed in detail in the appendix in Section B.1. After this optimisation was performed, the typical amount of time required for applying the b -tagging algorithm based on JetFitter on all jets in a typical $WH \rightarrow l\nu b\bar{b}$ or $t\bar{t}$ event was reduced to 30-50 ms. This is comparable to the amount of time needed to run the other single b -tagging algorithms, but still higher than the time required for the inclusive single-vertex secondary vertex based algorithm, which is on average 10 ms. One of the ideas which could be exploited to further improve the execution time of the fit would be remove the primary vertex position \vec{r}_{PV} from the variables of the state vector to update: a fixed position for the primary vertex can in fact be considered, in most of the cases, as a very good approximation for the decay chain vertex fit. This can be eventually implemented if there is interest in trying to use the algorithm also in the Trigger chain, for the on-line reconstruction and selection of b -jets.

6.7 Optimisation of the physics performance

The basic finding strategy in JetFitter was described in Section 6.6.3. The simplest possible approach to use the JetFitter algorithm would be to select tracks in a jet according to the loose selection criteria presented in Section 6.2 and then pass them to the JetFitter algorithm, to start the vertex clustering. However, the performance of the algorithm can be significantly improved by optimising the track selection criteria and, most important, by explicitly reconstructing conversions, K_S^0 mesons, Λ baryons and hadronic interactions.

6.7.1 Initial track selection

The performance of a secondary vertex based b -tagging algorithm is extremely sensitive to fake vertices formed occasionally in light jets, typically due to resolution effects or mismeasured track parameters. While in the impact parameter based b -tagging algorithms described in Section 6.3 the track selection is applied only on the absolute values of the transverse and longitudinal impact parameters and the impact parameter significances are then used as an input for the likelihood method, which is based both on prompt and secondary tracks, when looking for secondary vertices in b -jets it is critical to reject tracks which are likely to have their origin directly in the primary vertex.

This is achieved by using the three-dimensional impact parameter significance⁶, which is optimally evaluated by using the χ^2 contribution of a certain track to the primary vertex fit, which is condensed in Eq. 6.8. In case the track was used in the primary vertex fit of the event, the χ^2 obtained in the primary vertex fit after the *smoothing* step can be used directly⁷. Otherwise the χ^2 needs to be computed explicitly. The distribution of the $\text{Prob}(\chi^2)$ for prompt tracks in light-jets and for tracks from b - or c -hadron decays in both b - and c -jets is shown in Fig. 6.17. For completeness also the distribution expected for conversions and K_S^0 is shown. The plot in linear scale shows that the $\text{Prob}(\chi^2)$ distribution is nearly flatly distributed between 0 and 1 for prompt tracks, with a small peak at low probability values. This is expected due to the non-Gaussian behaviour of the resolution function, which was already seen in Fig. 6.7.

The distributions show clearly that an upper cut on the $\text{Prob}(\chi^2)$ distribution also rejects a certain amount of tracks stemming from the b - and c -hadron vertices. In order to keep the highest possible amount of such tracks, but still keep the contamination from prompt tracks as low as possible, the lifetime sign introduced in Eq. 6.3 can be used as an additional discriminator. The separate $\text{Prob}(\chi^2)$ distributions for negative and positive lifetime signed tracks is shown in Fig. 6.17 (bottom). It can be seen that most of tracks from b - and c -hadron decays have a positive lifetime sign, with a smaller $\text{Prob}(\chi^2)$. So, when applying an upper cut on $\text{Prob}(\chi^2)$ it is useful, while preserving a constant number of tracks from light-jets, to cut looser on this quantity when the lifetime sign is positive and tighter when it is negative. A cut value of 10% and 5% has been chosen, respectively, for positive and negative lifetime signed tracks (corresponding respectively to -1 and -1.3 on the Log_{10} scale).

The effect of such a track selection is made more explicit in Table 6.6: on average 15% of the tracks in light-jets pass the selection (2/3 of them still coming from prompt tracks),

⁶In order to take the different resolutions in the transverse plane and longitudinal direction into account, the three-dimensional impact parameter significance cannot be evaluated as $\frac{d}{\sigma_d}$, where d is the distance of closest approach of the track to the vertex in three dimensions, but the complete χ^2 contribution needs to be computed.

⁷The smoothed χ^2 contribution has also the advantage that it is an unbiased measurement of the displacement of the track with respect to the primary vertex, since the track under investigation is removed from the vertex before the χ^2 is computed.

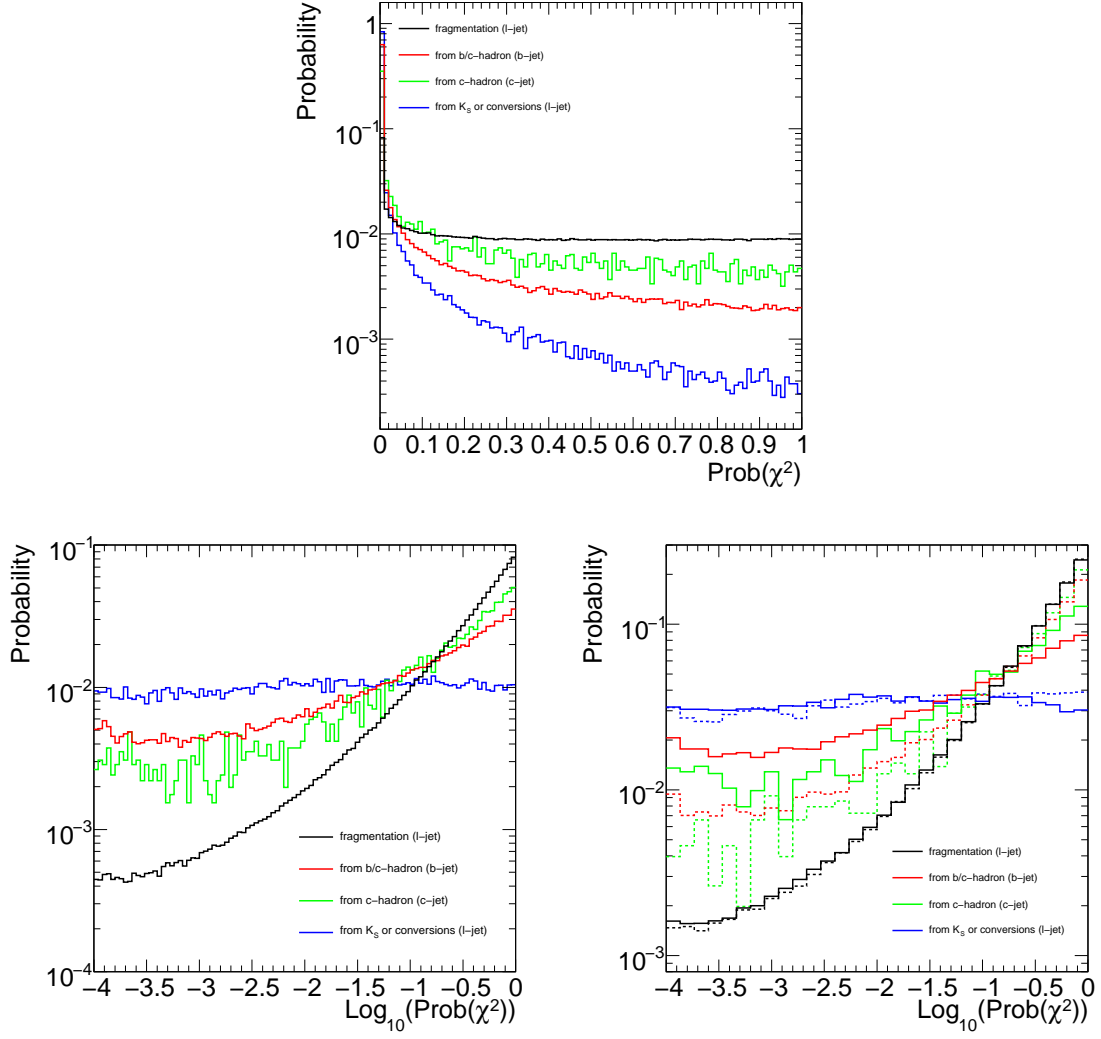


Figure 6.17: $\text{Prob}(\chi^2)$ distribution for tracks having different physics origins, measuring their compatibility with the primary interaction vertex (top plot). The distribution of $\text{Log}_{10}(\text{Prob}(\chi^2))$ is also shown (bottom left plot). In addition, the same quantity is shown separately for tracks with positive (full) and negative (dashed) lifetime sign (bottom right plot).

while 67% of the tracks from b - and c -hadron decays in b -jets. In addition, around 35% of the remaining contribution from light jets comes from conversions, K_S^0 , Λ decays and hadronic interactions.

6.7.2 Detection of two-track vertices

The explicit detection of two-track vertices can significantly decrease the contamination from secondary tracks in light-jets. This is realised by considering all combinations of two-track vertices and trying to obtain a common vertex out of them. Tracks corresponding to such vertices, if they are very likely to originate from light-jets, can be removed from the later fit of the b/c decay chain based on JetFitter. In addition, the two-track vertices which are not rejected by this procedure can be used as a first seed for starting the secondary vertex finding algorithm.

Initial IP cuts ($ d_0 < 7. \text{ mm}$ and $ z_0 < 10 \text{ mm}$)			
	light-jet	<i>c</i> -jet	<i>b</i> -jet
Average track multiplicity	5.4	6.1	6.9
from fragmentation	91.5 %	61.2 %	36.3 %
from <i>b</i> -hadron	0.0 %	0.0 %	26.4 %
from <i>c</i> -hadron	0.1 %	30.3 %	28.6 %
from K_S	1.7 %	2.0 %	1.9 %
from conversion	3.9 %	4.0 %	3.7 %
other	2.8 %	2.5 %	3.1 %

Prob(χ^2) < 10 % (5 %) for $\text{sign}_{3D} > 0$ (< 0)			
	light-jet	<i>c</i> -jet	<i>b</i> -jet
Average track multiplicity	0.8	1.4	3.2
from fragmentation	65.3 %	26.2 %	7.9 %
from <i>b</i> -hadron	0.0 %	0.0 %	34.6 %
from <i>c</i> -hadron	0.2 %	51.1 %	44.8 %
from K_S	8.0 %	6.1 %	3.1 %
from conversion	16.2 %	10.8 %	5.5 %
other	10.3 %	5.9 %	4.1 %

Table 6.6: Average track multiplicity and relative contribution of tracks originating from different physics processes in light, *c*– and *b*-quark jets, before and after the Prob(χ^2) cut used for selecting tracks intended to be used in JetFitter for the secondary vertex fit. The requirement of the loose track selection criteria of Table 6.3 has already been applied.

The Kalman filter based vertex reconstruction algorithm is used for the vertex fit of all pairs of displaced tracks (the same used in the sequential primary vertex finder). This is the best choice in this case, since the adaptive vertex fit can help (down-weighting outliers) only if more than two tracks are available in the fit. In order to avoid numerical problems and speed up the fit convergence, the fit is initialised by using the 3*d* seed finder introduced in Section 5.4.2. The mean point on the axis representing the minimum distance between the two tracks is used as an initial estimate for the vertex position. The fit is considered as having converged if, after a new iteration of the Kalman update steps, $|\chi_{new}^2 - \chi_{old}^2| < 0.001$. No constraint on the two-track mass is applied in the fit.

Some properties of the two-track vertices are shown in Fig. 6.18. Tracks which genuinely form a vertex have a value for the Prob(χ_{vtx}^2) provided by the vertex fit which is quite flatly distributed between 0 and 1. Wrong combinations of tracks (which do not form a real vertex) peak at low Prob(χ_{vtx}^2), however, when the resolution does not allow to separate the two tracks, values of Prob(χ_{vtx}^2) in proximity to one can still be reached. Tracks from fragmentation have an analogous behaviour, not because they do not form a real vertex, but because on each of the two tracks the requirement not to be compatibility with the primary vertex has already been applied. In addition, for all distributions presented here, the Prob(χ_{vtx}^2) has already been required to be higher than 0.34%. The second plot in the same figure shows the probability of both tracks at the vertex to be simultaneously compatible with the primary vertex (by summing up the corresponding χ_{trk}^2 contributions of Eq. 6.8). The two-track vertices formed by the residual tracks from fragmentation in light-jets clearly peak at higher values of compatibilities, however, due to the long non-Gaussian tails of the resolution function, they extend even up to probabilities of 10^{-5} or less. Tracks from *b*- and *c*-hadron decays are more significantly displaced. Random combinations of tracks which happen to be compatible with a single vertex are in between. The third plot of Fig. 6.18 shows the flight length significance distribution. This distribution is significantly correlated with the second,

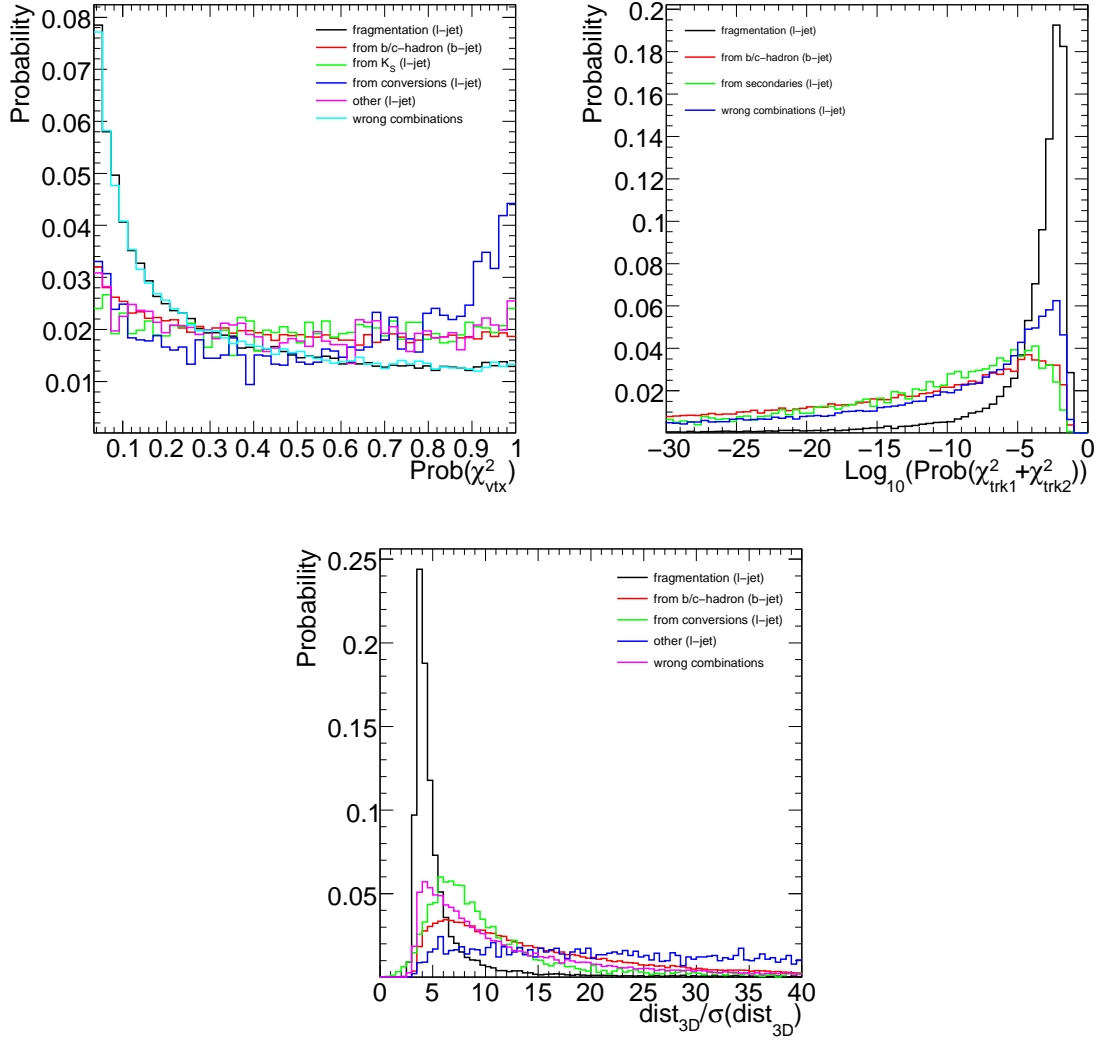


Figure 6.18: Top left distribution: vertex probability of two-track vertices. Top right distribution: probability that both tracks fitted in a vertex are compatible with the primary vertex. Bottom distribution: distance between primary vertex and two-track vertices, divided by fit error, after the requirement on $\text{Prob}(\chi^2_{trk1} + \chi^2_{trk2}) < 10^{-2}$ has already been applied.

but also offers additional discrimination power. The reason is that the lifetime significance evaluates the compatibility of the tracks to the primary vertex only along the direction of flight in which the two-track vertex is pointing, under the hypothesis that the two tracks form a single vertex, while the lateral displacement of the single tracks with respect to the primary vertex is not considered anymore. While the latter is usually more powerful, the flight length significance can still provide some additional discrimination power.

After the two-track vertices are reconstructed, the following additional selection criteria are applied:

- $\text{Prob}(\chi^2_{trk1} + \chi^2_{trk2}) < 10^{-2}$
- $\text{dist}_{3D}/\sigma(\text{dist}_{3D}) > 1.5$
- $(\vec{r}_{2tr-vtx} - \vec{r}_{PV}) \cdot \vec{p}_{jet} > 0$

- $|IP_{r\phi}| < 3.5$ mm and $|IP_z| < 5$ mm

These requirements are very loose and are intended to keep the highest possible amount of real displaced vertices.

After such a selection, the two-track vertices in light jets are dominated by:

1. hadronic interactions in the layers of the inner detector (Fig. 6.19)
2. remaining tracks from fragmentation (visible in Fig. 6.19 at small radii)
3. K_S^0 mesons decaying into a pair of oppositely charged pions (Fig. 6.20)
4. Λ baryons decaying into a proton and a negative charged pion (Fig. 6.21)
5. photons converting into electron-positron pairs (Fig. 6.22)

In all distributions of the selected two-track vertices both the contribution from light and from b -jets is shown. All distributions are normalised to the overall number of selected two-track vertices in b - and light jets. However, after the vertex selection cuts, the average number of two-track vertices in light jets is around 0.12, while it is 3.7 in b -jets and 0.54 in c -jets. So, if one would normalise the distributions to the average number of expected two-track vertices, the distributions for light jets would need to be rescaled by a factor of 30 lower with respect to b -jets.

Examining the radial distributions of the two-track vertices in Fig. 6.19 in more detail, some peaking structures (in particular in light jets) are clearly visible, corresponding to the positions in space where the hadronic interactions are more likely to occur, i.e. where layers of material are present in the detector (see Fig. 3.3). The first two peaks at a decay radius of ≈ 3 cm refer to the two-layer structure of the beam pipe⁸. The first pixel layer, usually called b -layer, is broader (since the structure is not circular, but the radius depends on the azimuthal angle ϕ) and situated at around 5 cm. The second pixel layer is visible at around 9 cm, while a further small peak is visible at a radius of 7 cm, most probably due to some pixel modules support structures. The resolution of hadronic interactions is in general very good, since tracks coming out of them have reasonable transverse momenta and, more importantly, the two tracks picked up by the two-track vertices (out of in general many more than just two tracks) have a significant opening angle with respect to each other.

The corresponding distribution in b -jets shows the expected distribution of two-track vertices originating mainly from b - and c -hadron decays, which essentially wash out the few candidates expected from hadronic interactions. Most of the b - and c -hadrons decay at few mm radius, well inside the beam pipe. This is the case only for jets of up to 150-200 GeV in transverse momentum, as considered in this study (since the average transverse decay radius is proportional to p_T/m). To reduce the contamination from hadronic interactions one can reject tracks belonging to two-track vertices corresponding to vertices in correspondence of pixel layers. This is however only reasonable if this does not remove a too high amount of tracks from b -jets, which would later reduce the vertex reconstruction efficiency. Therefore, only tracks of two-track vertices corresponding to interactions in the first and second layer of the pixel detector are flagged as originating from material interactions.

The invariant mass distribution for two-track vertices, assuming the mass of the two decaying particles to be the pion mass (Fig. 6.20) and selecting only vertices with oppositely charged tracks, shows that K_S^0 mesons can be reconstructed very well. A clear peak is seen

⁸The double-layer structure of the beam pipe is an artifact of the simulation, which will be corrected in new versions of the detector geometry.

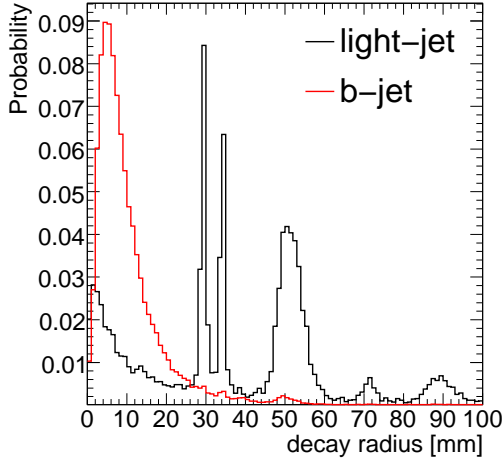


Figure 6.19: Radial distance of selected two-track vertices from the interaction region, in both light- and b -jets. The interactions in the beam pipe and in the various pixel layers are clearly visible.

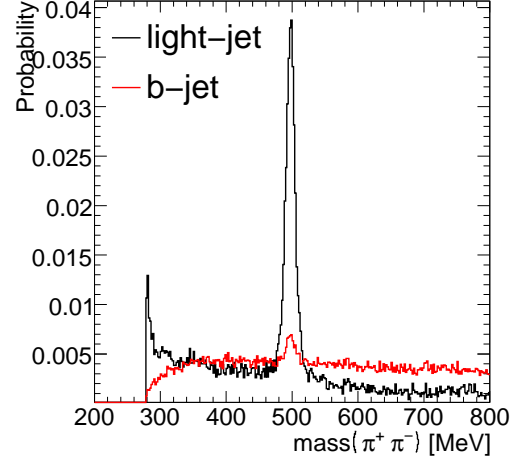


Figure 6.20: Invariant mass distribution of all selected two-track vertices, in both light- and b -jets, under the hypothesis of pion masses for both the decaying particles. Only combinations of oppositely charged tracks are considered here.

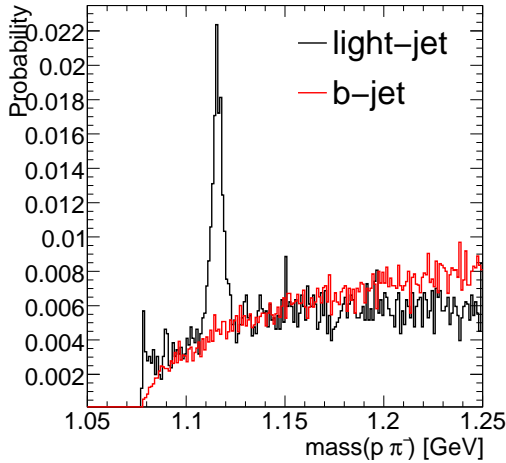


Figure 6.21: Invariant mass distribution of all selected two-track vertices, in both light- and b -jets, under the hypothesis of proton and pion masses for the decaying particles (the proton mass is assigned to the charged particle track with the largest momentum).

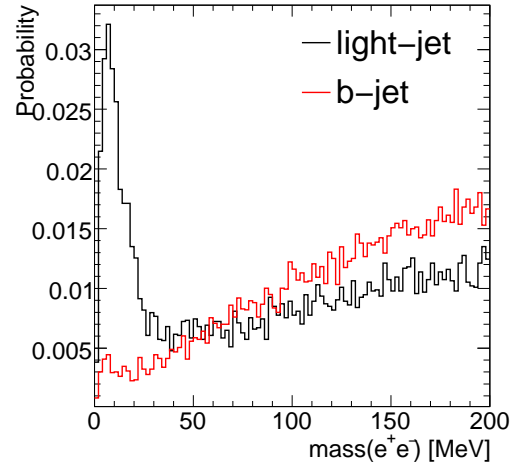


Figure 6.22: Invariant mass distribution of all selected two-track vertices, in both light- and b -jets, under the hypothesis of electron masses for both decaying particles.

on top of background coming from other sources of two-track vertices. In b -jets the K_S^0 peak is seen only as a small peak on top of a large continuum *background* of two-track vertices representing a subset of the b - and c -hadron decay products in a certain jet. In addition, the small peak is formed both by prompt K_S^0 mesons and by K_S^0 mesons emitted through the b - or c -hadron decay, where it would be convenient only to remove the first ones. Additional separation criteria will therefore be needed to distinguish these three categories further. A similar separation is desirable also for Λ baryons, where the fraction of Λ baryons in b -jets is negligible, but the removal of Λ baryons would again remove some two-track vertices cor-

responding to good tracks from b - and c -hadron decays. The situation is a bit better for conversions (Fig. 6.22), where an upper cut on the invariant mass removes only a very small fraction of nearly collinear tracks from b - and c -hadron decays.

6.7.3 Explicit reconstruction of neutral particles

Two-track vertices which are likely to represent a neutral particle are further processed. Beyond the general selection cuts for two-track vertices illustrated in the previous section, additional selection criteria are defined to select conversions, K_S^0 and Λ baryons, as illustrated in the following. As an exception, in order not to lose too many neutral candidates, the impact parameter cuts are loosened to $|IP_{r\phi}| < 7$ mm and $|IP_z| < 10$ mm. The additional selection criteria for conversions are:

- oppositely charged tracks;
- $m(e^+e^-) < 30$ MeV,

while for K_S^0 mesons and Λ baryons they are:

- opposite charged tracks;
- $|m(\pi^+\pi^-) - 497.6 \text{ MeV}| < 18 \text{ MeV}$ (for K_S^0 mesons only);
- $|m(p\pi^-) - 1115.7 \text{ MeV}| < 30 \text{ MeV}$ (for Λ baryons only);
- $\text{Prob}(\chi_{trk1}^2 + \chi_{trk2}^2) < 10^{-4}$;
- $\text{Prob}(\chi_{trk1,2}^2) < 10^{-3}$ (on the single tracks);
- $\text{dist}_{3D}/\sigma(\text{dist}_{3D}) > 4.5$;
- $\text{radius}(vtx) > 2 \text{ cm}$ OR at least one track with $|d_0| > 1 \text{ mm}$ and $|z_0| > 2 \text{ mm}$.

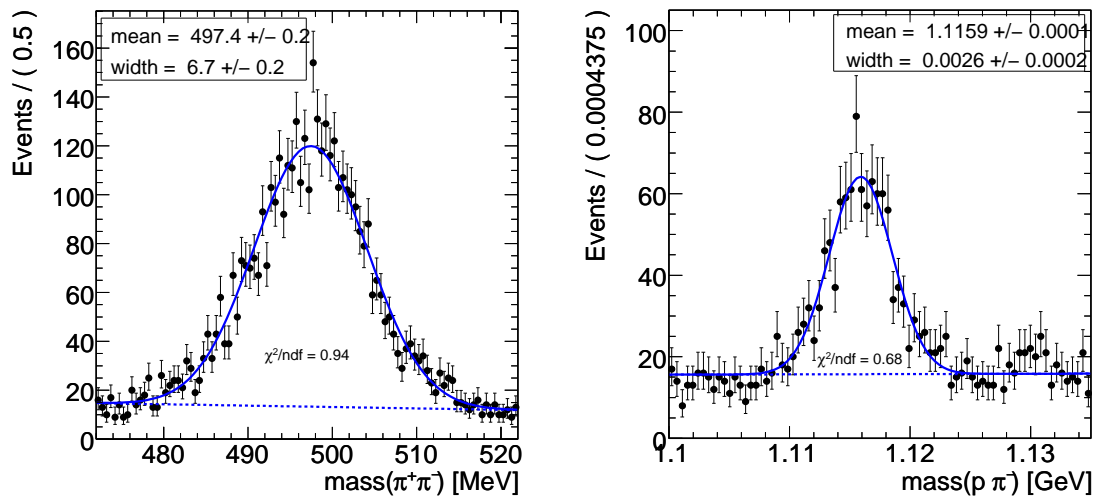


Figure 6.23: Mean and width of a binned likelihood fit to the $m(\pi^+\pi^-)$ (left plot) and $m(p\pi^-)$ (right plot) distributions in proximity of the K_S^0 (or Λ) peak, with a likelihood model given by a superposition of a Gaussian function and a first order polynomial.

The mass window cuts applied for K_S^0 and Λ particles are based on the achievable invariant mass resolution, which is studied more in detail in Fig. 6.23 (all remaining specific selection cuts for neutral candidates have been applied here). The reconstructed means are compatible with the PDG world averages [71] within statistical uncertainties. The width for the cuts have been chosen for the K_S^0 peak at around 2.2σ and for the Λ peak at 3.4σ around the nominal values. While it is sensible to choose a larger interval when the resolution is smaller, since a smaller amount of background can pass through, it should be stressed that these values were not optimised and could be thus refined.

While in the case of conversions, the amount of wrongly selected two-track vertices from b - and c -hadron decays and thus of potentially fake neutral candidates is relatively low, in the case of K_S^0 and Λ decay vertices it is non-negligible. Therefore in the latter case, a neutral candidate is accepted only when either the decay radius is higher than 2 cm (since most of the two-track vertices from b -jets can be found at smaller radii, as demonstrated in Fig. 6.19) or when at least one of the two tracks at the vertex does not pass the tight impact parameter cuts of Table 6.4.

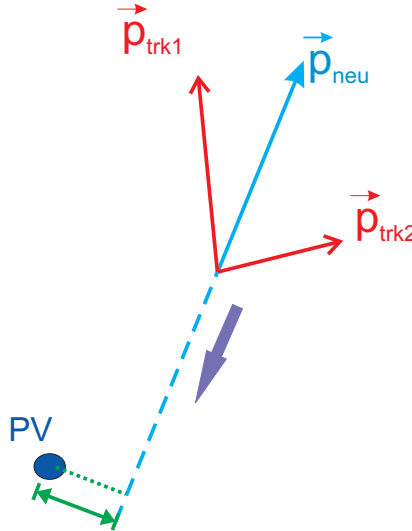


Figure 6.24: Schematic view of how a neutral track is created out of a two-track vertex and then extrapolated to the primary vertex.

Once conversions, K_S^0 mesons and Λ baryons are identified, a *neutral track* is created out of them (Fig 6.24). A neutral track is a mathematical artifact in the ATLAS reconstruction software which permits to handle such a neutral object with the same tools usually used for charged tracks. The neutral track is defined through a position in space (the decay vertex) and the momentum at the decay vertex, plus the complete position / momentum covariance matrix, whose extraction requires some care. The momentum of the neutral track is set to the sum of the two-track momenta defined at the vertex position, but with the additional constraint that the two tracks have to pass through the fitted vertex: this information is provided during the *smoothing* step by the Kalman based two-track vertex fit. The vertex constraint has the effect of correlating the vertex position with the momentum at the vertex both of the first and the second track. In principle, to take these correlations into account, the complete 9x9 fit covariance matrix, with vertex position, momentum of first and second track, is needed. However, in the case the only correlation between the two tracks in the fit comes from the correlation of the single tracks momenta with the vertex position, the correlation

matrix between the two tracks can be recovered a-posteriori, by simple propagation of the correlation matrices of the tracks momenta with the vertex position and of the vertex position itself:

$$COV(\vec{p}_{trk1}, \vec{p}_{trk2}) = COV(\vec{p}_{trk1}, \vec{r}) COV^{-1}(\vec{r}, \vec{r}) COV(\vec{r}, \vec{p}_{trk2}). \quad (6.21)$$

Once all covariance matrices are available, the complete covariance matrix for $[\vec{r}, \vec{p}_{neu}]$ can be obtained using error propagation (in the full matrix formalism) starting from $[\vec{r}, \vec{p}_{trk1}, \vec{p}_{trk2}]$ and $COV_{\vec{r}, \vec{p}_{trk1}, \vec{p}_{trk2}}$. The result can be then easily translated into the 5 parameters perigee representation of a neutral track.

In this way the neutral track is fully defined. If it corresponds to a neutral particle originating directly from the primary vertex, it should have a small impact parameter, even if the impact parameters of the two tracks at the decay vertex were very large. The impact parameters can be obtained, exactly as in the case of a charged track, by extrapolating the track to the point of closest approach in the transverse plane to the primary vertex. The extrapolation tool of the ATLAS reconstruction software will recognise that the track is of neutral nature and will avoid to bend it as an effect of the magnetic field. As a result the new track parameters $d_0, z_0, \phi, \theta, 1/p$ and related covariance matrix are obtained, where, as explained at the beginning of Section 5.3.4, the parameter q/p of the charged parametrisation is substituted by $1/p$. The impact parameters are, again as in the charged case, d_0 and $z_0 \cdot \sin(\theta)$, in the $x-y$ and $r\phi-z$ planes, respectively. The two impact parameters can be again combined into a single χ^2 discriminator with two degrees of freedom, expressing the compatibility with the primary vertex.

In order to validate the reconstruction of neutral particles, the reconstructed neutral track candidates, in terms of impact parameters, particle momenta and their associated uncertainties, have been compared with their related Monte Carlo truth based quantities. This validation is shown in the appendix in Section B.2. As a result, it can be concluded that the track parameters and their associated errors are correctly estimated.

The impact parameter significance or the combined χ^2 value expressing the compatibility of the neutral track to the primary vertex can be used to further separate neutral candidates whose origin lies in the primary vertex and either real neutral candidates or fake candidates originating in b or c -hadron decays. For K_S^0 mesons the impact parameter significance distributions are shown in Fig. 6.25, while the combined $\text{Prob}(\chi^2)$ distributions are shown in Fig. 6.26. The impact parameter significances have been signed with the three-dimensional lifetime sign of Eq. 6.3. The analogous distributions for conversions are shown in Figs. 6.27 and 6.28.

For both K_S^0 mesons and conversions a relative good separation can be obtained between prompt candidates and the remaining real or fake candidates from b or c -hadron decays. Due to the non-Gaussian tails in the impact parameter significance distribution from prompt tracks, the $\text{Prob}(\chi^2)$ is not perfectly flatly distributed, but analysing for example $\text{Log}_{10}(\text{Prob}(\chi^2))$, it can be seen that a good discrimination between the two mentioned components is still possible. At the same time it turns out to be much more difficult to separate real from fake candidates from b/c -decays, which are found in b -jets with a ratio of $n_{\text{real}}/n_{\text{fake}} \approx 0.5$.

Exploiting this new information, tracks corresponding to conversions, K_S^0 and Λ particles whose impact parameter is compatible with the primary vertex (with $\chi^2 < 36$, corresponding to $\text{Prob}(\chi^2) > 10^{-7.8}$) are flagged as neutral candidates really originating from the primary vertex. In the case of K_S^0 mesons and Λ baryons, the remaining ones are flagged as neutral tracks to be directly used later in the b/c -decay chain fit, but, in order to reduce the contamination of fake candidates from heavy hadron decays, the neutral candidates without at least one track failing the tight impact parameter cuts ($|d_0| < 1$ mm and $|z_0 \cdot \sin(\theta)| < 2$ mm)

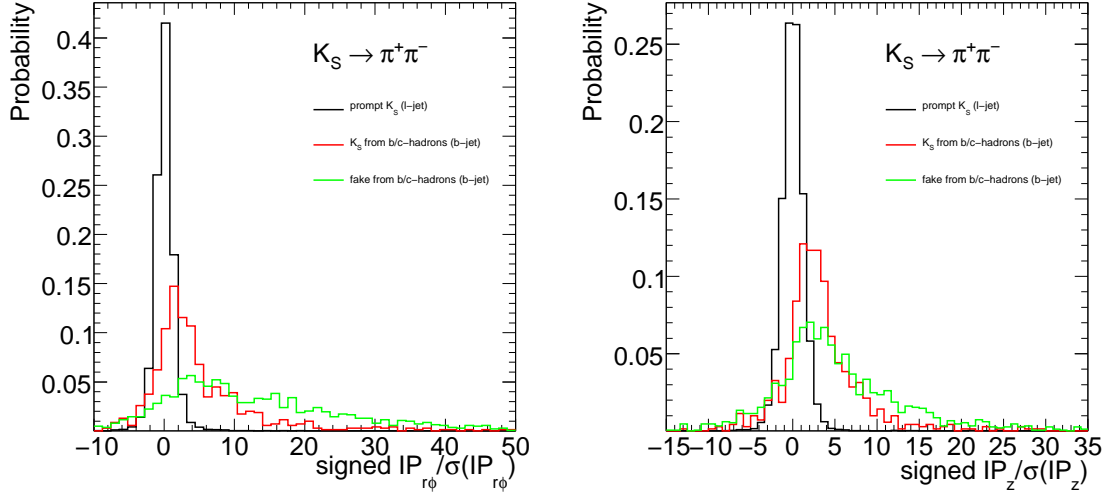


Figure 6.25: Distribution of the transverse (left) and longitudinal (right) impact parameter significances for K_S^0 mesons, separately for prompt K_S^0 mesons in light-jets, K_S^0 mesons from b/c -hadron decays and for fake candidates corresponding to tracks arising directly from b - or c -hadron decays in b -jets.

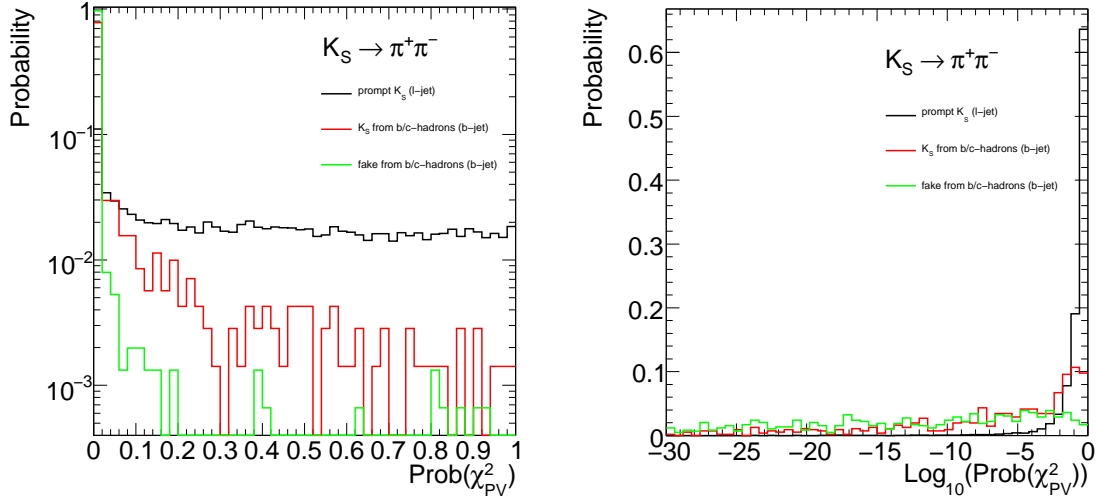


Figure 6.26: Distribution of $\text{Prob}(\chi^2)$ (compatibility to the primary vertex) for K_S^0 mesons, separately for prompt K_S^0 mesons in light-jets, K_S^0 mesons from b/c -hadron decays and for fake candidates corresponding to tracks arising directly from b or c -hadron decays in b -jets. The $\text{Prob}(\chi^2)$ distribution is shown with a linear scale on the x -axis (on the left) and with a logarithmic scale (on the right).

are removed from the neutral particle candidates. This requirement reduces the K_S^0 and Λ reconstruction efficiency by a factor of ≈ 4 and brings the ratio of real to fake vertices to approximately $n_{\text{real}}/n_{\text{fake}} \approx 1.6$.

When used later in the b/c multi-vertex fit, neutral tracks are inferior compared to charged tracks in terms of impact parameter resolution in both transverse and longitudinal directions (Fig. 6.29), but can still provide a useful additional constraint on the b/c vertex decay positions. The vertexing algorithms were designed to transparently handle both neutral and

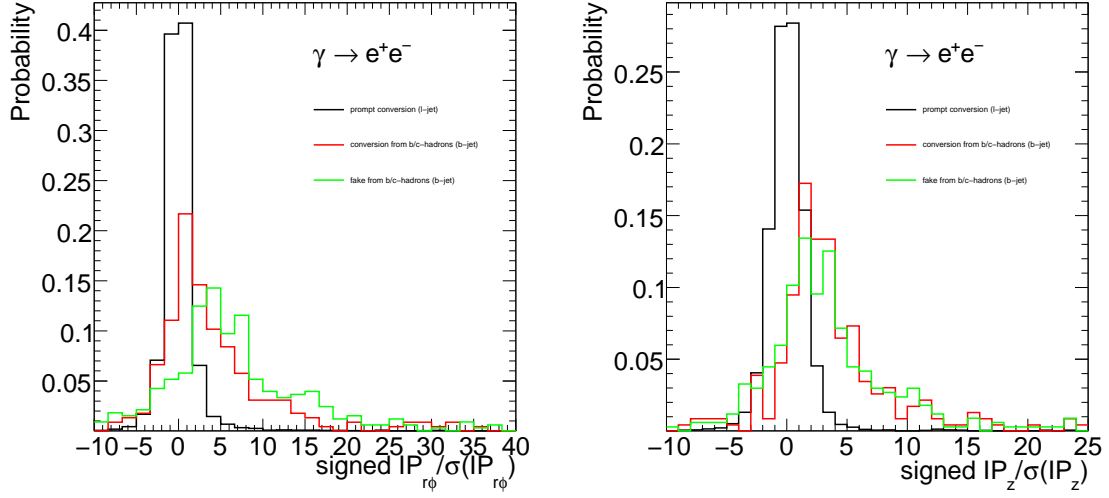


Figure 6.27: Distribution of transverse (left) and longitudinal (right) impact parameter significances for conversions, separately for prompt conversions in light-jets, conversions from b/c -hadron decays and for fake candidates corresponding to tracks arising directly from b or c -hadron decays in b -jets.

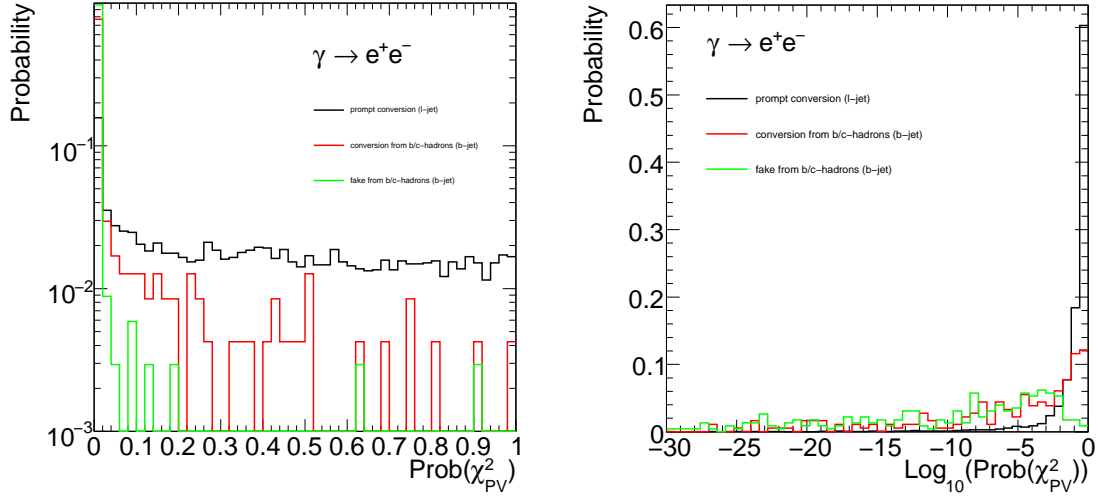


Figure 6.28: Distribution of $\text{Prob}(\chi^2)$ (compatibility to the primary vertex) for conversions, separately for prompt conversions in light-jets, conversions from b/c -hadron decays and for fake candidates corresponding to tracks arising directly from b or c -hadron decays in b -jets. The $\text{Prob}(\chi^2)$ distribution is shown with a linear scale on the x -axis (on the left) and with a logarithmic scale (on the right).

charged tracks. The only real change needed in JetFitter is during the initialization phase of the fit, when the initial positions of the single-track vertices along the initial b -flight axis are determined. In the case of a neutral track, the minimum distance between the flight axis and the neutral track needs to be determined: this involves the distance between two straight tracks, instead of a straight and a charged track. The former is however even simpler than the latter, since the minimum distance between two straight lines can be easily computed analytically.

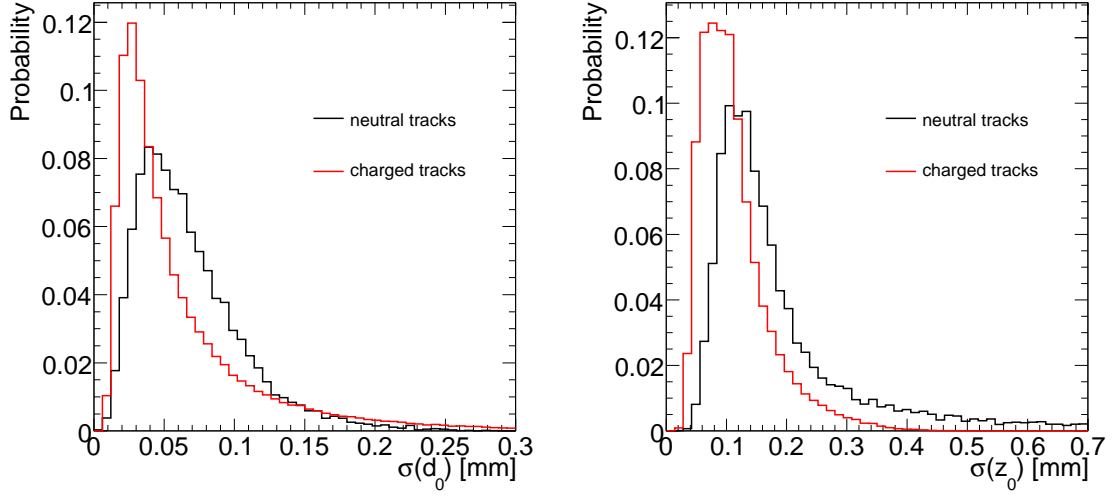


Figure 6.29: Distributions of transverse (left) and longitudinal (right) impact parameter uncertainties for displaced neutral and charged tracks selected in b -jets, used later in the $b \rightarrow c$ -hadron decay chain fit. Most of the neutral tracks correspond to K_S^0 and Λ candidates.

6.7.4 Final selection of tracks and finding procedure

Here the overall strategy used to select tracks to be used in the $PV \rightarrow b^- \rightarrow c$ -hadron decay chain fit is summarised briefly.

After the initial track selection (Section 6.7.1), all possible two-track vertices are reconstructed, according to the selection criteria defined in Section 6.7.2. Two-track vertices having a decay radius corresponding to the radial position of the first or the second layer of the pixel detector are flagged as being produced by hadronic interactions. At the same time, using different selection criteria, neutral particles, corresponding either to conversions, K_S^0 mesons or Λ baryons, are reconstructed (Section 6.7.3), and, using the impact parameter information of the associated neutral tracks, they are separated into prompt particles and into secondary neutral tracks most likely originating from b/c -hadron decays.

In addition to tracks corresponding to two-track vertices, additional single tracks (not attached to any of the selected two-track vertices) passing tighter selection criteria are also considered:

- $|IP_{r\phi}| < 1.5$ mm and $|IP_z| < 3$ mm
- $p_T > 750$ MeV
- $\text{Prob}(\chi_{PV}^2) < 5\%$ ($< 1\%$) for $\text{sign} > 0$ ($\text{sign} < 0$),

Afterwards, the real vertex finding procedure with JetFitter (Section 6.6.3) is started:

- A first decay chain fit is performed only with the tracks corresponding to the selected two-track vertices, after removing those flagged as material interactions and as prompt neutral particles and after adding the additional secondary neutral tracks flagged as originating from b/c -hadron decays.
- A second decay chain fit is performed, on top of the result of the first decay chain fit, by using also the additional selected single tracks.

The clustering procedure is then started, according to the finding strategy outlined in Section 6.6.3, in order to determine the decay chain topology. The complete fit result is stored, to be used later for b -tagging. The event data model used to store this information in the ATLAS event output and, in general, the way the JetFitter finding and fitting algorithms are implemented in the ATLAS reconstruction framework are described with some detail in the appendix in Section B.3.

6.8 Geometrical performance of the secondary vertex finders

In this Section, the performance of the inclusive secondary vertex reconstruction algorithms as described in Section 6.5 and 6.6 is discussed. Both the inclusive secondary vertex finder and the JetFitter algorithms are applied on jets reconstructed in a sample of WH (with $H \rightarrow b\bar{b}$, $u\bar{u}$ and $c\bar{c}$) and $t\bar{t}$ events.

Table 6.7 shows the rate of reconstructed secondary vertices in b -, c - and light quark jets in bins of jet transverse momentum and pseudo-rapidity for both secondary vertex reconstruction algorithms, as obtained on the $t\bar{t}$ and WH samples. In order to make a comparison possible, in the case of the JetFitter algorithm a secondary vertex is considered as found in the event only if at least one vertex with two or more associated tracks was reconstructed. Decay chains with two single-track vertices are left out of this comparison.

Above a certain transverse momentum, the fraction of b -jets where a secondary vertex is reconstructed, for both algorithms, approaches approximately a constant value around 75 – 80%, while the number of vertices reconstructed in light-jets significantly increases with higher jet transverse momentum. These effects have to do with the ability to separate real or fake secondary tracks from the primary vertex: charged particle tracks from secondary particles are selected, as described in the previous sections, according to their impact parameter significance with respect to the primary vertex.

In the case of b -jets, the impact parameters of the tracks originating from displaced b - and c -hadron vertices and the vertex flight length significance are nearly invariant under Lorentz boost; however, the transverse momenta of secondary particles increase with the transverse momentum of the original b -quark and thus the error on both the impact parameter and on the flight length decreases, increasing progressively the fraction of cases where a secondary vertex from b/c -hadron decays can be distinguished from the primary vertex.

	0 < $ \eta $ < 0.5			0.5 < $ \eta $ < 1.0			1.0 < $ \eta $ < 1.5			1.5 < $ \eta $ < 2.0			2.0 < $ \eta $ < 2.5		
p_T [GeV]	b	c	l	b	c	l	b	c	l	b	c	l	b	c	l
15–30	50.0 55.4	13.3 18.0	1.3 2.0	49.0 54.4	13.6 17.4	1.3 1.8	47.4 52.6	11.6 15.4	1.2 1.8	43.2 48.7	10.4 13.3	1.3 1.9	34.1 39.5	8.0 11.1	1.4 2.2
30–50	66.9 71.4	21.0 26.1	2.5 3.8	65.9 70.6	19.9 24.6	2.3 3.2	63.6 68.7	19.3 24.2	2.3 3.4	58.8 63.7	15.6 19.8	2.7 3.7	50.3 56.2	14.1 18.2	3.0 4.6
50–80	74.2 77.3	25.9 30.6	3.8 5.5	73.7 76.6	24.7 29.4	3.6 4.7	71.5 74.3	23.7 27.4	3.7 4.9	66.5 69.6	21.2 24.4	4.3 5.8	59.7 63.6	18.6 22.9	5.2 6.9
80–120	77.8 79.1	30.7 34.8	5.7 7.8	77.3 78.4	29.3 32.3	5.4 6.8	75.2 76.6	27.7 31.2	5.4 6.9	70.8 72.2	25.5 28.3	6.2 7.8	64.2 65.9	24.4 25.7	7.4 9.8
120–200	78.8 79.5	34.2 37.2	8.5 11.2	77.4 78.3	31.6 34.5	8.0 10.1	76.0 76.1	31.0 34.5	8.4 10.1	70.1 70.2	29.8 30.5	9.5 11.7	63.0 63.5	26.7 28.7	11.6 14.2

Table 6.7: The fraction of jets with at least one reconstructed secondary vertex passing the selection criteria to be used by the b -tagging algorithms in b -, c - and light quark jets for the inclusive secondary vertex finder (top) and JetFitter (bottom). These numbers, given in percent, have been obtained on the WH and $t\bar{t}$ samples.

The rising amount of fake vertices in light quark jets with increasing jet transverse momenta is partly due to the nature of the fragmentation process in light quark jets, which produces a higher number of tracks coming from the primary interaction vertex the higher the transverse jet momentum is: in fact, being the probability for a reconstructed primary

track to yield an outlying impact parameter measurement approximately constant in p_T up to a first approximation, the probability of having one or more charged particles track which mimic the presence of secondary particles increases roughly linearly with the amount of primary tracks. The real secondary vertices produced in light quark jets, like conversions and V0 decays, represent an additional source of displaced tracks in light jets.

The dependence of the vertex reconstruction efficiency on the jet pseudo-rapidity is mainly a consequence of the different resolutions which can be achieved in different regions of the inner detector: at increasing rapidities, the track resolutions in proximity of the interaction point get worse, the track reconstruction efficiency starts to decrease and the amount of badly reconstructed tracks rises.

	15 < p_T < 30			30 < p_T < 50			50 < p_T < 80			80 < p_T < 120			120 < p_T < 200		
	b	c	l	b	c	l	b	c	l	b	c	l	b	c	l
Nothing	30.0	66.8	89.0	17.5	56.6	84.4	12.8	51.0	81.1	10.9	45.5	77.4	10.3	42.2	71.2
1 Single Track	17.0	16.9	8.8	12.8	18.9	11.6	10.9	19.8	13.0	10.0	21.2	14.3	10.2	21.0	16.3
2 Single Tracks	1.6	0.7	0.2	1.9	1.0	0.3	2.2	1.3	0.5	2.8	1.7	0.8	3.3	2.5	1.2
1 Single Vertex	44.0	14.8	1.9	53.0	21.6	3.4	52.8	25.1	4.9	49.5	27.6	6.7	45.2	28.2	9.6
1 Vertex + 1 Track	5.7	0.8	0.07	10.7	1.6	0.2	14.7	2.5	0.4	18.2	3.5	0.8	21.3	5.2	1.4
2 Vertices	1.6	0.08	0.00	4.1	0.2	0.01	6.6	0.3	0.03	8.6	0.6	0.09	9.6	0.9	0.2

	0.0 < $ \eta $ < 0.5			0.5 < $ \eta $ < 1.0			1.0 < $ \eta $ < 1.5			1.5 < $ \eta $ < 2.0			2.0 < $ \eta $ < 2.5		
	b	c	l	b	c	l	b	c	l	b	c	l	b	c	l
Nothing	14.3	54.4	89.3	15.1	55.8	89.1	16.6	56.9	88.2	20.0	60.2	87.0	25.3	63.5	86.3
1 Single Track	10.7	18.0	7.6	11.0	18.1	8.0	12.2	18.7	8.7	14.4	19.1	9.7	17.2	18.8	10.1
2 Single Tracks	2.2	1.2	0.3	2.3	1.2	0.3	2.3	1.1	0.3	2.5	1.3	0.3	2.5	1.2	0.3
1 Single Vertex	49.8	23.5	2.6	50.2	22.3	2.4	49.7	21.0	2.5	47.1	17.4	2.8	42.4	14.7	3.0
1 Vertex + 1 Track	15.8	2.5	0.2	14.7	2.3	0.2	13.5	2.1	0.2	11.6	1.8	0.2	9.4	1.5	0.3
2 Vertices	7.3	0.4	0.03	6.7	0.4	0.02	5.7	0.3	0.03	4.3	0.3	0.02	3.1	0.2	0.02

Table 6.8: Population of the different specific decay chain topologies reconstructed by *JetFitter* applying the vertex reconstruction on b -, c - and light quark jets, shown as a function of p_T (GeV) in the top table and as a function of η in the bottom table. These numbers, given in percent, have been obtained on the WH and $t\bar{t}$ samples.

As stated in Section 6.6, the key feature of *JetFitter* is the ability to distinguish between several decay chain topologies, in addition to recognising the presence of a single inclusive decay vertex. Table 6.8 shows the population of the various different reconstructed topologies for b -quark, c -quark and light-quark jets, as a function of the jet transverse momentum and pseudo-rapidity, while Table 6.9 illustrates the average behaviour.

The fractions of b -jets and light jets landing in the different categories of vertex topologies show that most of the gain achievable with *JetFitter* in terms of b -jet identification performance is provided by the ‘1 Vertex + 1 Track’ and ‘2 Vertices’ categories, where, due to the expected different relative population of the these topologies in b - and light-jets, a gain in rejection against light quark jets of a factor ~ 3 -4 and ~ 10 can be obtained on approximately $\sim 14\%$ and $\sim 6\%$ of the reconstructed b -jets, respectively. Another small improvement comes from the two single track category, which potentially gives access to one prong b -/one prong c -decays, which the fit of a single secondary vertex misses completely: this however affects only $\approx 2\%$ of the b -jets, with a rejection this time around a factor 2-3 lower than in the single vertex category. The dependence on p_T and η follows the same pattern as already described for the inclusive vertex reconstruction.

6.8.1 Vertex association efficiency and purity

Another good criterion to estimate the algorithmic performance of the secondary vertex finders is to look at the:

- *association efficiency*: fraction of the charged particles stemming from the decays of the

b - or c -hadrons and reconstructed as tracks in the inner detector which are correctly associated to a displaced vertex;

- *association purity*: fraction of the charged particle tracks assigned by the secondary vertex finders to one or more displaced vertices which are really stemming from b - or c -hadron decays.

A b -quark hadronizes into a b -hadron, which then subsequently decays, most of the times, into one or more c -hadrons. The average charged particle multiplicity at these decay vertices as obtained from the Monte Carlo generator PYTHIA was analysed, where a minimum transverse momentum of 500 MeV was required for the charged particles. The decay products of a strongly decaying c -hadron (e.g. D^*) are considered as stemming from the b -hadron vertex, while only the decay products of weakly decaying c -hadrons are considered as stemming from the c -hadron vertex. The average charged particle multiplicity of the inclusive b/c -hadron vertex is ~ 4.7 , as shown in Fig. 6.30 (left), distributed in ~ 2.3 charged particles stemming from the b -hadron vertex and ~ 2.4 charged particles stemming from the c -hadron vertex. Around 80% of the b - and c -hadron charged decay products are correctly reconstructed as tracks and pass the loose track selection cuts (defined in Table 6.3): as shown in Fig. 6.30 (right), the overall average charged particle multiplicity decreases to ~ 3.8 , the b -hadron decay multiplicity to ~ 1.8 and the c -hadron decay multiplicity to ~ 2 .

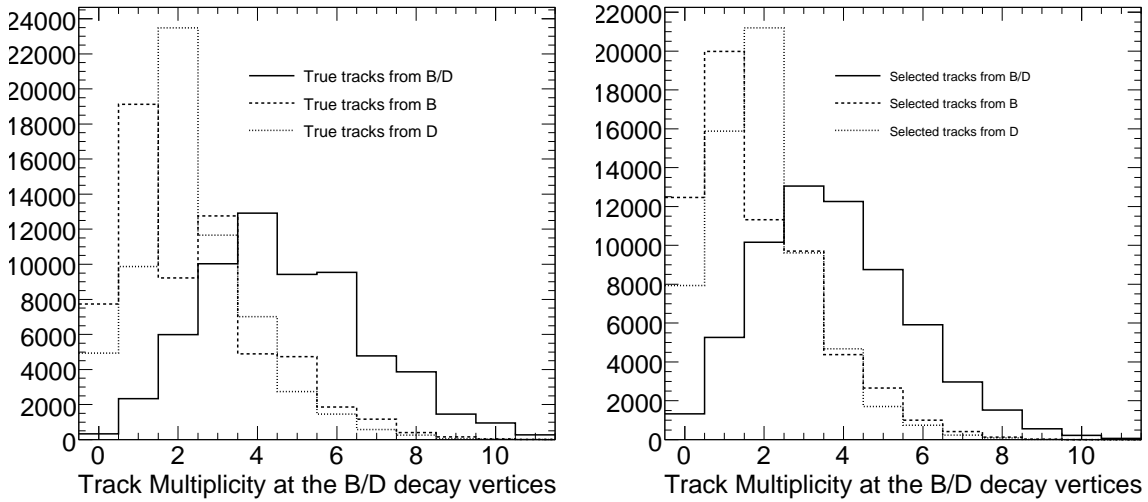


Figure 6.30: Average charged particles decay multiplicity of the b - and/or c -hadrons at generator level (on the left), compared with the amount of charged particles stemming from the b - and/or c -hadron vertices reconstructed as tracks in the inner detector and passing the loose track selection criteria defined in Table 6.3 (on the right). The normalisation of the histograms reflects the available Monte Carlo statistics. The $WH(H \rightarrow b\bar{b})$ sample has been used here.

In order to strongly reduce the contamination of the reconstructed secondary vertices by other charged particles arising from the jet, in particular those produced directly by the fragmentation process and thus originating from the primary interaction point, tracks stemming from the b - or c -hadron vertices, but not distinguishable from primary tracks, have to be suppressed, according to the selection procedure summarised in Section 6.7.4. This further reduces the b - and c -hadron decay products reconstruction efficiency, resulting in a compromise between the highest possible *efficiency* and a reasonable *purity*.

Algorithm	Topology	Population in b -jets	Association efficiency	Association purity
Inclusive vertex fit	1 inclusive B/D vertex	67 %	69 %	92 %
JetFitter	1 track + 1 track	2.3 %	58 %	77 %
	1 vertex	48.8 %	76 %	92 %
	1 vertex + 1 track	13.8 %	81 %	85 %
	2 vertices	5.9 %	84 %	89 %

Table 6.9: Track association efficiency and track association purity, as defined in the text, for the two considered secondary vertex finders, acting on different topologies, for which also the relative population in b -jets is shown. The $WH(H \rightarrow b\bar{b})$ sample has been used here.

In Table 6.9 the track association efficiencies and purities of the displaced tracks stemming from the b - or c - hadrons for the inclusive secondary vertex finder and for the JetFitter algorithms are shown. While the fit of a single inclusive b/c -hadron decay vertex allows to obtain a high purity, at the cost of starting to lose some tracks when the distance between the b - and c -hadron vertices starts to be relevant, the JetFitter algorithm is partially able to recover this inefficiency, thanks to the ability of reconstructing more complex decay chain topologies, at the cost of a slightly lower purity.

It is interesting to understand whether the difference seen between the behaviour of the fit of a single inclusive vertex and the decay chain fit can be understood in terms of the real distance between the weakly decaying b - and c -hadrons and in terms of the experimental resolutions of their respective tracks. First, the number of tracks arising from b - and c -hadron decays, after the loose track selection defined in Table 6.3 and after the tracks compatible with the primary vertex are removed according to the additional criteria defined in Section 6.7.1, are analysed, as shown in Table 6.10. As a first consequence, in 27.5% of the b -jets no vertex can be found, independently on the vertexing algorithm, while in around 5% and 15% of cases only respectively a single b - and c -vertex can be found. In these cases the decay chain fit has nearly no advantage and is fully equivalent to a single-vertex fit⁹.

Category	Population	Result
no selected tracks from B or D	27.5%	no vertex can be found
≥ 2 tracks from B	5.3%	single b -vertex
≥ 2 tracks from D	15.6%	single c -vertex
≥ 1 tracks from both B and D	51.5%	further analysed in Table 6.11

Table 6.10: Categories defined according to the amount of tracks from b - or c -hadron decays. The $WH(H \rightarrow b\bar{b})$ sample has been used here.

The remaining $\approx 52\%$ of b -jets is further analysed in Table 6.11. Here the various possible topologies of the true $b \rightarrow c$ -hadron decay chains are listed and their degree of compatibility with a single vertex. If the compatibility is better than 1%, the decay chain is assumed to be indistinguishable from a single vertex, and again the single-vertex fit can be considered perfectly equivalent to the decay chain fit, while if this compatibility is below 0.1% it is assumed that the two vertices can be efficiently separated in the decay chain fit. Considering all fractions where a single vertex fit is still good enough, this accounts for an additional fraction of 31% of the b -jets, which is given by summing up all cases where tracks from both b - and c -vertices are present but they cannot be efficiently distinguished ($\text{Prob}(\chi_{\text{single vtx}}^2 > 1\%)$).

⁹An exception is given by single-track vertices found in the decay chain fit out of the intersection of a single displaced track and of the jet axis, which however are only very weakly constrained and are therefore quite frequently found also in light-jets

Tracks from b -hadron	Tracks from c -hadron	Population	Single good vertex ($\text{Prob}(\chi_{vtx}^2) > 1\%$)	Single loose vertex ($\text{Prob}(\chi_{vtx}^2) > 0.1\%$)	Separation possible ($\text{Prob}(\chi_{vtx}^2) < 0.1\%$)
1	1	10.7%	77%	84%	16%
> 1	1	9.5%	66%	73%	27%
1	> 1	15.3%	56%	63%	37%
> 1	> 1	16.1%	49%	57%	43%

Table 6.11: Categories with ≥ 1 track originating from b and c -hadron vertices and degree of compatibility (in terms of $\text{Prob}(\chi_{vtx}^2)$) with a single vertex fit. The $\text{WH}(H \rightarrow b\bar{b})$ sample has been used.

By using a looser vertex probability requirement ($\text{Prob}(\chi_{\text{single vtx}}^2) > 0.1\%$), which is what both the inclusive vertex finder and JetFitter do, this fraction can be increased to $\approx 35\%$. In the remaining 17% of cases the decay chain fit has a clear advantage over the single-vertex fit, in two respects:

- In $\approx 15\%$ of b -jets a vertex plus an additional track or two vertices are found¹⁰. This additional track or vertex is lost in the inclusive single-vertex fit (this is the reason for the higher association efficiencies seen in Table 6.9).
- In another $\approx 2\%$ of cases, two single tracks along the b -flight axis are reconstructed¹¹. These candidates are completely lost in the single-vertex fit.

As a result the inclusive vertex finder should find a single inclusive vertex finder in $\approx 71\%$ of b -jets: in a subset of this fraction, $\approx 15\%$, it will however lose one or more tracks from b/c -decays. The decay chain fit is expected to find a single vertex in $\approx 56\%$ of b -jets, a single vertex and an additional track in $\approx 8\%$ of b -jets, two vertices in $\approx 7\%$ of b -jets and two single-track vertices in $\approx 2\%$. With respect to this study, some additional selection requirements are applied both in the inclusive secondary vertex and in the decay chain fit (according to Section 6.7.4), which are difficult to decouple from the finding procedure itself. This considered, these numbers can explain relatively well the fraction of different topologies in b -jets found by the inclusive vertex finder and by JetFitter (Table 6.9), with the clear exception of the relative fraction of '1 vertex' and '1 vertex + 1 track' topologies. In particular, in this case, it has been checked that some of the candidates which should go into the single-vertex category gain one track either from conversions, K_S^0 or Λ mesons, or directly from fragmentation, and migrate into the '1 vertex + 1 track' category.

6.8.2 Separation and resolution of b/c -decay vertices

In the case JetFitter finds a single vertex, its position will be typically in between the position of the real b -hadron vertex and the real c -hadron vertex. This is confirmed by Fig. 6.31, where the reconstructed vertex is expected to peak at 0 when it converges to the b -hadron decay vertex or at 1 in case it converges to the c -hadron vertex. Up to the effect of resolution, the vertex is always found between the b -hadron and c -hadron vertices and, in a significant fraction of cases, as expected, it converges either to the b - or c -hadron vertex.

In the case of all remaining decay chain topologies, two vertices, with one or more tracks, are found: analysing the same kind of distributions as before, it can be easily investigated how well the two b - and c -hadron vertices are separated in the decay chain fit (Fig. 6.32,

¹⁰This corresponds to the last three lines of Table 6.11, considering only the fraction of cases where $\text{Prob}(\chi_{vtx}^2) < 0.1\%$.

¹¹This corresponds to the first line of Table 6.11, considering only the fraction of cases where $\text{Prob}(\chi_{vtx}^2) < 0.1\%$.

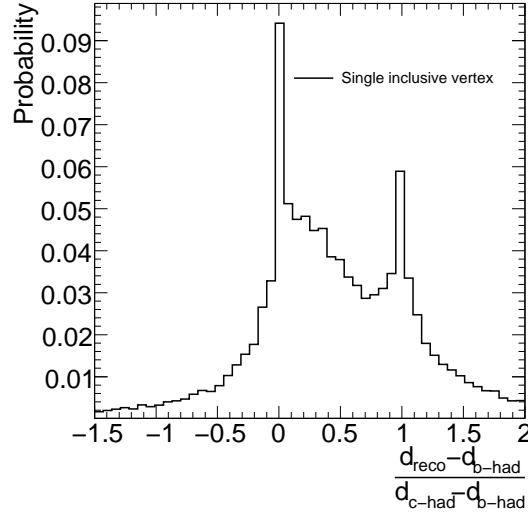


Figure 6.31: Distribution of distances of the reconstructed vertices to the b -vertex along the b -flight axis, divided by the distance between the b - and c -hadron vertices, again projected on the b -flight axis, for the cases where decay chain fit ends up in a single vertex.

left side). The vertex nearest to the primary vertex is taken as a candidate for the b -hadron vertex, the other one for the c -hadron vertex. Two peaks can be seen, which means that the vertices are in most cases separated fairly well and that the two vertex topology is correctly reconstructed by the decay chain fit. Both peaks have some tails towards the wrong decay vertex assignment: this is essentially due to the fact that the decay chain fit is not always able to cleanly disentangle the b - and c -vertices, so that from time to time some tracks get shared between the two vertices. In addition, for the c -hadron decay vertex candidates there is a small very sharp artificial peak at the b -hadron position: these candidates correspond to cases where there is no c -hadron vertex in the event, since the distance for such non existing c -hadron decay vertices is set to $d_{c\text{-hadr}} = -100$. In these few cases the b -hadron is a badly reconstructed vertex, in most of the times caused by one or more tracks from fragmentation, and the c -hadron position represents the real b -hadron decay vertex position. The subset of cases where the association of the b - and c -hadron candidates to their respective vertices is correct (purity of both vertices higher than 80%) is analysed in Fig. 6.32 (right side), where the two b - and c -hadron peaks can be perfectly disentangled: this corresponds to about half of the population with two vertex candidates.

It can be noticed that, in the case of the topologies with two vertices, the b -hadron position is biased towards smaller distances. The pulls of the identified b - and c -hadron decay vertex positions, for both the cases of one or two reconstructed vertices, is analysed in Fig. 6.33. In the case of one reconstructed vertex, the b -hadron pull has a big tail on the right due to the c -hadron contamination, while the c -hadron pull shows the same tail on the other side. In the case two vertices are found in the decay chain fit, the bias on the pull for the c -hadron vertex nearly disappears, while in the case of the b -hadron vertex a small bias is seen to smaller values, together with quite a significant left tail. The fact that, even if two vertices are correctly reconstructed in the jet, the pulls still do not have width equal to one, is most probably due to the fact that the hypothesis that the lateral displacement of the c -hadron decay vertex with respect to the b -hadron flight path is not perfectly realised, biasing the fit χ^2 during the minimisation procedure. The reason why this causes a bias on the b -

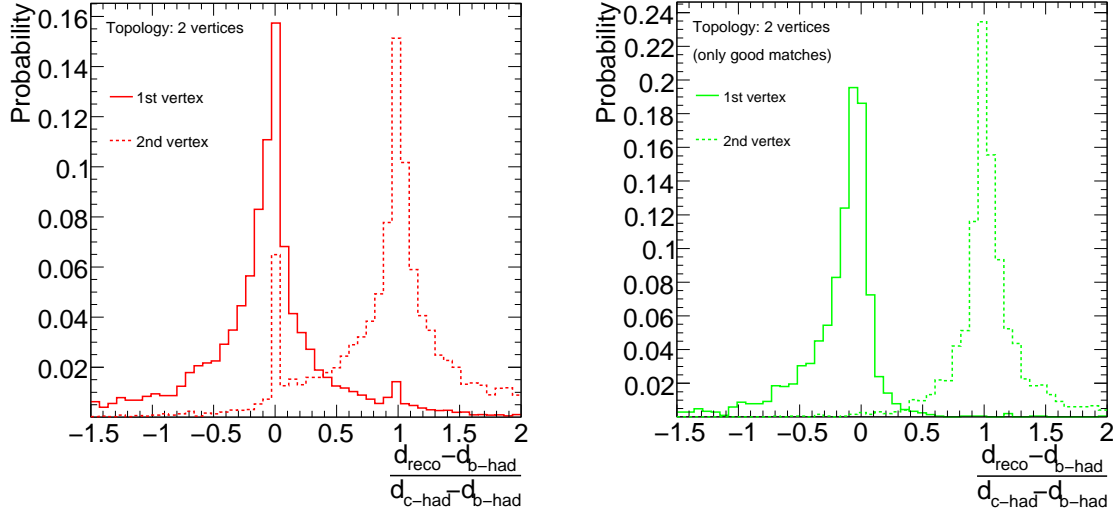


Figure 6.32: Distribution of distances of first and second vertex reconstructed in the decay chain fit to the b -vertex along the b -flight axis, divided by the distance between the b - and c -hadron vertices, again projected on the same axis. The same distribution is shown on the left for all topologies with two vertices with at least 1 track each, while on the right only candidates corresponding to topologies where the tracks are correctly associated to the real b - and c -hadron decay vertices are shown, as explained in the text.

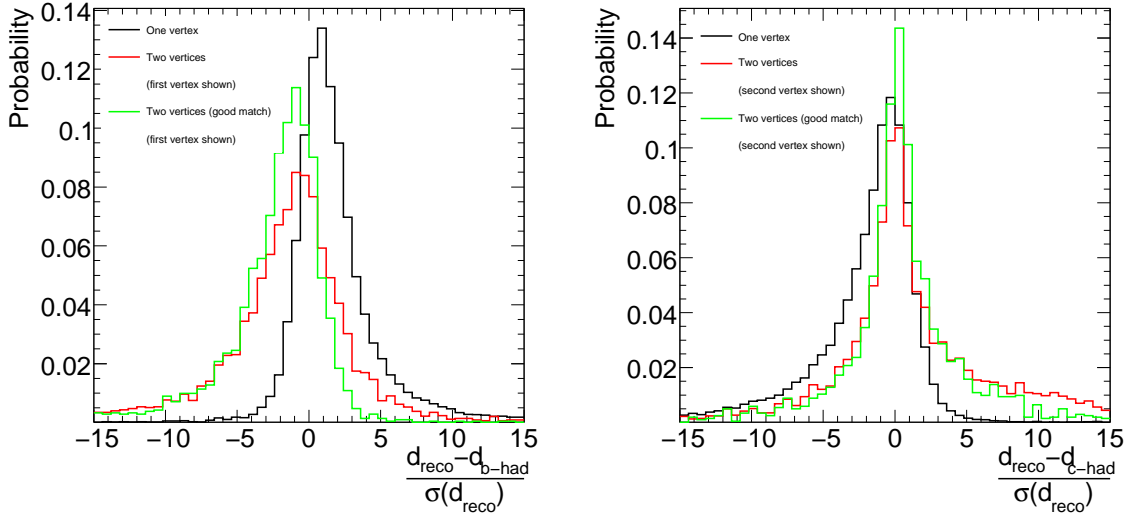


Figure 6.33: Distribution of residuals of the b - (left) and c -hadron (right) positions along the b -flight axis, divided by their respective fitted errors. In the case of the one vertex category, the single fitted vertex position is taken as a reference for both b - and c -hadron decay vertices. For the case of two reconstructed vertices, the first and second vertex, ordered in increasing distances from the primary vertex, are taken as the b - and c -hadron vertex candidates, respectively. The pulls are shown also for the case the b - and c -hadron vertices have been correctly associated to their respective tracks.

hadron decay position, and not on the c -hadron decay position, is most probably that tracks originating from the c -hadron have a longer lever arm (corresponding to the distance from the c -hadron decay position to the primary vertex) with respect to tracks originating from the b -hadron: so, if the lateral displacement of the c -hadron with respect to the b -hadron flight axis is beyond the experimental resolutions (Fig. 6.12), the fit will in most cases try to pull the b -hadron position towards the axis defined between the primary vertex and the c -hadron decay vertex position. It is worth stressing that for b -tagging applications the most important point is to distinguish the two vertices, in order to disentangle different topologies, and this - as demonstrated - works pretty well. However, in the case the JetFitter algorithm will be used also for future b -physics applications in ATLAS, it would be certainly worth to revise the fit formalism to include the error due to the lateral displacement of the c -hadron with respect to the b -direction directly into the fit.

In order to get an idea of the overall achievable vertex resolution, Fig. 6.34 shows the resolution of the b -hadron decay vertex position, as obtained on a sample of WH events, both for single vertex and two vertices topologies of the decay chain fit. For the latter category the weighted average position of the two vertices has been considered as the b -hadron position. In both cases a significant tail on the right in the distributions can be observed, which is due to tracks originating from the c -hadron decay vertex: this tail nearly disappears, as expected, if only vertices with tracks originating from real b -hadron decays are considered, as shown in the same figure.

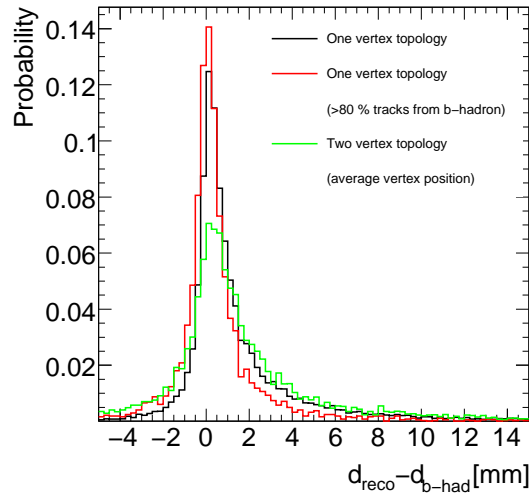


Figure 6.34: Distribution of residuals of the reconstructed three-dimensional flight length with respect to the real b -hadron flight path. This is shown first for events where the decay chain fit returns a single vertex (both inclusively and for real b -hadron vertices) and, then, for events where the decay chain fit returns two vertices: when two vertices are found, the resolution is computed by considering a weighted average position.

6.9 Secondary vertex b -tagging algorithms

In this section the application of the *JetFitter* vertex reconstruction algorithm to b -tagging will be illustrated. The information provided by the vertex reconstruction algorithms need to be condensed in few variables, typically chosen to be those which show the best discrimination between b - and light or c -jets. Then a method needs to be defined to combine these variables

into a single discriminating variable.

Before introducing the new methods defined for the JetFitter algorithm, the method already used in ATLAS for the single inclusive vertex reconstruction algorithm will be briefly mentioned, too.

6.9.1 Likelihood method based on a single inclusive vertex

The result of applying the inclusive vertex finding algorithm described in Section 6.5 on a jet is condensed in the secondary vertex, if eventually found, and in its associated tracks. Based on this, the following discriminating variables are defined:

1. The mass of the reconstructed secondary vertex M ;
2. The ratio of the total energy of tracks included into the secondary vertex to the total energy of all tracks in the jet R ;
3. Number of accepted (excluding V^0 s or hadronic interactions) two-track vertices in the jet N .

Since this b -tagging algorithm was mainly intended to be used in combination with the impact parameter based algorithms described in Section 6.3, the three-dimensional distance between the primary vertex and the secondary vertex is not used as an additional discriminating variable. In addition, the vertex reconstruction efficiency $\epsilon_{b,l}$ in b - or light-jets is also used as an additional information. Out of these four variables, a likelihood function is defined, both for b -jets and light-jets:

$$PDF_{b,l} = (1 - \epsilon_{b,l}) \cdot \delta(M, R, N) + \epsilon_{b,l} \cdot F_{b,l}(M, R)G_{b,l}(N) \quad (6.22)$$

The PDFs F and G for b - and light-jets are extracted from Monte Carlo, with the PDF F based on the two continuous variables M and R obtained through the use of a multidimensional smoothing method, to take the correlations between M and R properly into account. More details can be found in Ref. [75]. This algorithm is usually denoted in ATLAS as SV1.

6.9.2 Likelihood method based on JetFitter

The b -tagging algorithm implemented as an application of the decay chain fit operated by *JetFitter* is also based on separating b -jets from c - and light-quark (u, d, s) jets by means of the definition of a likelihood function.

The decay topology is described by the following discrete variables:

1. Number of vertices with at least two tracks.
2. Total number of tracks at these vertices.
3. Number of additional single track vertices on the b -hadron flight axis.

The vertex information is condensed in the following variables:

1. Mass: the invariant mass of all charged particle tracks attached to the decay chain (characterised by all multi- and single-track displaced vertices).
2. Energy fraction: the energy of these charged particles divided by the sum of the energies of all charged particles associated to the jet.

3. Flight length significance $\frac{d}{\sigma(d)}$: the weighted average vertex position along the flight axis divided by its error.

While the latter three variables can be used to define continuous PDFs, the former three can be used to define discrete PDFs. A finite set of categories can first be defined, which are defined by specific values of the three variables. A discrete PDF can be considered as a coefficient which is a function of the considered categories and expresses the probability, either for a b -, c - or light-jet, that the decay chain fit results in a certain category. While the simplest thing would be to define a coefficient by the product of three independent coefficients in the three different variables, like:

$$coeff(cat) = coeff(vertices) \cdot coeff(tracks) \cdot coeff(add. tracks), \quad (6.23)$$

this is not ideal since it would consider the three discrete variables as uncorrelated, which is absolutely unjustified here. However, considering all three discrete variables as completely correlated, would require a number of coefficients equal to $N_1 \cdot N_2 \cdot N_3$, where N_i is defined as the number of possible categories for variable i , which is impractical. The adopted solution is to group categories together when their distinction does not significantly increase the overall separation power and assuming, for specific values of the $N_{vertices}$ variable, the N_{tracks} and $N_{add. tracks}$ variables to be uncorrelated. The specific combinations for which coefficients are defined are illustrated in Table 6.12. This permits to reduce the number of considered coefficients for the topologies to 12, which, in terms of number of degrees of freedom, is equivalent to a histogram with 13 bins of a continuous variable.

$N_{vertices}$	$N_{add.tracks}$	N_{tracks}
0	0	
	1	
	2	
	≥ 3	
1	0	$\left\{ \begin{array}{l} 2 \\ 3 \\ 4 \\ 5-6 \\ \geq 7 \end{array} \right\}$
	1	
	≥ 2	
2		4
		≥ 5

Table 6.12: The table illustrates how the three discrete variables are combined into categories, corresponding to certain topological configurations of the decay chain fit. In the subset of cases where one single vertex with at least two tracks ($N_{vertices}$) is present in the decay chain, the other two variables, the number of single-track vertices ($N_{add. tracks}$) and the number of tracks at the vertex(ices) (N_{tracks}), are considered as uncorrelated: in this case their corresponding coefficients are just multiplied (which is indicated by the symbol \times)

The invariant mass is partially corrected for the presence of neutral particles at the b - or c -hadron decay vertex, which are not directly detectable in the tracking device. In fact, due to their presence, the sum of all momenta of the decaying charged particles \vec{p}_{ch} at the b/c -hadron decay vertex does not necessarily point into the b -flight direction. The relative p_T with respect to the b -flight direction, $p_{T,ch}$, can be exploited to take some of the missing momentum from the neutral particles into account. In the b -hadron rest frame, the momentum of the decay

vertices	tracks at vtx	add. single tracks	b -jet	light-jet	c -jet
0	0	0	15.6 %	83.4 %	54.2 %
0	0	1	12.0 %	11.6 %	19.3 %
0	0	2	2.0 %	0.5 %	1.2 %
0	0	≥ 3	0.4 %	0.0 %	0.1 %
1	2	any	25.4 %	3.9 %	17.6 %
1	3	any	19.5 %	0.5 %	6.0 %
1	4	any	10.9 %	0.1 %	1.1 %
1	5-6	any	7.3 %	0.0 %	0.2 %
1	≥ 7	any	0.9 %	0.0 %	0.0 %
1	any	0	49.8 %	4.1 %	22.5 %
1	any	1	12.1 %	0.0 %	2.2 %
1	any	≥ 2	2.1 %	0.0 %	0.2 %
2	4	any	2.1 %	0.0 %	0.2 %
2	≥ 5	any	4.0 %	0.0 %	0.1 %

Table 6.13: Population of the different topologies in b -, c - and light jets. In the case where a single vertex is found, not all categories are disjoint, as illustrated in Table 6.12. Due to this and due to the fact that the probability is by definition normalised to one, 12 coefficients are needed to describe the 14 listed topologies.

products can be separated into a charged component \vec{p}_{ch} and a neutral component \vec{p}_{neu} and the invariant mass of the b -hadron can be written as:

$$m_B = \sqrt{m_{ch}^2 + p_{T,ch}^2 + p_{\text{long},ch}^2} + \sqrt{m_{neu}^2 + p_{T,neu}^2 + p_{\text{long},neu}^2}, \quad (6.24)$$

where both the masses and the transverse momenta are invariant under Lorentz boost and are therefore equal to the quantities which can be measured in the detector frame. Since there is no way to access the longitudinal momentum of both charged and neutral components and since the p_T of the neutral and charged components must balance each other, the minimum m_B which the observed decaying particle must have is:

$$m_B = \sqrt{m_{ch}^2 + p_{T,ch}^2} + |p_{T,ch}|, \quad (6.25)$$

where both m_{ch} and $p_{T,ch}$ can be directly accessed. This definition for the mass, which is adopted in the present study, was exploited for the first time in Ref. [74]. While the use of this definition seems to improve the separation between b - and c -jets, the improvement in light-jet rejection is only marginal, since fake vertices in light jets have usually a significant amount of $p_{T,ch}$.

The energy fraction exploits the hardness of the b -quark fragmentation function. Most of the energy of the original b -quark is expected to be transmitted to the b -hadron, while, few tracks occasionally forming a vertex in a light jet usually carry a very low fraction of the original jet energy. The separation of b -jets from c -jets due to the different distributions of the energy fraction is smaller, since the fragmentation function of c -hadrons is still softer, but much closer to the fragmentation function for b -hadrons.

The use of these variables allows the definition of a likelihood function of the form:

$$L^{b,l,c}(x) = \text{coeff}(\text{cat}) \cdot \text{PDF}_{cat}(\text{Mass}) \cdot \text{PDF}_{cat}(\text{Energy fraction}) \cdot \text{PDF}_{cat}\left(\frac{d}{\sigma(d)}\right), \quad (6.26)$$

which has to be parametrised separately for each of the three different flavours.

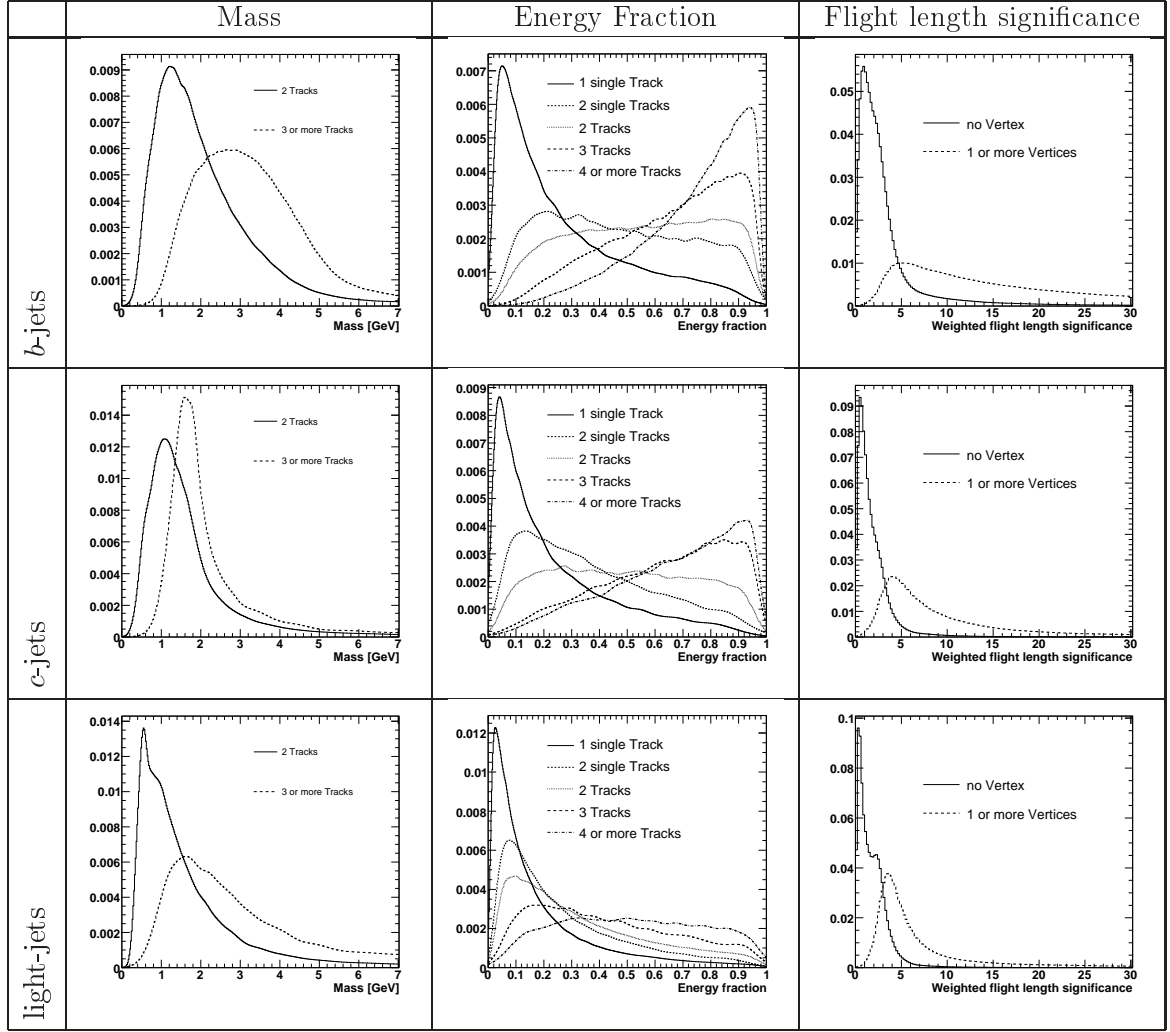


Figure 6.35: The $PDFs$ for the mass, the energy fraction and the flight length significance are shown, separately for the three different jet-flavours and split according to the decay chain topology found by *JetFitter*.

The information about the decay topology of the jet as reconstructed by *JetFitter* is represented by the category (the coefficient $coeff(cat)$) while the vertex information is contained in the probability distribution functions ($PDFs$). In order to reduce the correlations between the decay topology and vertex related variables and to increase the discrimination power, the $PDFs$ are made category dependent. This splitting of the $PDFs$ is done only when needed. The templates for the vertex variables are shown in Fig. 6.35 for all three flavours, while the populations of all defined categories are listed in Table 6.13. Each PDF was split independently into the categories it was found to be most correlated with, in order to keep the number of split $PDFs$ to be determined on Monte Carlo simulated events as low as possible. They were determined on a large set of $WH(m_H = 120 \text{ GeV})$ Monte Carlo events with $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$ and $H \rightarrow u\bar{u}$ and on $t\bar{t}$ Monte Carlo events.

The *JetFitter* based b -tagging algorithm can be either used as a stand-alone algorithm or in combination with pure impact parameter based algorithms (see Section 6.9.4).

6.9.3 Neural network discriminator based on JetFitter

Application of neural networks to classification problems

An explicit definition of the likelihood function in the multi-dimensional space of the input variables allows, by definition, to condense the maximum possible amount of information into a single likelihood function. By using a test statistics based on their likelihood ratio, the best possible separation between two different jet flavours can be obtained, as stated also by the Neyman-Pearson lemma. This is, however, only valid under the assumption that the full N -dimensional space of the likelihood function is filled without assuming subsets of such a space to be uncorrelated (unless they really turn out to be uncorrelated). As described in the previous chapter, in order to be able to calibrate the likelihood function with a reasonable amount of data, either from Monte Carlo simulation or later - hopefully - from data, this assumption must be necessarily partially made.

A discriminator based on the use of an artificial neural network can provide, under certain circumstances up to a good approximation, a discriminator which is equivalent to the likelihood ratio based on the full N -dimensional likelihood functions for the hypotheses considered. The goodness of such an approximation depends crucially on few assumptions, as will be explained briefly here.

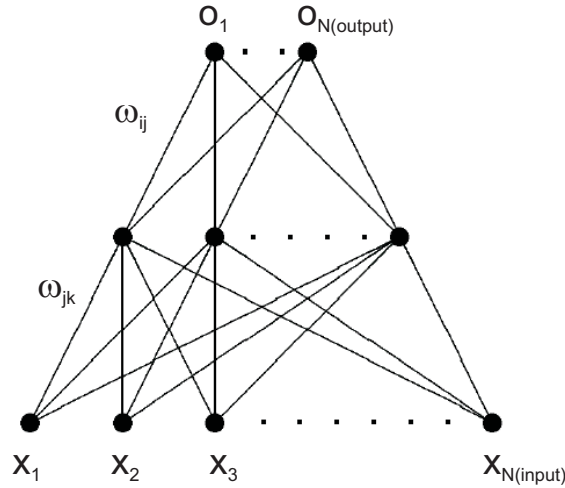


Figure 6.36: Layout of a neural network with one hidden layer.

Artificial neural networks are widely used in the high energy community and a general introduction is outside the scope of the present thesis. In general a neural network is an interconnected group of artificial neurons and represents a parametric mapping (function) between the N_{inp} input nodes and the N_{out} output nodes. Given a neural network with one input layer, one hidden layer and one output layer, as illustrated in Fig. 6.36, the value of the output node i can be written as:

$$F_i(\vec{x}) = g \left[\sum_j \omega_{ij} g \left[\sum_k \omega_{jk} x_k + \theta_j \right] + \theta_i \right], \quad (6.27)$$

where ω_{jk} and θ_j denote the weights and thresholds of the nodes j of the first hidden layer, ω_{ij} and θ_i denote the weights and the thresholds of the nodes of the output layer and the

sum runs over all input nodes k and over all nodes j of the hidden layer. In the following the function g is chosen to be:

$$g[x] = \frac{1}{2} (1 + \tanh(x)) = (1 + e^{-2x})^{-1}, \quad (6.28)$$

which is the sigmoid function. The weights ω_{ij} , ω_{jk} and the thresholds θ_i , θ_j fully characterise the behaviour of the neural network. An artificial neural network of this kind is usually called a multi-layer perceptron.

It can be demonstrated that such a multi-layer perceptron with one hidden layer and based on the sigmoid function, if a sufficient number of nodes in the hidden layer is used, can approximate any continuous, multivariate function to any desired degree of accuracy and that a failure in obtaining this is necessarily due to a poor choice of the parameters of the neural network, which are determined as a result of the training process.

The most delicate phase of the definition of a neural network is the training phase. This consists in varying the parameters $\omega_{ij}, \theta_i, \omega_{jk}, \theta_j$ in order to minimise a certain error measure, which is typically chosen to be the quadratic sum over all patterns of the residuals of the output values for the output node i with respect to the target values t_i , which would correspond to the correct classification:

$$E = \frac{1}{N_{pat}} \sum_{p=1}^{N_{pat}} \sum_{i=1}^{N_{out}} \left(o_i^{(p)} - t_i^{(p)} \right)^2 \quad (6.29)$$

In order to use the neural network for classification purposes, a number of output nodes equal to the number of possible classification categories is needed. In the case of b -tagging, these are three: b -, c - and light-jets. Then, for each pattern, the target values of the output nodes are set, depending on the true flavour of the jet corresponding to that specific pattern, as shown in Table 6.14.

Category	t_1	t_2	t_3
b -jet	1	0	0
c -jet	0	1	0
light-jet	0	0	1

Table 6.14: Target values used for the three output nodes o_i ($i = 1, 2, 3$) during the neural network training phase.

The Bayes theorem provides a strong theoretical handle to interpret the result of the training process. If $P(\vec{x}|C_i)$ is the probability to observe the set \vec{x} of values for the input variables, given a pattern belonging to the category C_i (this is equivalent to the likelihood $L_{C_i}(\vec{x})$ used in the likelihood based approach), while $P(\vec{C}_i|\vec{x})$ is the a-posteriori probability to have an event of category C_i given a set of values for the input variables \vec{x} , the Bayes theorem states that:

$$P(C_i|\vec{x}) = \frac{P(\vec{x}|C_i) \cdot P(C_i)}{\sum_i P(\vec{x}|C_i) P(C_i)}, \quad (6.30)$$

where $P(C_i)$ represents the a-priori probability to observe an event of the category $P(C_i)$. It can be demonstrated, provided a quadratic error measure of the type of Eq. 6.29 is used, the complexity of the network in terms of amount of nodes and layers is sufficient and the training process succeeds in finding the global minimum, that the values of the output nodes represent

the correct a-posteriori probability $P(C_i|\vec{x})$. The a-priori probabilities $P(C_i)$ adopted by the network will depend only on the relative amount of patterns of the various true flavours found in the training sample. For this statement to hold, it is crucial that a sigmoid function is used not only for the nodes of the intermediate layer, but also for all nodes of the output layer, in such a way that the output values are constrained to be inside the $[0, 1]$ interval.

The fact that the output nodes represent a good approximation for the a-posteriori probability $P(C_i|\vec{x})$ is equivalent to stating that the neural network provides a discriminator which is to a good approximation equivalent to the likelihood ratio if the full N_{inp} -dimensional likelihood function could be properly defined. In fact the ratio of the two a-posteriori probabilities $P(C_i|\vec{x})$ and $P(C_j|\vec{x})$ is essentially:

$$\begin{aligned} \frac{P(C_i|\vec{x})}{P(C_j|\vec{x})} &= \frac{P(\vec{x}|C_i) \cdot P(C_i)}{P(\vec{x}|C_j) \cdot P(C_j)} = \\ &= \text{LR}_{i/j}(\vec{x}) \frac{P(C_i)}{P(C_j)}. \end{aligned} \quad (6.31)$$

Since the a-priori probabilities $P(C_i)$ and $P(C_j)$ are positive definite, the ratio of the two a-posteriori probabilities is a monotonic increasing function of the likelihood ratio. A cut on the likelihood ratio based on two of the flavours (e.g. b over light) is therefore equivalent to a cut on the ratio of the two respective a-posteriori probabilities, defined by the neural network output nodes.

In order to achieve this equivalence between neural network based discriminator and N_{inp} -dimensional likelihood ratio, the training process must be able to find the global minimum of the error function in the space of all possible values for the network parameters $(\omega_{ij}, \theta_i, \omega_{jk}, \theta_j)$. The most classical method, adopted also in this thesis, is the so called *back-propagation* method. One of the main advantages of neural networks is that the function F_i is analytically defined, it is continuous and can be easily differentiated: this means that also the derivatives of the error function with respect to the network parameters can be easily computed. The back-propagation acts then iteratively, performing one update step after each pattern, trying to change the network parameters (generically called $\vec{\phi}$ here) in such a way to reduce the overall error, with the following learning rule:

$$\vec{\phi}_{it+1} = \vec{\phi}_{it} + \Delta\vec{\phi}_{it}, \quad (6.32)$$

where the increment $\Delta\vec{\phi}_{it}$ is defined as:

$$\Delta\vec{\phi}_{it+1} = -\eta \frac{\partial E_{it}}{\partial \vec{\phi}} + \alpha \Delta\vec{\phi}_{it}. \quad (6.33)$$

The first term tries to minimise the error function, moving the parameters $\vec{\phi}$ in the opposite direction of the gradient, where the η parameter is called learning rate and governs the speed of the training. The second term, called momentum term, is intended to avoid the minimisation procedure to fall in a local minimum, providing a certain *inertia* to the points moving in the network parameter space, based on the previous minimisation step. For all neural networks trained in this thesis, a value for the learning rate $\eta = 0.5$ and for the momentum term $\alpha = 0.2$ has been chosen.

While the *back-propagation* method for training is extremely simple, it turns out to be, at least for the application to most of the typical high energy physics classification problems, sufficiently fast and very robust in case of ill shaped and complex distributions for the input variables. More sophisticated methods directly or indirectly based on the use of second-order

derivatives of the error function are also available and allow in general a faster convergence, but they are less robust and can easily break down if the complexity of the network and of the input variable distributions increases. The FORTRAN based JetNet package [76] was used for the training of the multi-layer perceptrons used in this thesis. An interface, written in C++, was written, in order to integrate this package into ROOT, the most common analysis framework used in high energy physics.

Application to b -tagging

The most natural way of applying the neural network technique to b -tagging is to define a network topology with six input nodes, corresponding to the six variables of the output of the decay chain fit already defined in Section 6.9.2 and then used for the likelihood function, and with three output nodes, corresponding to the b -, c - and light-jet hypotheses.

There is one more dependence that has not been accounted for, which is the dependence of the input variables on the jet kinematics, defined by p_T and η . An example of this dependence was illustrated in Table 6.8, where it has been shown that the amount of decay chain fits belonging to a certain vertex topology in b -, c - and light-jets depends strongly on the jet kinematics.

There are two different consequences of this kinematic dependence:

- The p_T and η distributions of the light, b - and c -jets used for the training (Fig 6.37) are not identical. Due to the correlations with the input variables, a single global training would effectively include some of the information about the different kinematics of light and b - or c -jets in the training sample in the discriminating variable. This affects the training of the likelihood as well.
- The dependence on p_T and η of the input variables is not accounted for in the discriminator, potentially reducing the separation power.

The first problem can be solved by re-weighting the events in the Monte Carlo samples used for training in such a way to obtain the same distributions for the jet transverse momenta and pseudo-rapidity for all three different flavours. Since these two quantities, p_T and η , are correlated, the re-weighting is applied directly on the two-dimensional distribution. In order to avoid the need of a smoothing procedure, a two-dimensional histogram with coarse binning is used for the re-weighting. The p_T intervals have been chosen to be (15-25, 25-35, 35-50, 50-80, 80-120, 120-200, 200- ∞) GeV, while the $|\eta|$ intervals are (0.-0.7, 0.7-1.5, 1.5-2.5). This induces a two-dimensional re-weighting histogram with 21 bins. The one-dimensional distributions for jet p_T and η , before and after re-weighting, are shown in Fig. 6.37. After the re-weighting all three flavours have the same relative population in each of the $[p_T, \eta]$ bins. The weights have then been used in the neural network training procedure.

The second issue was dealt with by including the jet p_T and η variables directly into the neural network as additional input variables. In fact, since after the re-weighting procedure the two-dimensional p_T, η distribution is the same for all three jet flavours, the neural network, when trained on the Monte Carlo sample, cannot anymore use the p_T and η variables directly to try to separate the three different flavours, but it still can exploit the conditional dependence of the remaining variables on p_T and η (i.e. the correlations of these variables with jet p_T and η) to try to improve the discriminating power. In order to avoid any boundary effect due to the binning used for the re-weighting procedure, only a limited set of discrete values representing the bin in jet p_T and η on which the re-weighting has been applied are fed into the neural network, rather than directly the values of jet p_T and η .

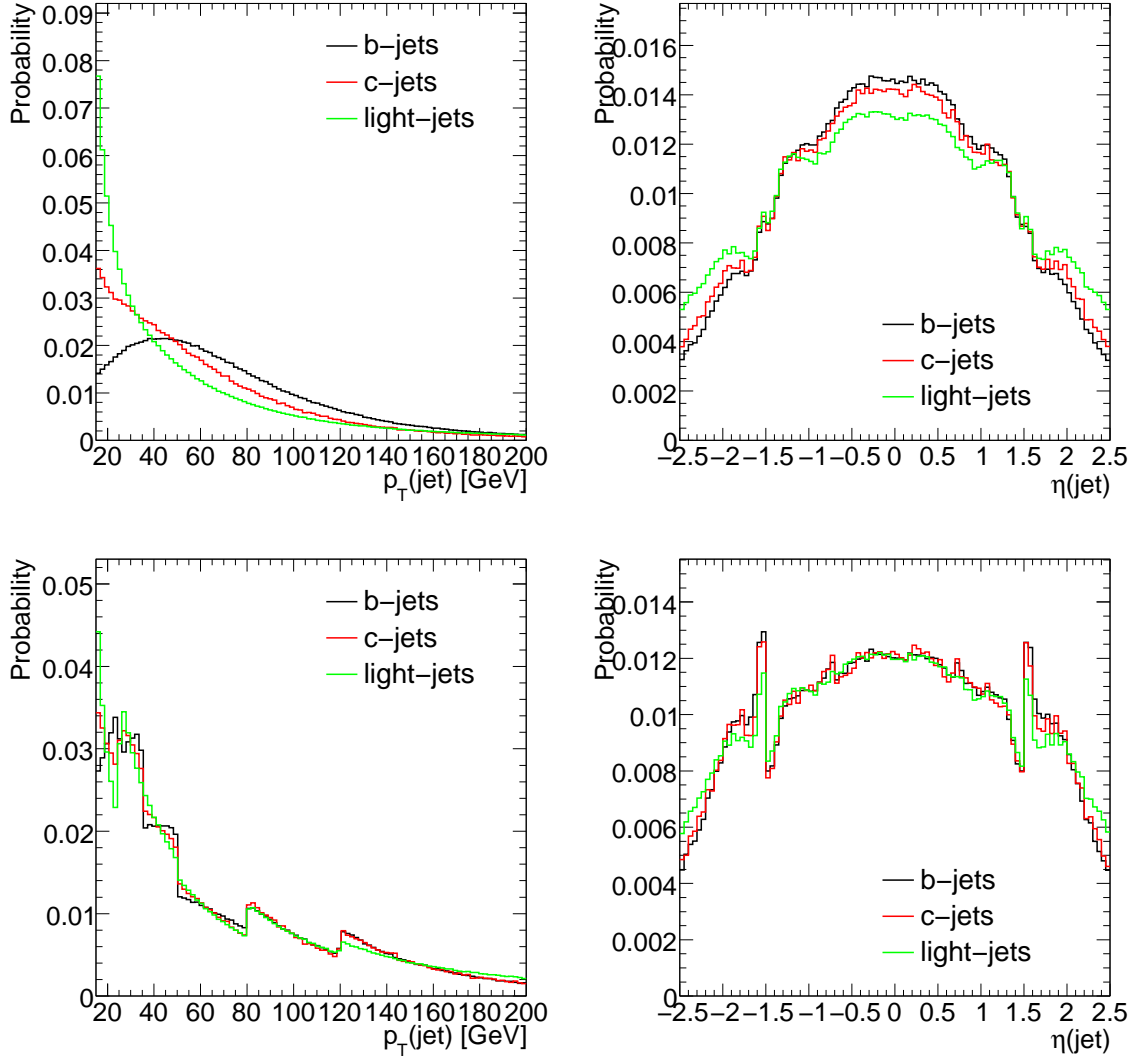


Figure 6.37: Transverse momentum (left column) and pseudo-rapidity (right column) of b -, c - and light jets, before (top row) and after (bottom row) the re-weighting procedure is applied.

The Monte Carlo sample used to train the neural network is divided into two parts, a training sample and a testing sample. The back-propagation training algorithm is used iterating over the training sample, aiming at minimising the error function defined on the training patterns, and then the error function is measured on the testing patterns. Every ten training cycles a testing cycle is performed. The evolution of the error function on both testing and training samples is monitored and a convergence criterion is defined to stop the training when the neural network is likely to have reached the global minimum:

- The training continues until the minimum in the testing error is reached.
- If the testing error increases, 300 further training iterations before stopping the procedure are performed, to monitor whether the increase in the testing error is only temporary due to statistical fluctuations.
- After each testing iteration, the neural network configuration of the weights and thresh-

olds is stored.

- The configuration corresponding to the minimum testing error is selected as the final configuration of the neural network.

This procedure avoids that the training is stopped too early due to statistical fluctuations in the training sample. It proved to work reliably on all tested networks. As an example, a training result is illustrated in Fig. 6.38, where the evolution of the error function is shown as a function of the training iteration.

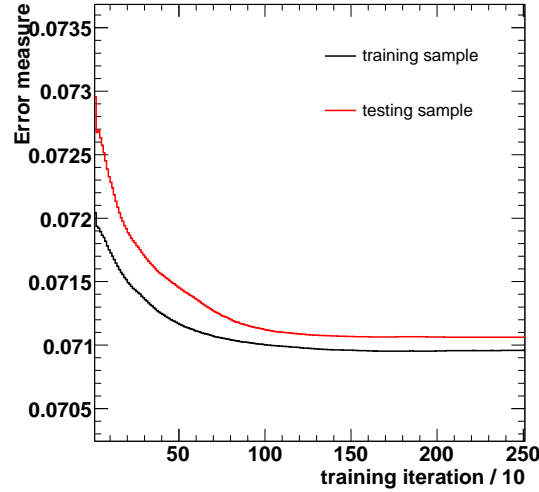


Figure 6.38: Error measure as a function of the training iteration during the training phase of the neural network, as computed on the testing and training samples.

Discriminating variables to select b -jets and reject either light- or c -jets can be defined according to the a-posteriori probabilities for a jet to be a b -jet, c -jet and light-jet, which are provided, according to the Bayesian interpretation of these discriminating variables, by the values of the output nodes:

$$\text{Discr}_l = \frac{P(b)}{P(b) + P(l)} \quad (6.34)$$

$$\text{Discr}_c = \frac{P(b)}{P(b) + P(c)} \quad (6.35)$$

The distributions for these variables are shown in Fig. 6.39, using as input values for the neural network the events in the testing sample. A distinctive signature that the training process was successful is that the distributions for the discriminating variables, Discr_l (Discr_c) for b - and for light (c -) jets cross at 0.5, where the a-posteriori probabilities for b - and for light (c -) jets have to be the same. This is the case both for Discr_l and Discr_c .

The a-posteriori probability for each of the three jet flavours, according to the Bayes theorem (Eq. 6.30), depends on the a-priori probabilities $P(b)$, $P(c)$ and $P(\text{light})$ that a b -, c - or light-jet is found, which, through the neural network training process, are set to the relative composition of the training sample in terms of b -, c - and light jets. Since the composition of the training sample has no direct physical interpretation, one is rather interested at defining the corresponding a-posteriori probabilities where the a-priori probabilities $P(b)$, $P(c)$

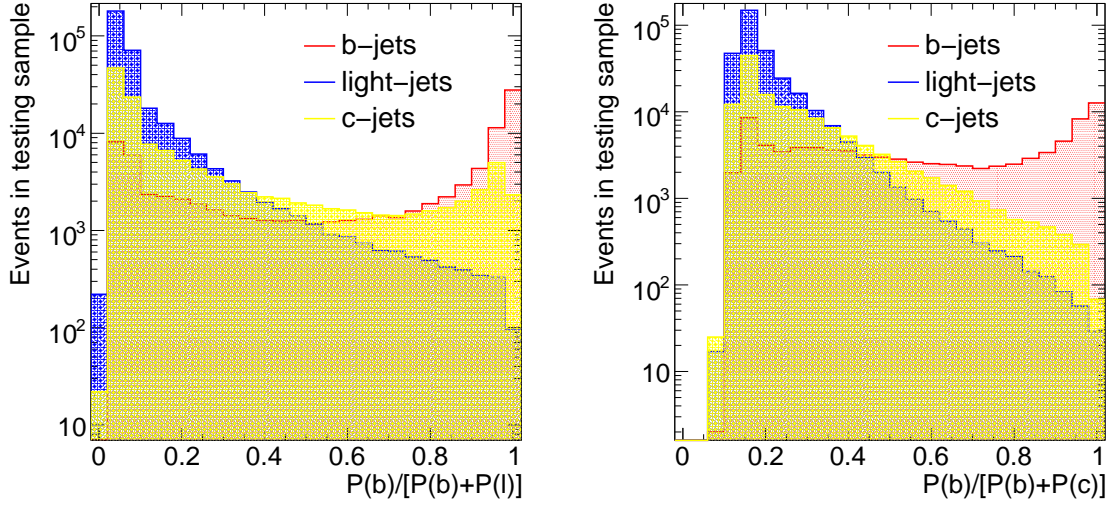


Figure 6.39: Distribution for the neural network discriminating variables against light- (on the left) and c -jets (on the right), as obtained from the testing sample.

and $P(\text{light})$ are set to $1/3$, which allows a probabilistic interpretation of the discriminating variables when b -tagging is later applied. This can be done by exploiting the fact that the a-posteriori probabilities over all possible flavours or classes C_i have to sum up to one. Using the notation introduced in Eq. 6.30, defining the $P_2(C_i|x)$ as the *correct* a-posteriori probabilities and $P_2(C_i)$ as the correct prior probabilities (to be set to $1/3$) and defining $P_1(C_i|x)$ as the a-posteriori probabilities obtained from training and $P_1(C_i)$ the prior probabilities corresponding to the flavour composition of the training sample, the new a-posteriori probabilities $P_2(C_i|\vec{x})$ can be obtained with the following formula:

$$P_2(C_i|\vec{x}) = \frac{P_1(C_i|\vec{x}) \frac{P_2(C_i)}{P_1(C_i)}}{P_1(C_i|\vec{x}) \frac{P_2(C_i)}{P_1(C_i)} + \sum_{j \neq i} P_1(C_j|\vec{x}) \frac{P_2(C_j)}{P_1(C_j)}}, \quad (6.36)$$

which extends the result for two output nodes which can be found in Ref. [77].

6.9.4 Combination with impact parameter based algorithms

The efficiency of a secondary vertex based b -tagging algorithm is intrinsically limited by the vertex (or decay chain) reconstruction efficiency. Therefore, to increase the b -tagging performance, this class of algorithms can be combined with a purely impact parameter based one, e.g. IP3D.

In the case of the use of the likelihood formalism, the combination can be obtained in a very simple way, if it is assumed that the likelihood functions based on the secondary vertex and impact parameter informations are uncorrelated. The combination of the two different discriminators is then just provided by the sum of their log-likelihood ratios,

$$\log(\text{LR}_{\text{comb}}) = \log(\text{LR}_{\text{IP}}) + \log(\text{LR}_{\text{SV}}) \quad (6.37)$$

which is equivalent to multiplying their PDFs. Even if the hypothesis of uncorrelated likelihood functions is not correct, a significant increase in performance can be obtained.

In the case of the neural network based b -tagging algorithm based on the output of the JetFitter decay chain fit, the combination with the impact parameter based algorithm is realised by adding the $\log(\text{LR}_{IP3D})$ as an additional input variable to the neural network.

6.10 Evaluation of b -tagging performance

In this section, the performance of the b -tagging algorithms described in this thesis is discussed. For the performance studies presented in the following, an iterative cone algorithm with a cone size of $\Delta R = 0.4$ using topological clusters of energy in the calorimeter as input, has been used. In addition, only jets within the acceptance of the tracking detectors ($|\eta| < 2.5$) and with a reconstructed transverse momenta above 15 GeV are used in the performance studies.

6.10.1 Definitions of efficiency and rejection

The efficiency to tag a jet of flavour q as b -jet, ϵ_q , is defined as:

$$\epsilon_q = \frac{\text{Number of jets of real flavour } q \text{ tagged as } b}{\text{Number of jets of real flavour } q}. \quad (6.38)$$

Usually ϵ_b is called *tagging efficiency* and ϵ_{udsc} *mistagging rate*. The inverse of the mistagging rate $r_{udsc} = 1/\epsilon_{udsc}$ is called *b -tagging rejection*. The assignment of a certain flavour to a jet in the Monte Carlo simulation is not unambiguously defined. The following definition has been introduced: the jet flavour is the flavour of the heaviest quark after gluon radiation and splitting within a cone of $\Delta R < 0.3$ around the jet direction. To test the pure algorithmic performance of the algorithms, a procedure called “purification” can be applied. Here, a light quark jet is only considered if there is no heavy parton (b - or c -quark) or τ lepton within a cone of $\Delta R = 0.8$ around the jet direction.

6.10.2 Physics events samples used

The following fully simulated Monte Carlo samples have been used for the performance studies described in this section:

- Higgs boson production in association with a W boson. The W boson was forced to decay into a muon and its anti-neutrino, $W \rightarrow \mu\nu_\mu$, the Higgs boson was forced to decay into pairs of b -, c - or u -quarks: $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$, $H \rightarrow u\bar{u}$. This ensures that the jets of different quark flavour have very similar kinematics. To cover a wide range of jet transverse momenta, samples with Higgs boson masses of $m_H = 120$ GeV and $m_H = 400$ GeV have been produced.
- Events with pairs of top quarks, $t\bar{t}$, where one or both W bosons decayed leptonically.
- Events with pairs of top quarks and at least two additional jets: $t\bar{t}jj$. These events show a large hadronic activity with a high probability of overlapping jets.

The performance was evaluated by using the b -tagging algorithms and calibration files as implemented in release 14.0.0 of the ATLAS reconstruction software.

6.10.3 Performance in rejecting light-jets

The performance of the b -tagging algorithm developed in this thesis, based on the decay chain fit performed by *JetFitter*, was analysed by comparing it to the other inclusive secondary vertex based algorithm available in ATLAS (SV1). Both algorithms are considered also in combination with IP3D, the most powerful impact parameter based algorithm available in ATLAS. For JetFitter also the results based on the use of the neural network are considered.

The b -tagging performance is therefore analysed for the following algorithms:

- IP3D (algorithm based on transverse and longitudinal impact parameter, presented in Section 6.3)
- SV1 (based on the inclusive single-vertex fit presented in Section 6.5)
- JetFitter (likelihood based on the decay chain fit presented in Section 6.6)
- SV1+IP3D (likelihood combination of SV1 and IP3D)
- JetFitter+IP3D (likelihood combination of *JetFitter* and IP3D)
- JetFitter (NN) (neural network based on the decay chain fit)
- JetFitter+IP3D (NN) (neural network based on the combination of JetFitter and IP3D)

Table 6.15 shows the rejection rates against light quark jets for fixed b -tagging efficiencies, for different tagging algorithms and data samples. In the case of the neural network based approach, the performance corresponding to the discriminator rejecting light-quark jets ($P_b/(P_b + P_l)$) is shown. The errors quoted for all rejection values are due to the limited statistics of the Monte Carlo samples.

Figure 6.40 shows the rejection of light-quark (u, d, s) jets, r_u , versus the b -tagging efficiency ϵ_b for the pure secondary vertex based algorithms and for the impact parameter based algorithm, based on the use of the $t\bar{t}$ data sample only. It can be seen that the maximum achievable b -tagging efficiency of the vertex based algorithms is limited by the efficiency to reconstruct a displaced vertex or b -hadron decay topology. It is higher in the case of *JetFitter* which imposes only very soft requirements on the topology. The best performance over the full range of b -tagging efficiencies is provided by the b -tagging algorithm based on the *JetFitter* vertexing algorithm and on the use of a neural network discriminator. There is also a significant difference between *SV1* and *JetFitter*, the latter one showing better performance. It has to be noted, however, that *JetFitter* explicitly uses lifetime information (the flight distance significance), which is not used by *SV1* to decorrelate the algorithm better from *IP3D*. At the same time in *SV1* the correlations between the mass and energy fraction variables are explicitly considered, while in the likelihood approach based on *JetFitter* they are not; they are, however, considered in the neural network based approach, where the best overall performance is obtained. The right plot of Figure 6.40 exhibits a steep structure at b -tagging efficiencies of about 52%. This can be related to a sudden drop in the rejection rate of the *IP3D* algorithm for this b -tagging efficiency, as can be seen in the left part of Figure 6.40. This behaviour can be explained mainly by the presence of jets with only a single track associated to them.

Fig. 6.41 shows the rejections against light quark jets for the secondary vertex based tagging algorithms after combination with the *IP3D* algorithm, again compared with *IP3D*. This combination results in a significant increase in rejection power. Again, *JetFitter+IP3D*, in particular when based on a neural network discriminator, shows the best overall performance.

Sample	ϵ_b	IP3D	SV1	JetFitter (LH)	JetFitter (NN)
WH(120)	50% raw	162 ± 3	154 ± 2	164 ± 3	224 ± 4
	60% raw	53 ± 1	43 ± 0	52 ± 0	63 ± 1
	70% raw	19 ± 0	-	16 ± 0	21 ± 0
WH(120)	50% purified	171 ± 3	163 ± 3	168 ± 3	230 ± 5
	60% purified	55 ± 1	44 ± 0	52 ± 0	64 ± 1
	70% purified	19 ± 0	-	16 ± 0	22 ± 0
WH(400)	50% raw	161 ± 13	142 ± 2	141 ± 2	172 ± 3
	60% raw	54 ± 0	55 ± 1	56 ± 1	57 ± 1
	70% raw	19 ± 0	-	20 ± 0	20 ± 0
WH(400)	50% purified	187 ± 3	154 ± 2	143 ± 2	175 ± 3
	60% purified	60 ± 1	58 ± 1	56 ± 1	58 ± 1
	70% purified	21 ± 0	-	20 ± 0	20 ± 0
$t\bar{t} + t\bar{t}jj$	50% raw	236 ± 2	200 ± 1	257 ± 2	326 ± 3
	60% raw	67 ± 0	77 ± 0	89 ± 0	99 ± 0
	70% raw	22 ± 0	-	27 ± 0	31 ± 0
$t\bar{t} + t\bar{t}jj$	50% purified	316 ± 3	291 ± 2	288 ± 2	376 ± 4
	60% purified	76 ± 0	95 ± 0	94 ± 0	107 ± 1
	70% purified	24 ± 0	-	28 ± 0	33 ± 0

Sample	ϵ_b	SV1+IP3D	JetFitter+IP3D (LH)	JetFitter+IP3D (NN)
WH(120)	50% raw	404 ± 10	459 ± 12	616 ± 19
	60% raw	97 ± 1	111 ± 1	143 ± 2
	70% raw	25 ± 0	29 ± 0	35 ± 0
WH(120)	50% purified	455 ± 13	495 ± 14	674 ± 23
	60% purified	102 ± 1	114 ± 1.6	150 ± 2.4
	70% purified	25 ± 0	30 ± 0	36 ± 0
WH(400)	50% raw	348 ± 8	386 ± 9	442 ± 11
	60% raw	121 ± 2	123 ± 2	134 ± 2
	70% raw	37 ± 0	39 ± 0	40 ± 0
WH(400)	50% purified	410 ± 11	418 ± 11	483 ± 14
	60% purified	131 ± 2	129 ± 2	141 ± 2
	70% purified	39 ± 0	39 ± 0	41 ± 0
$t\bar{t} + t\bar{t}jj$	50% raw	469 ± 5	658 ± 7	798 ± 10
	60% raw	165 ± 1	199 ± 1	233 ± 2
	70% raw	41 ± 0	45 ± 0	61 ± 0
$t\bar{t} + t\bar{t}jj$	50% purified	924 ± 14	980 ± 15	1164 ± 19
	60% purified	237 ± 2	241 ± 2	291 ± 2
	70% purified	47 ± 0	48 ± 0	69 ± 0

Table 6.15: The rejection against light quark jets for fixed b -tagging efficiencies of 50%, 60% and 70% for different data samples and for the standalone secondary vertex based b -tagging algorithms (top table) and for the combined b -tagging algorithms (bottom table), without and with applying the purification procedure, denoted as *raw* and *purified*, respectively.

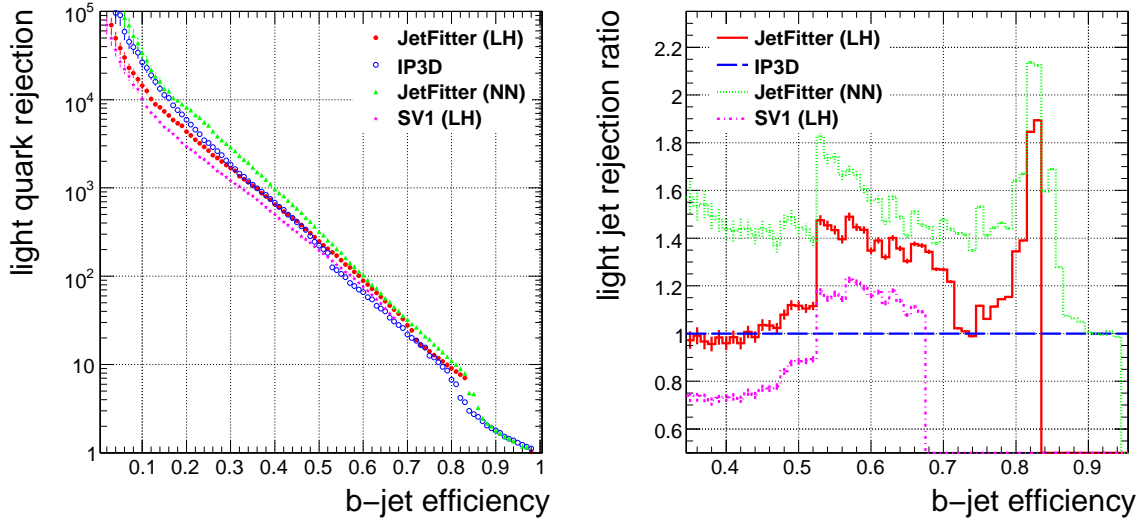


Figure 6.40: Left: Light jet rejection versus b -tagging efficiency for $IP3D$, $SV1$ and $JetFitter$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $IP3D$ for $SV1$ and $JetFitter$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ sample. No purification of light quark jets has been applied.

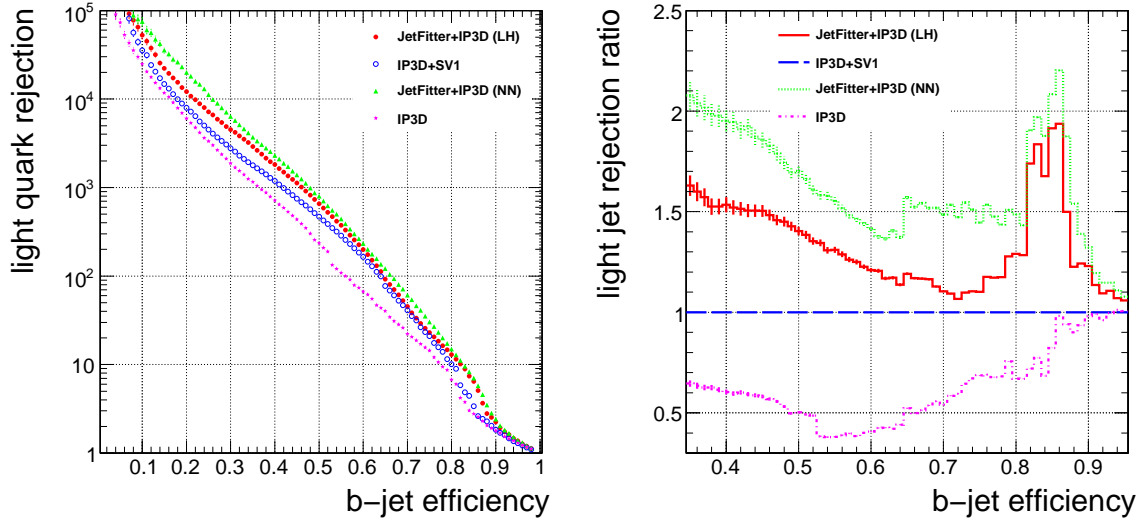


Figure 6.41: Left: Light jet rejection versus b -tagging efficiency for $IP3D$, $SV1+IP3D$ and $JetFitter+IP3D$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $SV1+IP3D$ for $IP3D$ and $JetFitter+IP3D$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ sample. No purification of light quark jets has been applied.

The algorithmic improvement of the JetFitter algorithm is particular evident when applied on the $t\bar{t}$ and $t\bar{t}jj$ samples, it is still significant for the WH(120) sample and much less pronounced for the WH(400) sample, where the average jet transverse momentum is around 200 GeV. Most of these differences can be understood by looking more in detail at jet kinematics (in particular different distributions of the jet transverse momenta and pseudo-rapidity) and at the level of contamination of light jets due to nearby lying b - or c -jets.

Dependence on jet kinematics

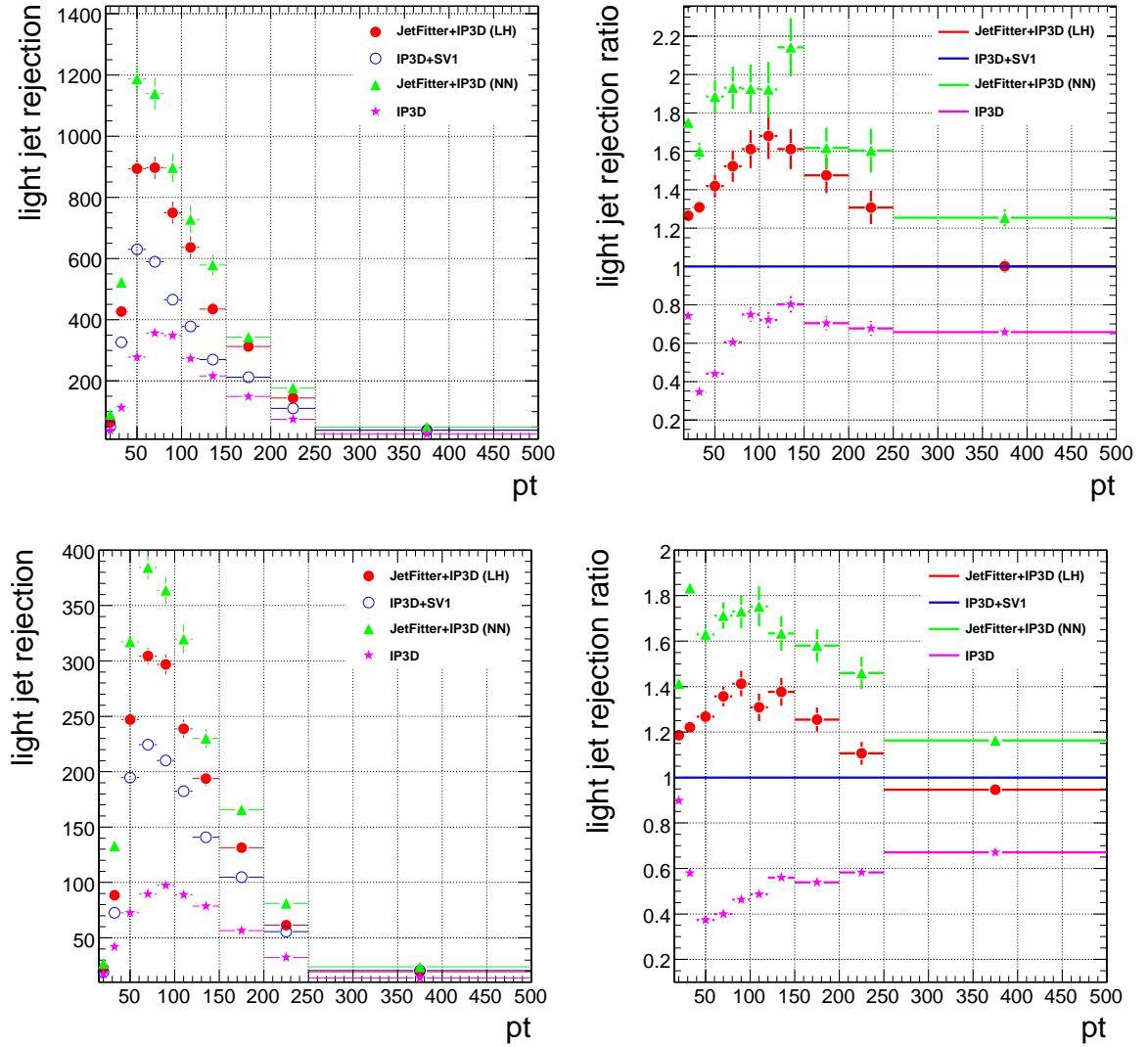


Figure 6.42: Left: Light jet rejection for a fixed b -tagging efficiency of 50% (top) or 60% (bottom) as function of the jet transverse momentum p_T for $IP3D$, $SV1+IP3D$ and $JetFitter+IP3D$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $SV1+IP3D$ for $IP3D$ and $JetFitter+IP3D$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples. No purification of light quark jets has been applied. The jet p_T is expressed in GeV.

The dependence of the performance on the jet kinematics is of particular importance. Figs. 6.42 and 6.43 show the rejection against light quark jets for fixed b -tagging efficiencies

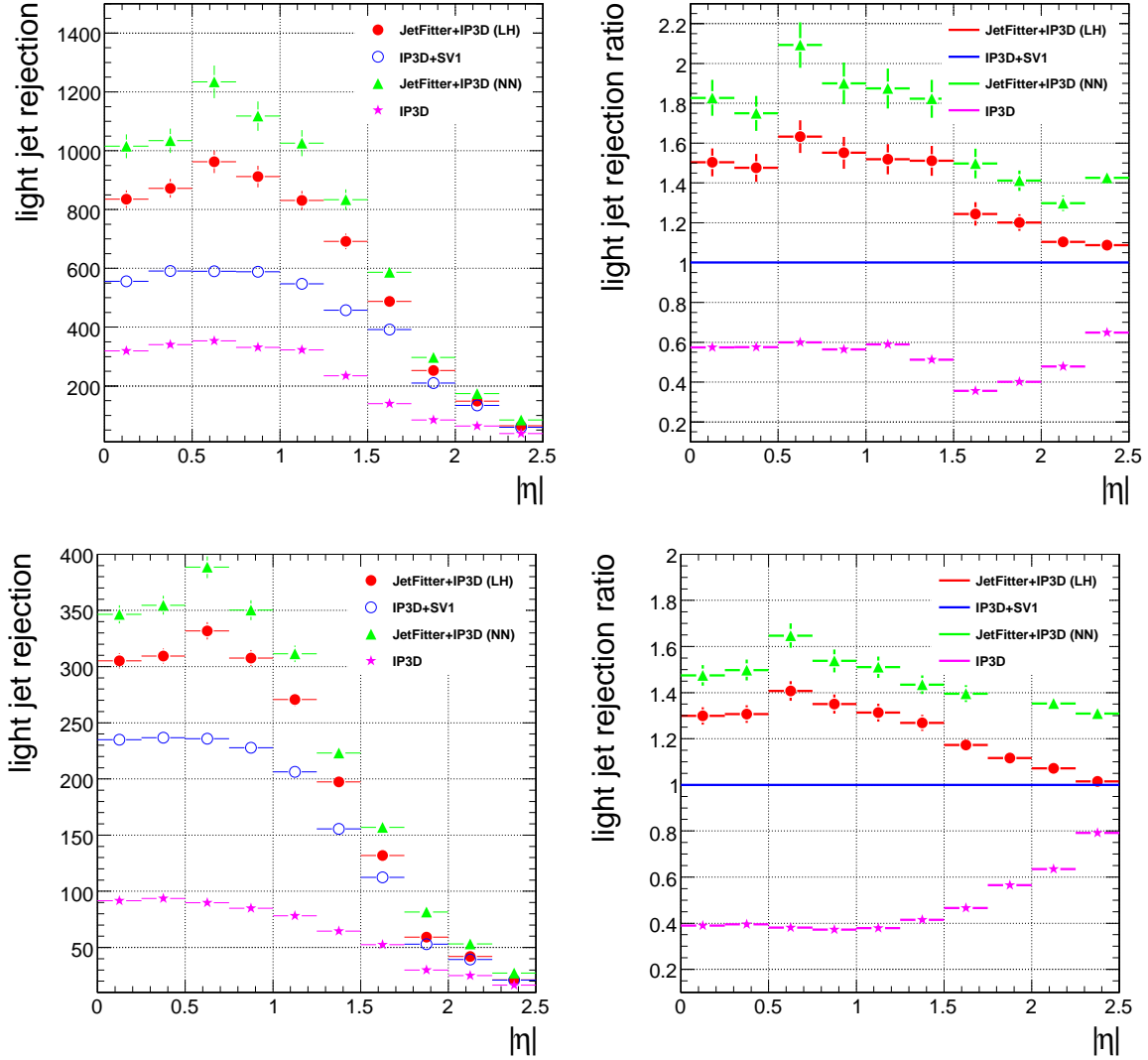


Figure 6.43: Left: Light jet rejection for a fixed b -tagging efficiency of 50% (top) or 60% (bottom) as function of the jet pseudo-rapidity η for $IP3D$, $SV1+IP3D$ and $JetFitter+IP3D$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $IP3D+SV1$ for $IP3D$ and $JetFitter+IP3D$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples. No purification of light quark jets has been applied.

of 50% and 60% versus the jet transverse momentum and jet pseudo-rapidity, respectively. The performance plots as a function of the jet transverse momentum provides also a more fair comparison among the various b -tagging algorithms, since, when comparing the performance of the inclusive samples, it should be considered, as already shown in Fig. 6.37, that the transverse momenta distributions for the light- and c - or b -jets are considerably different. This does not only affect the absolute performance, but can eventually even artificially increase the performance of the likelihood based algorithms with respect to the neural network based ones, since the likelihood functions were actually extracted using samples where the light and c - and b -jets are affected by the same behaviour.

Improvements of up to a factor of around two in light quark rejection can be obtained, both at 50% and at 60% b -tagging efficiency, by using the neural network discriminator based on

the *JetFitter* output, with respect to the combined algorithm based on *IP3D+SV1*. Especially in the kinematic regions where the best b -tagging performance is attained, the improvements are particularly significant. The reasons for this behaviour are manifold. For larger pseudo-rapidities, the particles pass through more material and thus the detector resolution is worse. Another reason for the degradation of the longitudinal impact parameter at large pseudo-rapidities is the dramatic increase of the extrapolation distance from the b -layer (first pixel layer) hit to the vertex. The degradation for high jet transverse momenta is caused by an increased fragmentation multiplicity, resulting in an increased combinatorics in the secondary vertex finding stage when trying to find the tracks stemming from the b -hadron decay. Furthermore, the pattern recognition in the track reconstruction becomes more difficult in the very dense environment of jets with very large transverse momenta. The steep fall for low jet transverse momenta is mainly due to the strongly enhanced multiple scattering of low momentum charged particle tracks, leading to significantly degraded impact parameter resolutions. The observed dependence of the b -tagging performance on the jet kinematics can be related to the performance of the secondary vertex reconstruction as discussed in Section 6.8. It can be seen there, that the secondary vertex reconstruction efficiency in b -quark jets drops in the same kinematic regions as the resulting b -tagging performance, with an increased rate of (fake) vertices in light quark jets.

Contamination from nearby heavy flavour jets

Apart from the kinematic dependence, the b -tagging performance also depends critically on the contamination of light-quark jets by displaced tracks stemming from nearby b - or c -quark jets, as can already be concluded comparing the b -tagging performance before and after application of the purification procedure in Table 6.15. It is however worth looking at this dependence in more detail, analysing the b -tagging performance as a function of the distance ΔR of the light quark jets to the nearest b - or c -quark or τ lepton.

ϵ_b	Algorithm	[0.3; 0.35]	[0.35; 0.4]	[0.4; 0.45]	[0.45; 0.5]	[0.5; 0.6]	[0.6; 0.7]
50%	<i>SV1</i>	9.1 ± 0.1	21 ± 1	70 ± 3	188 ± 11	362 ± 15	369 ± 13
	<i>JetFitter</i>	26.6 ± 0.6	86 ± 4	203 ± 14	374 ± 32	431 ± 20	418 ± 16
60%	<i>SV1</i>	5.7 ± 0.1	11 ± 0	25 ± 1	54 ± 2	109 ± 3	119 ± 2
	<i>JetFitter</i>	13.8 ± 0.2	36 ± 1	71 ± 3	97 ± 4	115 ± 3	113 ± 2

ϵ_b	Algorithm	[0.3; 0.35]	[0.35; 0.4]	[0.4; 0.45]	[0.45; 0.5]	[0.5; 0.6]	[0.6; 0.7]
50%	<i>SV1+IP3D</i>	12 ± 0.2	37 ± 1	166 ± 11	484 ± 47	989 ± 70	1059 ± 74
	<i>JetFitter+IP3D</i>	29 ± 1	114 ± 6	379 ± 38	915 ± 123	1205 ± 94	1287 ± 87
60%	<i>SV1+IP3D</i>	7.3 ± 0.1	16 ± 0	48 ± 1	111 ± 5	240 ± 8	261 ± 8
	<i>JetFitter+IP3D</i>	13.4 ± 0.2	40 ± 1	112 ± 6	191 ± 11	267 ± 10	283 ± 9

Table 6.16: The rejection against light-jets for different b -tagging efficiencies (ϵ_b) in various intervals $[min; max]$ of the difference in ΔR of the light quark jets to the nearest b -, c - or τ -jet. For *JetFitter* only the neural network based algorithm is considered. The top table shows the result on the secondary vertex based b -tagging algorithms used stand-alone, while the bottom table shows the result for the combined algorithms.

Table 6.16 shows, both for the pure secondary vertex based algorithms and after combination with the *IP3D* algorithm, that the more the light quark jets are contaminated by b - or c -hadron decay products (smaller values of ΔR), the stronger the degradation of the b -tagging performance is. At very small values of ΔR the light quark-jets can be barely distinguished from the nearby heavy flavour jet. A small kinematic bias (slightly different p_T and η distributions of the jets for different ΔR intervals) should also be taken into account when interpreting these results.

There is a noticeable difference in the behaviour of *JetFitter* compared to *SV1*, the second being less robust against building up vertices which catch up contributions from nearby heavy flavour jets. The additional requirement in *JetFitter* that the displaced tracks have to intersect the hypothetical b -hadron flight direction, which is constrained by the jet direction within the allowed error, makes this algorithm more efficient in discarding tracks originating from nearby jets.

6.10.4 Performance in rejecting charm-jets

Rejecting c -jets is significantly more difficult than rejecting light-jets, due to the presence of c -hadrons with real lifetime in c -jets. The performance of the secondary vertex based algorithms in rejecting c -jets is presented in Table 6.17 (top table) and in Fig. 6.44. The combination with the impact parameters significance based b -tagging algorithms does not yield a significant improvement if compared with the purely secondary vertex based b -tagging algorithms, as shown in Table 6.17 (bottom table) and Fig. 6.45.

Sample	ϵ_b	IP3D	SV1	JetFitter (LH)	JetFitter (NN)
WH(120)	50%	13 ± 0	14 ± 0	14 ± 0	15 ± 0
	60%	7.3 ± 0.1	7.4 ± 0.1	7.7 ± 0.1	7.5 ± 0.1
	70%	4.3 ± 0.0	-	4.4 ± 0	4.5 ± 0.0
WH(400)	50%	11 ± 0	13 ± 0	12 ± 0	13 ± 0
	60%	6.5 ± 0.1	6.8 ± 0.1	7.2 ± 0.1	7.0 ± 0.1
	70%	3.9 ± 0.0	-	4.3 ± 0.0	4.2 ± 0.0
$t\bar{t} + t\bar{t}jj$	50%	10 ± 0	12 ± 0	11 ± 0	13 ± 0
	60%	6.4 ± 0.0	6.5 ± 0	6.8 ± 0	7.5 ± 0.0
	70%	4.0 ± 0.0	-	4.3 ± 0	4.6 ± 0.0

Sample	ϵ_b	SV1+IP3D	JetFitter+IP3D (LH)	JetFitter+IP3D (NN)
WH(120)	50% raw	16 ± 0	15 ± 0	16 ± 0
	60% raw	8.6 ± 0.1	8.5 ± 0.1	8.7 ± 0
	70% raw	4.8 ± 0.1	4.9 ± 0.1	5.0 ± 0.1
WH(400)	50% raw	14 ± 0	14 ± 0	14 ± 0
	60% raw	8.0 ± 0.1	8.0 ± 0.1	8.1 ± 0.1
	70% raw	4.8 ± 0.0	4.8 ± 0.0	4.9 ± 0.0
$t\bar{t} + t\bar{t}jj$	50% raw	12 ± 0	12 ± 0	13 ± 0
	60% raw	7.6 ± 0.0	7.5 ± 0.0	7.9 ± 0
	70% raw	4.6 ± 0.0	4.6 ± 0.0	4.8 ± 0

Table 6.17: The rejection against charm quark jets for fixed b -tagging efficiencies of 50%, 60% and 70% for different data samples and for the standalone secondary vertex based b -tagging algorithms (top table) and for the combined b -tagging algorithms (bottom table).

A more significant improvement can be obtained at the cost of decreasing the light-jet rejection, by explicitly calibrating the b -tagging algorithms to reject c -jets instead of light-jets. The additional information on the decay chain topology can eventually also increase the performance in rejecting c -jets, at least in the region of low b -jet efficiencies. In the case of the likelihood formalism, in order to obtain the best combined rejection, also a dedicated calibration for the impact parameter based algorithm is needed. Since this calibration is missing at the moment in ATLAS, it is difficult to define a combined discriminator against

c -jets for IP3D+JetFitter based on the likelihood formalism.

A dedicated discriminator against c -jets was obtained, as described in Section 6.9.3, by adding the likelihood ratio of $IP3D$ (discriminating against light-jets) to the neural network, but training it simultaneously against c - and light-jets by using three output nodes in the neural network (corresponding to b -, c - and light-jets). Even if this is not yet optimal, it can already increase significantly the discrimination power against c -jets, as shown in Table 6.18 and Fig. 6.46. As previously mentioned, the discriminating variable against light jets is $P(b)/(P(b) + P(l))$, whereas the one against c -jets is $P(b)/(P(b) + P(c))$.

Rejection against	ϵ_b	$IP3D+SV1$	$JetFitter+IP3D$ (likelihood)	$JetFitter+IP3D$ (NN against light)	$JetFitter+IP3D$ (NN against charm)
light-jets	40%	1186 ± 18	1820 ± 35	2314 ± 50	1025 ± 15
	50%	469 ± 5	658 ± 7	798 ± 10	329 ± 3
	60%	165 ± 1	199 ± 1	233 ± 2	109 ± 1
	70%	41 ± 0	45 ± 0	61 ± 0	33 ± 0.1
c -jets	40%	23 ± 0	22 ± 0	28 ± 0	48 ± 1
	50%	12.7 ± 0.1	12.3 ± 0.1	12.9 ± 0.1	19.6 ± 0.1
	60%	7.6 ± 0.0	7.5 ± 0.0	7.7 ± 0.0	9.7 ± 0.1
	70%	4.6 ± 0.0	4.6 ± 0.0	4.8 ± 0.0	5.2 ± 0.0

Table 6.18: The rejection against light- and c -quark jets for fixed b -tagging efficiencies of 40%, 50%, 60% and 70%, by using both neural networks trained to reject either c - or light jets, compared to the combined $IP3D+SV1$ algorithm. The purification procedure has not been applied here. The results refer to the $t\bar{t}$ and $t\bar{t}jj$ samples.

At b -tagging efficiencies below 70% a significant increase in c -jet rejection can be obtained using the neural network discriminator against c -jets: for example at 50% b -efficiency, taking the $IP3D + SV1$ algorithm as a reference, the c -jet rejection can be increased by 50%, at the cost of reducing the light-jet rejection also by around 50%. It is worth studying the c -jet rejection in more detail as a function of the jet kinematics. This is illustrated in Fig. 6.47 and in Fig. 6.48. The most significant improvement in c -jet rejection can be obtained in the kinematic regions where b -tagging works best: for example at a b -tagging efficiency of 50% an improvement between 40 and 80% can be obtained for transverse momenta above 30 and below 200 GeV and for pseudo-rapidities below 1.75. In a typical analysis the best combined light and c -jet rejection will depend on the flavour composition of the background. Therefore, in general, neither using the discriminator against c -jets or the one against light-jets is the optimal choice. The optimal combined rejection can be obtained by defining a combined discriminator of the type:

$$\text{discr} = \frac{P(b)}{P(b) + c_l P(l) + (1 - c_l) P(c)}, \quad (6.39)$$

where c_l represents the relative fraction of light-jets in the background. In the case the neural network is optimally trained, thanks to the Bayes interpretation of $P(b)$, $P(c)$ and $P(l)$ as a-posteriori probabilities, using such a c_l allows to obtain the best signal-to-background ratio and thus overall background rejection. The c -jet and light-jet rejections achievable for fixed b -tagging efficiencies of 40, 50, 60 and 70% are shown in Fig. 6.49. Due to the different kinematic distributions of the c -, b - and light-jets in the sample used to estimate the performance, the best overall combined background rejection is not obtained at exactly the correspondent value of c_l : this can only be the case for jets with the same kinematic distribution. In the case

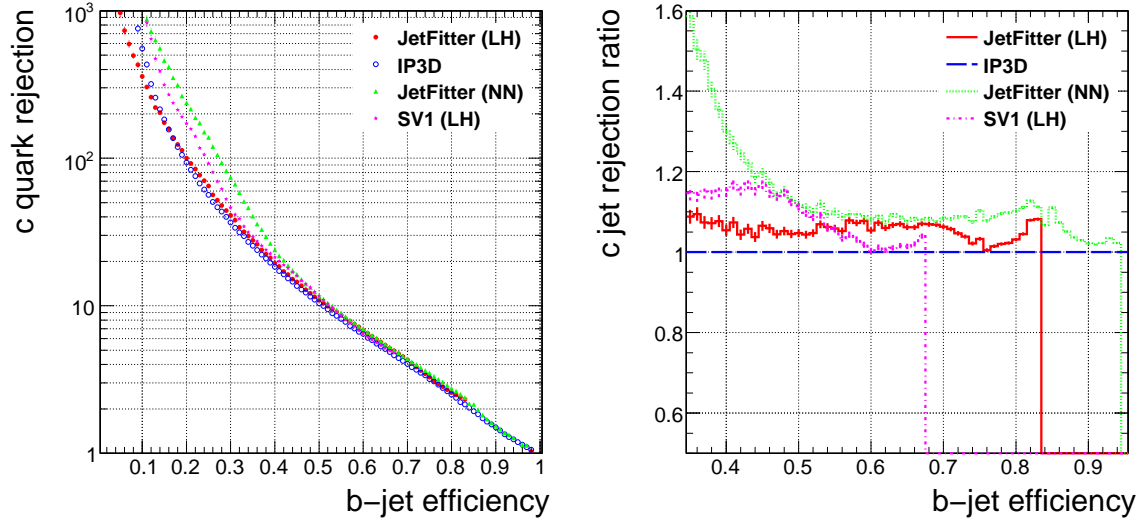


Figure 6.44: Left: Charm jet rejection versus b -tagging efficiency for $IP3D$, $SV1$ and $JetFitter$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $IP3D$ for $SV1$ and $JetFitter$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ sample.

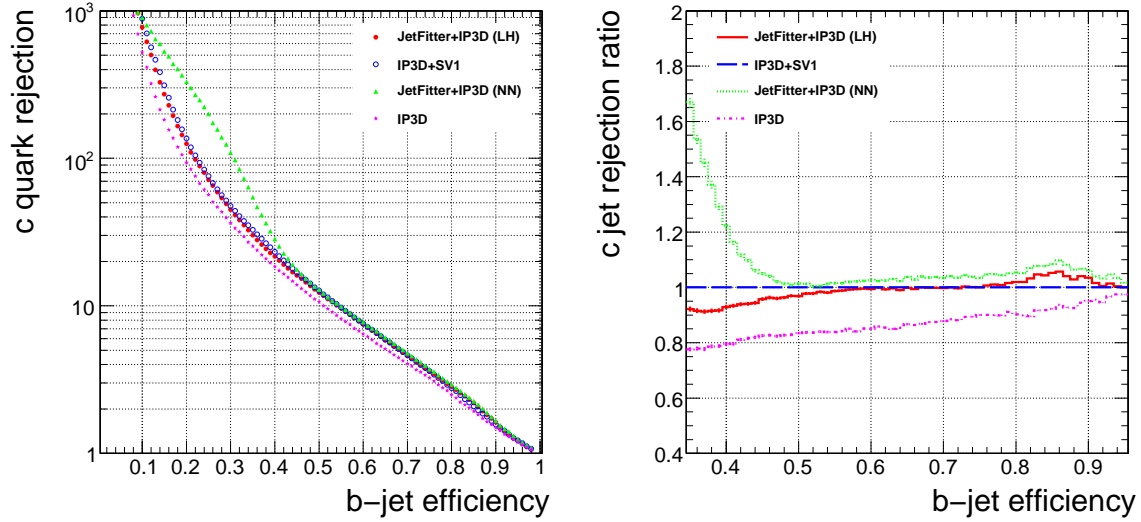


Figure 6.45: Left: Charm jet rejection versus b -tagging efficiency for $IP3D$, $SV1+IP3D$ and $JetFitter+IP3D$ (based on both likelihood and neural network approaches). Right: The ratio with respect to $SV1+IP3D$ for $IP3D$, $JetFitter+IP3D$ (both likelihood and neural network based). These results have been obtained on the $t\bar{t}$ sample.

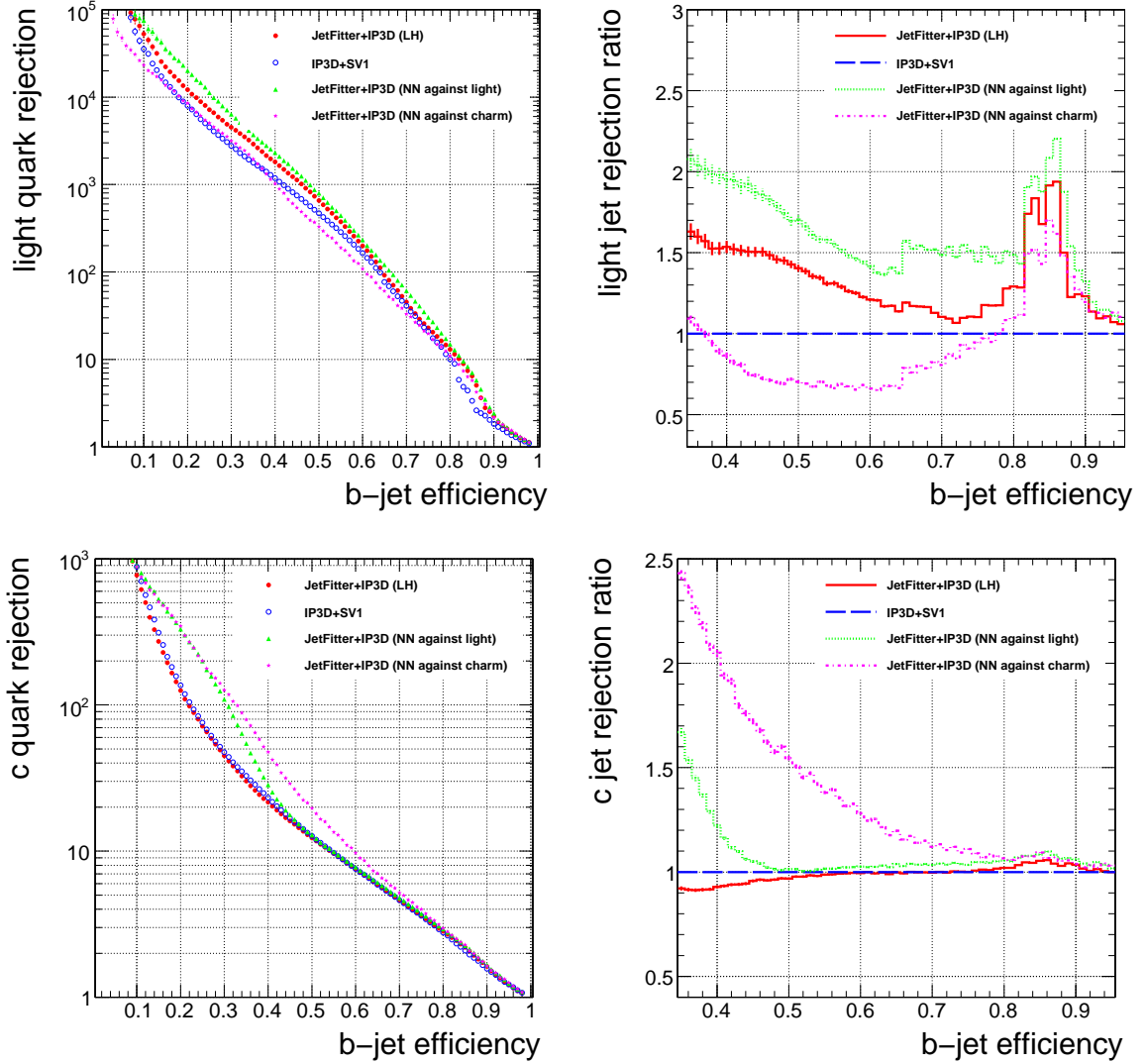


Figure 6.46: Left: Light jet (top) or charm jet (bottom) rejection as a function of the b -tagging efficiency for *JetFitter+IP3D*, based on the neural network discriminator against either c - or light jets, and for the remaining combined b -tagging algorithms. Right: The ratio with respect to *IP3D+SV1* for the same algorithms. These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples. No purification of light quark jets (see Section 6.10.1) has been applied.

of a real physics analysis with b -jets in the signal and with a strong contamination of both light and c -jets, it is best to scan all possible values of the parameter c_l in order to obtain the best overall background rejection for a certain signal efficiency. In the performance plots as a function of c_l it can be noticed that the c - or light-jet rejection does often drop only for extreme values of c_l close either to 0 or 1: therefore, by tuning the parameter c_l , quite a significant improvement in the overall background rejection can be obtained. Most of the gain is however still achievable at b -tagging efficiencies below 70%, where there is more room to specifically optimise against c - or light-jets: this is the region where the impact parameter significance or decay length significance starts to play a more limited role, and other variables, like the vertex mass or the decay chain topology, which yield a good discrimination between c - and b -jets, start to have a higher weight in the discriminator.

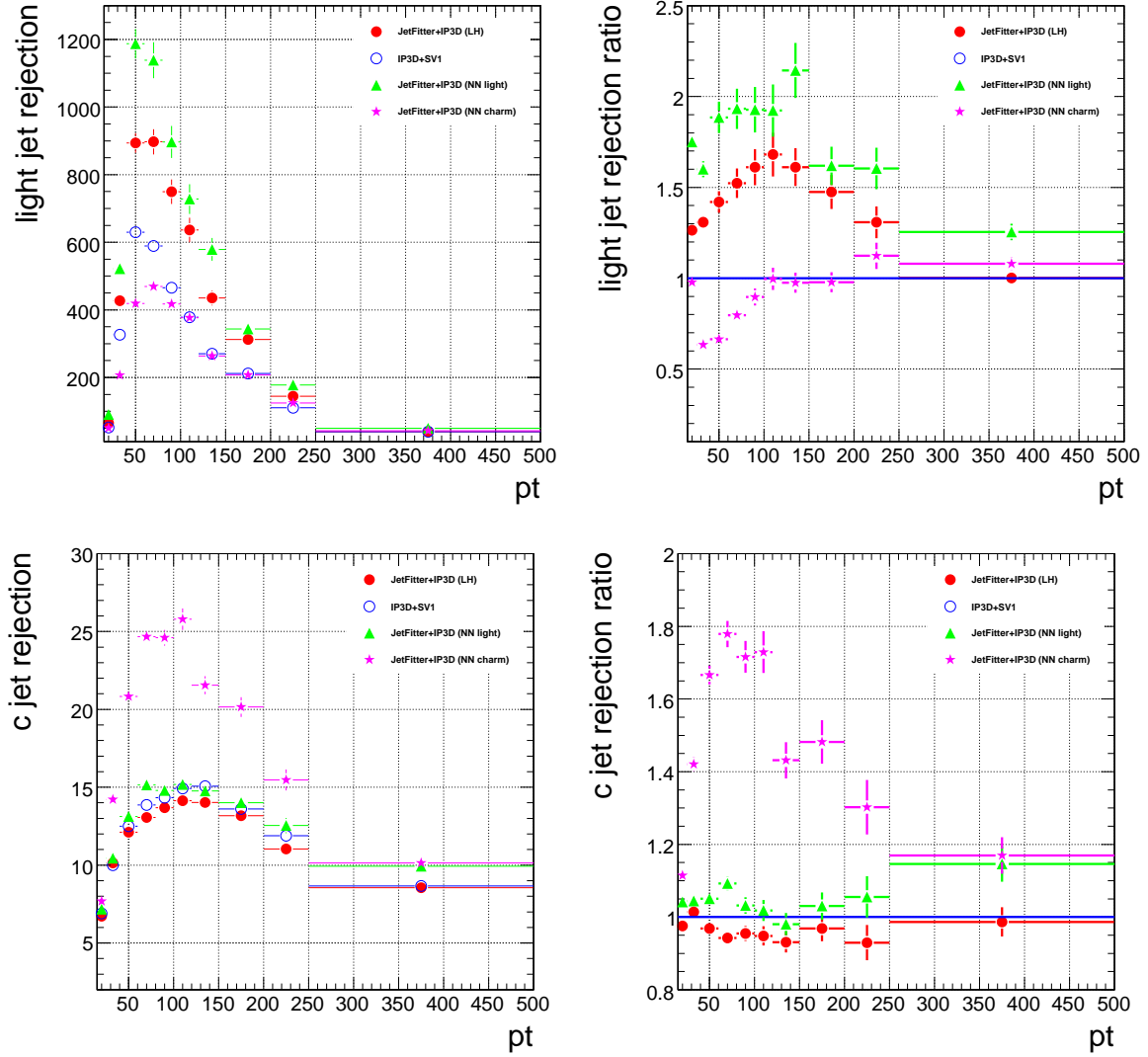


Figure 6.47: Left: Light jet (top) or charm jet (bottom) rejection for a fixed b -tagging efficiency of 50% as a function of the jet transverse momentum for $JetFitter+IP3D$, based on the neural network discriminator against either c - or light jets, and for the remaining combined b -tagging algorithms. Right: The ratio with respect to $IP3D+SV1$ for the same algorithms. These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples. No purification of light quark jets has been applied. The jet p_T is expressed in GeV.

The improvement in the combined c - and light-jet background rejection will play a crucial role in the analysis which will be presented in the last part of the present thesis work, which is about the WH Higgs boson search channel at high transverse momenta, with the W boson decaying leptonically and the Higgs boson decaying to a pair of b -quarks.

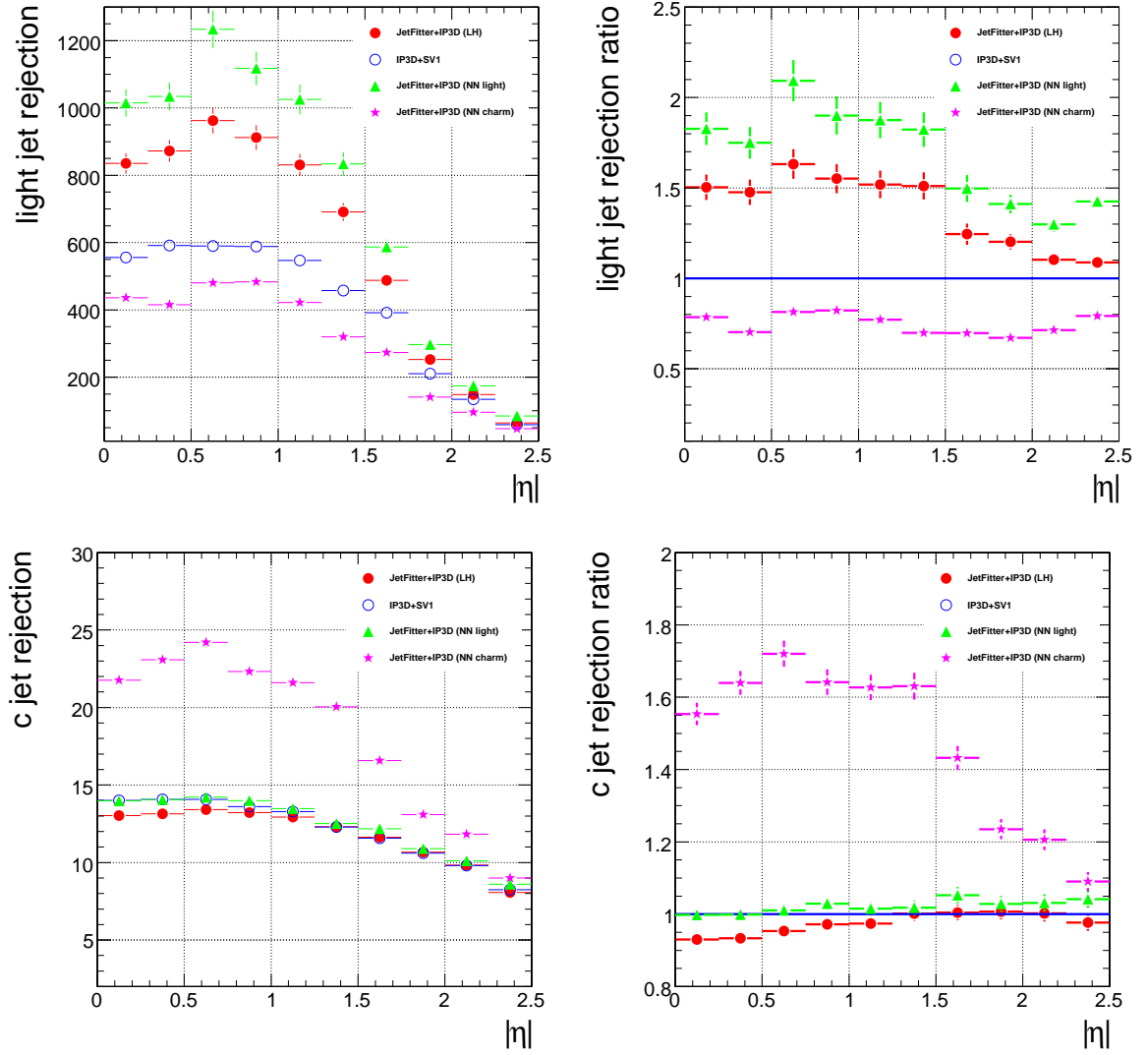


Figure 6.48: Left: Light jet (top) or charm jet (bottom) rejection for a fixed b -tagging efficiency of 50% as a function of the jet pseudo-rapidity for *JetFitter+IP3D*, based on the neural network discriminator against either c - or light jets, and for the remaining combined b -tagging algorithms. Right: The ratio with respect to *IP3D+SV1* for the same algorithms. These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples. No purification of light quark jets has been applied.

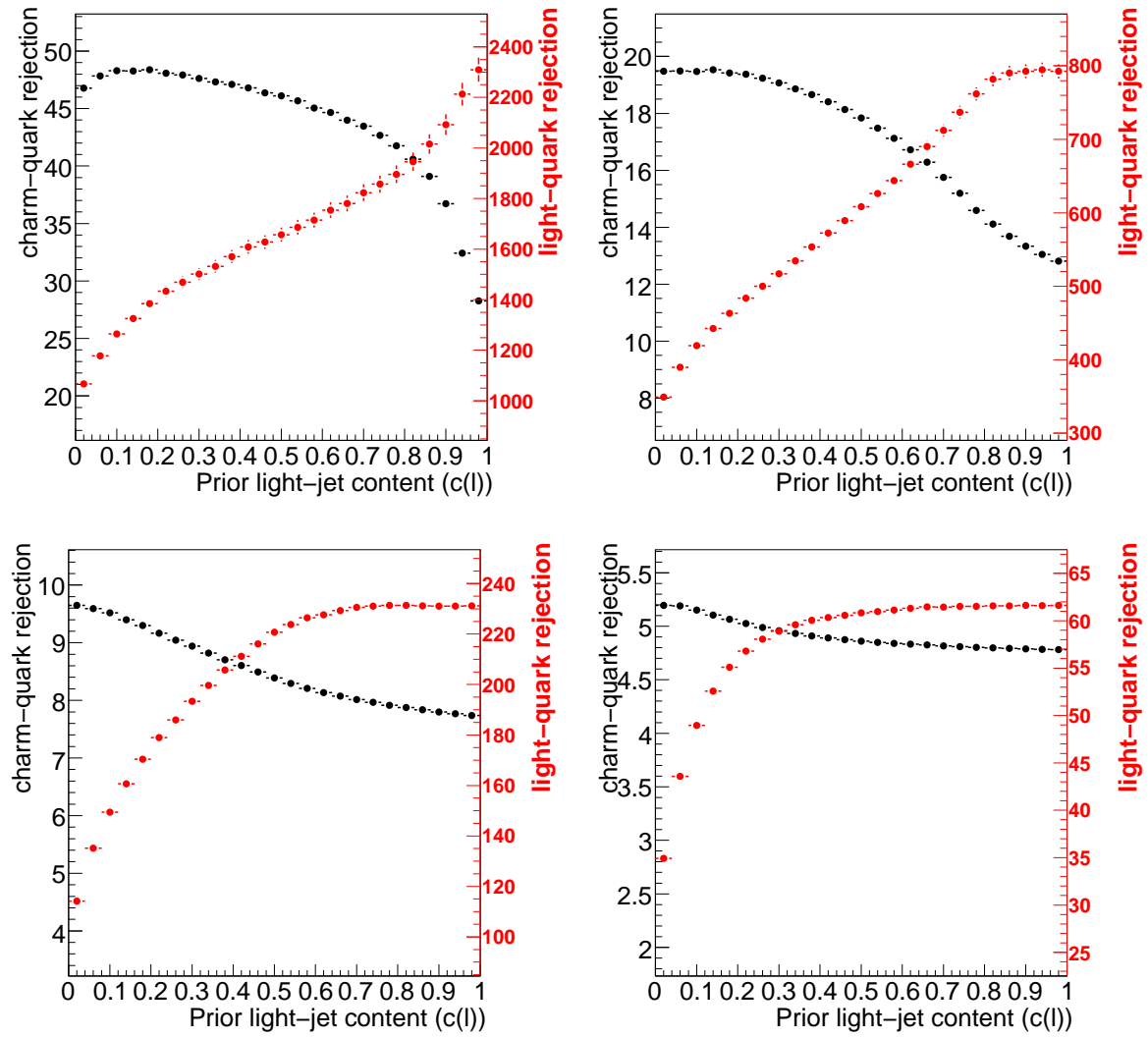


Figure 6.49: Left: Charm jet and light jet rejections versus the assumed prior light-jet background composition (c_l), for fixed b -tagging efficiencies of 40% (top left), 50% (top right), 60% (bottom left) and 70% (bottom right). These results have been obtained on the $t\bar{t}$ and $t\bar{t}jj$ samples.

7 The $WH \rightarrow \ell\nu b\bar{b}$ analysis at high transverse momenta

7.1 Introduction

The mechanism of mass generation is one of the fundamental open problems of particle physics, as mentioned in Chapter 1. In order to prove that the Standard Model of fundamental interactions describes this mechanism correctly, it needs to be proven that the Higgs boson exists in nature. This is one of the key aims of the ATLAS and CMS experiments at the LHC. As described in Chapter 2, the discovery of a Standard Model Higgs boson with ATLAS is particularly challenging in case the mass of the Higgs boson is close to the LEP limit of ≈ 115 GeV (a Higgs boson in this mass region will be denoted in the following as a *light* Higgs boson). In this case, the discovery sensitivity to a Higgs boson of the ATLAS experiment is more limited than at higher Higgs boson masses and relies on a combination of several channels, in particular $H \rightarrow \gamma\gamma$ and $H \rightarrow \tau\tau$ (see Fig. 2.5). There are several uncertainties connected with these searches, which depend both on the level of understanding of the detector performance (so far limited to Monte Carlo studies, even if partially based on test beam data) and on the modelling of the physics processes used in the Monte Carlo studies (event generation, photon fragmentation function, amount of statistics for background processes,...). It is therefore highly desirable to rely on the widest possible range of physics signatures, in order to gain robustness in the search of a low mass Higgs boson and, at the same time, to increase the combined discovery sensitivity.

Two important channels that have generally been considered less promising at the LHC are those where a Higgs boson is produced in association with a vector boson, $pp \rightarrow WH$ or ZH , followed by the dominant light Higgs boson decay, to two b -quark jets. The online selection of these event would typically rely on the identification of a leptonically decaying W or Z boson. The identification of the Higgs boson would be based on the presence of two b -jets in the event, with invariant mass corresponding to the a-priori not known Higgs boson mass. A first study of the WH channel was performed in ATLAS at the time of the Technical Design Report (TDR) [42], while a more detailed study was performed a few years later in Ref. [78]. The extraction of the signal is difficult because of the large backgrounds and very low signal acceptance, as the invariant mass distribution expected for the Higgs boson candidate in both signal and backgrounds after analysis selection of Fig. 7.1 illustrates: in the study presented in Ref. [78], the extraction of the $WH \rightarrow \ell\nu b\bar{b}$ signal results in a significance expressed in terms of S/\sqrt{B} ¹ of 2.1 with 30 fb^{-1} of integrated luminosity, with in addition a very low signal-to-background ratio of $S/B \approx 1.3\%$. The low signal-to-background ratio makes the analysis extremely sensitive to any uncertainty on the knowledge of the amount of background events, as shown in Fig. 7.2, where the significance is shown as a function of the assumed uncertainty on the knowledge of the expected number of background events $\Delta B/B$ ⁽²⁾; even in the most optimistic scenario where the backgrounds can be extracted from data

¹Here S is the number of expected signal events and B the number of expected background events.

²The corrected significance is based on the significance in terms of S/σ_B , where the uncertainty level on the background σ_B is computed by summing up quadratically the Poisson uncertainty \sqrt{B} and the systematic

by extrapolating them from the sidebands using their invariant mass distribution, it will be extremely hard to reach a precision $\Delta B/B$ at the percent level, which is required not to let the discovery sensitivity decrease drastically.

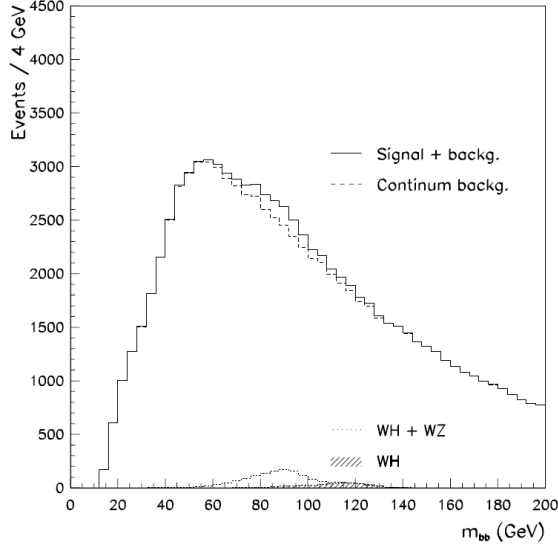


Figure 7.1: Mass distribution for the WH signal, WZ and all remaining backgrounds in the inclusive $WH \rightarrow \ell \nu b \bar{b}$ analysis. Taken from Ref. [78].

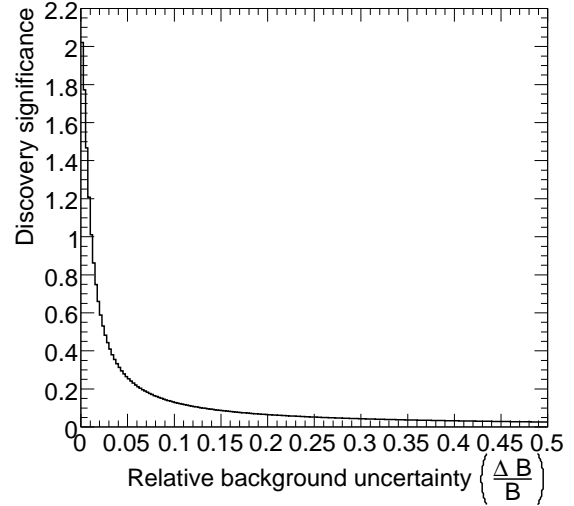


Figure 7.2: Corrected discovery sensitivity to a SM Higgs boson in the inclusive $WH \rightarrow \ell \nu b \bar{b}$ analysis ($S/\sqrt{B + \Delta B^2}$), as a function of the relative systematic uncertainty of the background contribution.

A recent study [1] argued that focusing the WH analysis only on the kinematic configurations where the W and H bosons are emitted at large transverse momenta and are basically back-to-back, has significant advantages over a more inclusive search. This region corresponds only to a small fraction of the total WH cross section (about 5% for $p_T > 200$ GeV), but the signal acceptance is larger, while the backgrounds are significantly reduced. Another advantage is that looking at configurations with back-to-back bosons, also the ZH production mode should become accessible, not only in the $Z \rightarrow \ell \ell$ decay mode, but also in the $Z \rightarrow \nu \nu$ mode, which is experimentally more difficult to access.

While the identification of a high p_T W or Z boson does not present particular challenges, the reconstruction of a high p_T Higgs boson decaying into a pair of b -quark jets does. In fact, due to the high boost, the two decay products will be increasingly close in ΔR for increasing p_T , as shown in Fig. 7.3, by approximately:

$$\Delta R(b, \bar{b}) \approx \frac{2 \cdot m_H}{p_{T,H}} \quad (7.1)$$

For example for $p_{T,H} = 300$ GeV ΔR is of the order of ≈ 0.8 . This makes the identification of the Higgs boson candidate more challenging and eventually requires both jet finding and b -tagging algorithms to be specifically optimised for this specific kinematic region.

For this reason, in the same study, a new jet algorithm was proposed, to identify the characteristic structure of a boosted Higgs to $b\bar{b}$ decay, and a first hadron level study was performed, in order to evaluate the discovery potential of the WH and ZH channels in this specific kinematic region. The effect of b -tagging was emulated by applying a fixed b -tagging efficiency of 60% and a fixed non b -jet rejection of ≈ 50 , where the b -tagging efficiency and rejection are defined as in Section 6.10.1. The effect of finite mass resolution was partially

uncertainty ΔB ($\sigma_B = \sqrt{B + \Delta B^2}$).

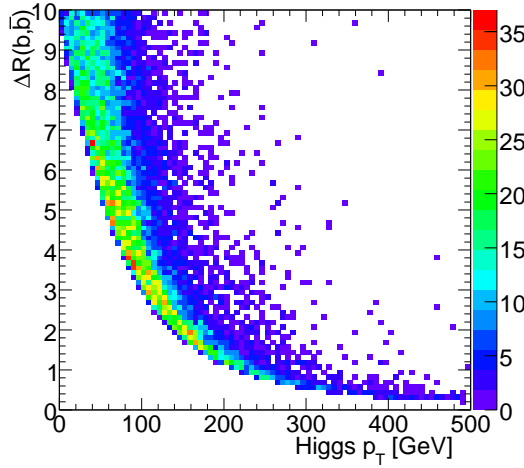


Figure 7.3: Distance in $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ between the two b -quarks from the Higgs boson (with $m_H = 120$ GeV) as a function of Higgs boson transverse momentum. $\Delta\phi$ represents the distance in azimuthal angle, while $\Delta\eta$ is the distance in pseudo-rapidity. The normalisation is arbitrary.

taken into account by selecting Higgs bosons with an invariant mass m_H within a $(1 \pm 10\%)m_H$ mass window.

There are several reasons why a detailed study based on a realistic detector simulation is crucial for confirming that extracting a Higgs boson signal from these channels is indeed possible. Some of the reasons will be briefly mentioned here:

- Possible impact of both calorimeter granularity and energy resolution and bending of charged particles in the magnetic field on the jet finding efficiency and invariant mass resolution;
- First realistic evaluation whether b -tagging can be really applied efficiently and separately on the two jets from the Higgs boson candidate;
- Evaluation of the effect of taking into account the presence of c -jets, in particular in the $t\bar{t}$ background, which was ignored in the hadron level study;
- Effect of trigger selection;
- Impact of realistic signal acceptance.

The most crucial points are certainly the ones concerning b -tagging. In particular, if the two jets from the Higgs boson candidate cannot be b -tagged separately, the rejection factor against background from u , d or s -jets would be R_{uds} , where R_{uds} is the rejection factor against a single light jet, instead of R_{uds}^2 , making the proposed search completely hopeless. Equally important is the impact of the presence of c -jets in the background, since for them the rejection, even for conventional jets (as was analysed in Section 6.10.4), is $R_c \approx 13$ for a b -tagging efficiency of $\epsilon_B \approx 60\%$, significantly lower than the rejection of $R \approx 50$ assumed for all non b -jets in the hadron level study.

In this chapter the first detector level study performed in ATLAS for the WH channel is presented.

7.2 Detector simulation

A realistic study of the discovery potential in the WH channel in the *boosted* scenario must necessarily rely on a realistic simulation of the detector response: with a complex experimental apparatus like ATLAS, this can be done only by using detailed Monte Carlo simulations.

In order to profit from the detector simulation, a fixed-order parton level calculation of the signal and background processes expected to enter the signal selection is not sufficient. As described in Chapter 1, the use of parton-shower based Monte Carlo generators results in the production of a set of unweighted events which contain all final state particles and can be used directly as input to the detector simulation.

This study relies on the use of the ATLFAST-II detector simulation [79]. Particles in the inner detectors and muons passing through the calorimeter and muon system are described using a full GEANT4 detector simulation, while the interactions of particles with the calorimeter is described by using a faster simulation of the detector response (FastCaloSim [80]), based on a full parametrisation of the calorimeter response in its full granularity to single final state particles entering the calorimeter surface. The use of a full GEANT4 based simulation for all sub-detectors is unpractical for this analysis, due to very low signal rate and the very high amount of events which need to be simulated to make reliable predictions for the background contributions: the use of ATLFAST-II is absolutely crucial, since it permits to decrease the average time needed to simulate a single physics event from ≈ 20 minutes to ≈ 1 minute and makes a realistic study of this channel possible.

The two most crucial ingredients in this analysis, where the detector simulation needs to provide a realistic answer, are jet finding and b -tagging. The jet finding algorithm implemented in this study is sensitive to the details of the jet-substructure. While the Monte Carlo generator used in the study should describe the jet shapes correctly at least up to leading logarithmic accuracy, it is extremely important that the fast calorimeter simulation reproduces realistically the calorimeter response in this scenario. A detailed comparison between a set of signal events simulated with ATLFAST-II and complete full simulation is presented in Section 7.10, where a sufficiently good agreement between the two simulations is found. The second crucial ingredient of the analysis, the use of b -tagging, relies mainly on the charged particles reconstructed as tracks in the inner detector and only marginally on the jets as reconstructed in the calorimeter, which provides only the direction used to associate tracks to the jet. The simulation of tracks in the inner detector in ATLFAST-II relies also on the GEANT4 full simulation, which is the most realistic Monte Carlo based simulation particle physicists have at their disposal.

7.3 Event samples

All potential signal and background contributions must be considered. The WH analysis is based on the identification of a high p_T lepton from the W boson, transverse missing energy (E_T^{miss}) from the associated neutrino and, on the opposite side of the event, on two b -jets from the Higgs boson. The main signal contribution is expected from the WH process, but a minor contribution is expected from ZH , where the Z boson decays leptonically into two leptons, but one of them is missed in the detector or is outside acceptance and gives therefore rise to a high amount of E_T^{miss} . Given the $\ell\nu b\bar{b}$ signature, the dominant background contributions are expected from WZ , W +jet, $t\bar{t}$ and single top (Wt channel). Further minor contributions are expected from Z +jets and the remaining di-boson production processes ZZ and WW .

The matrix element calculations used in the Monte Carlo generators for all these processes are based on fixed order calculations at leading order in the perturbative expansion in the

strong coupling constant α_s : the reason for not using Monte Carlo generators based on the next-to-leading order expansion in α_s is that no such a Monte Carlo generator is actually available for one of the main backgrounds, $Wb\bar{b}$ (or more in general Wjj). All results presented in this study are therefore based on leading order predictions and cross sections only.

For all expected signal and background contributions event samples were produced with the HERWIG Monte Carlo generator [33, 81] using JIMMY [82] to simulate the underlying event, assuming a centre-of-mass energy of 14 TeV, and then passed through the ATLFAS-II simulation of the ATLAS detector. An exception is that the AcerMC [83] Monte Carlo generator was used to produce single top (Wt) events, as well as additional $Wb\bar{b}$ samples used as a cross-check. For the $Wb\bar{b}$ sample AcerMC was interfaced to HERWIG for the parton showering and to JIMMY for the underlying event, while for the single-top sample PYTHIA [32] was used. For all processes, the CTEQ6L1 (leading order) PDFs were used [84].

For some of the background processes it is worth analysing if a leading order based Monte Carlo generator can be expected to reproduce them sufficiently well. This is certainly the case for the WH and ZH signal and for the di-boson WW , ZZ and WZ processes. These processes originate all from electroweak interactions. Eventual additional jets in the event stem exclusively from initial state radiation: since the Monte Carlo is leading order in the strong coupling constant only, this additional radiation is described only by the parton shower algorithm, on top of the leading order matrix element. The parton shower approximation is expected to work well for initial state radiation only if the additional radiation is soft or collinear to the incoming partons. The analysis selection is such that two high p_T bosons (where typically $p_T > 200$ GeV), in opposite directions, will be selected, while an eventual jet escaping the veto will have typically low transverse momentum, so it is expected to be described well by the parton shower approximation. The same argument applies also to the $t\bar{t}$ background, where the analysis typically selects two very boosted top quarks and the third jet must necessarily be soft, or it will again fail the jet veto.

Unfortunately the same argument does not apply to the W +jet background, and, analogously, to the Z +jet background. In this case, in fact, a single weak boson is present in the event and, in order to fake the high p_T Higgs boson candidate, at least two additional hard jets, preferably b -jets, must be present. While the first jet in HERWIG for these processes is produced directly according to the matrix element calculation, the second one is only produced by the parton shower algorithm: in the case the second jet is not sufficiently collinear to the first or softer, the W +jet (or Z +jet) background can be significantly underestimated. This needs to be carefully checked and is therefore worth a more detailed discussion.

In addition, the contributions expected from the s - and t -channels of single top production, from the $t\bar{t}Z$ process and from pure QCD di-jet production are considered to be negligible. This needs to be justified and is also discussed in more detail in the following.

W +jet background

The W +jet background, after b -tagging is applied, is dominated by the irreducible $q\bar{q} \rightarrow Wg \rightarrow Wb\bar{b}$ component. The kinematic region of this analysis, in which a $b\bar{b}$ pair is present at high p_T , is such that the leading-logarithmic parton-shower approximation as implemented in HERWIG (or PYTHIA), which is defined in the following as PS, can be expected to work well. In general, these approximations include the kinematically dominant terms where the ratio of scales involved (in this case the p_T of the $b\bar{b}$ system and either its mass or the relative p_T of the b and \bar{b}) is large. A large ratio for these scales would in fact correspond respectively to the collinear and soft kinematic limits where the parton shower approximation is expected to be reliable. However, since in the final selection the $b\bar{b}$ mass is relatively high (i.e. around the

Higgs boson candidate mass), this ratio may for the lowest p_T events be around two, which is not very large. While there is good evidence (see for example Ref. [85,86]) that parton showers do reproduce many event properties very well over a wide kinematic range, this can not be directly extended to $b\bar{b}$ production. In order to cross-check the parton shower approximation in this kinematic region, the rate and kinematics of $b\bar{b}$ production in HERWIG was compared to a fixed order α_s calculation of $Wb\bar{b}$ production, as implemented in AcerMC (which we define as ME). AcerMC uses the full LO matrix element for $Wb\bar{b}$ and $Zb\bar{b}$ production, including the effect of the b -quark mass and a matching to the initial-state parton-shower where the $g \rightarrow b\bar{b}$ splitting can be significant. However, it does not include a complete matching of the matrix element on to the final-state parton-shower which can include the same splitting, and which is expected to be an important contribution in the kinematic region of this analysis: in the AcerMC sample the final-state parton-shower will act directly on the two b -quarks produced by the matrix element. This means that the AcerMC samples will be missing some fraction of the parton shower, since, due to fact that the $b\bar{b}$ pair is produced by the ME itself and due to angular ordering, the $b\bar{b}$ is in effect forced to be the widest angle splitting, whereas in the full parton shower in HERWIG contributions where a $g \rightarrow gg$ splitting is the widest in angle, but where a $b\bar{b}$ pair is nevertheless still developed within the shower, will also be present.

In order to have a reference, first the AcerMC (ME) and HERWIG (PS) samples are compared, after applying the generator cuts which limit the $p_T(W)$ to be higher than 150 GeV. Some kinematic distributions of the $b\bar{b}$ system as produced by HERWIG and AcerMC are shown in Fig. 7.4. It can be clearly seen that at low $b\bar{b}$ invariant mass or low relative p_T , the PS prediction is enhanced with respect to the ME, due to higher order $g \rightarrow b\bar{b}$ splitting contributions, while, at high masses or high relative p_T the PS approach breaks down, and only the ME approach can fill that region of phase space. This means at at high gluon virtuality the PS result cannot be trusted.

However, in the specific region of phase space which is relevant for the present analysis, after requiring $p_T(b\bar{b}) > 200$ GeV and imposing $1.3 > \Delta R > 0.3$, the picture changes significantly, as the kinematic distributions of the $b\bar{b}$ system of Fig. 7.5 illustrate: most of the distributions from AcerMC (ME) and HERWIG (PS) are now in sufficiently good agreement. This is certainly the case for the mass of the b and \bar{b} system, in particular in the mass region around $m_H \approx 120$ GeV, and for the relative p_T of the b - and \bar{b} -quarks (first and second rows of plots). Some small differences can be noticed in the plots in the third row: the p_T of the $b\bar{b}$ system is slightly harder for the AcerMC simulation, consistent with the present of an extra gluon radiation in HERWIG, where an extra gluon may carry off some of the p_T , and the ΔR distribution, which shows some fluctuations beyond statistical uncertainty, with HERWIG having a tendency to fall off faster at large ΔR as might be expected.

From this comparison it is possible to conclude that both the ME and PS approach give very similar results for the kinematics for the $b\bar{b}$ system in the relevant region of phase space for the present analysis. This is not too surprising, since, after the requirement of having $p_T(b\bar{b}) > 200$ GeV and imposing $\Delta R > 0.3$, the region where the PS contribution is enhanced with respect to the ME is cut out, while, imposing $\Delta R < 1.3$, the region of highest gluon virtuality (in particular where the two b -quarks tend to be more back-to-back) is removed as well.

While the kinematic configuration of the $b\bar{b}$ system is very similar in both the HERWIG and AcerMC samples, the lack of ME/PS matching in the final state for AcerMC will lead to a depletion of gluon radiation, which will have an impact on the reconstruction efficiency of the hadronic $b\bar{b}$ system: therefore the HERWIG samples were used for the mainstream analysis³.

³In addition, in order to make the AcerMC $Wb\bar{b}$ sample usable for some of the optimisation and verification studies, the events of this sample were added to the HERWIG sample (which contains also all non b -

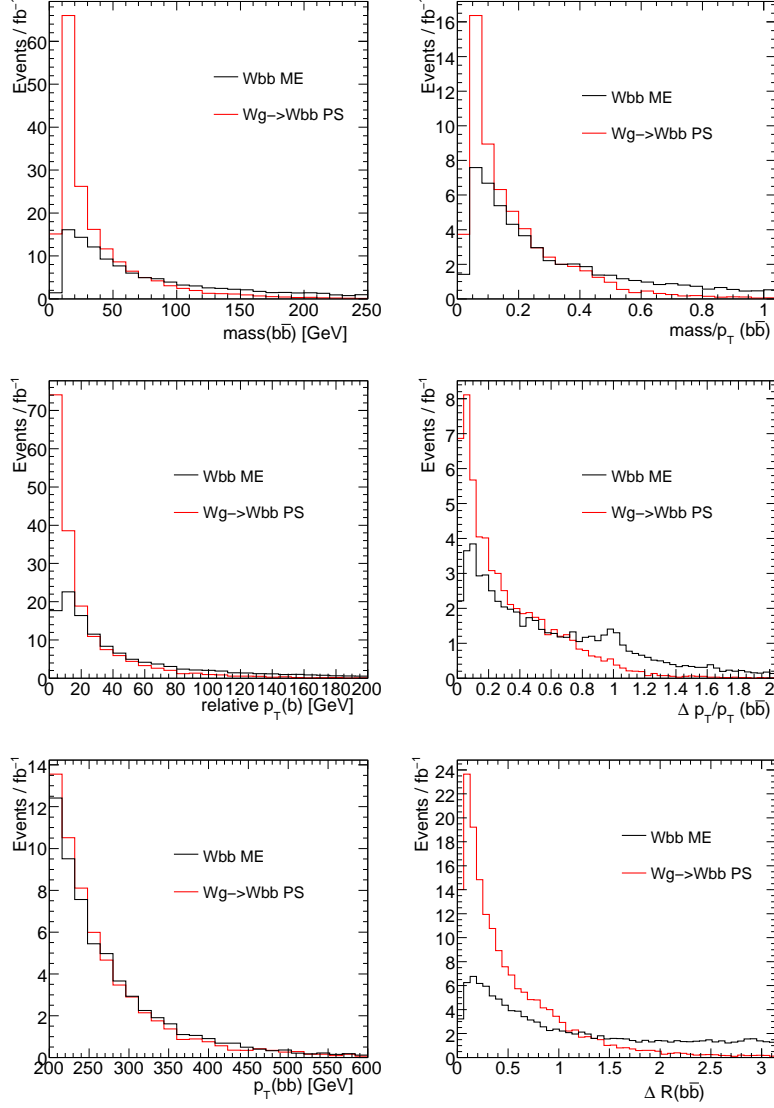


Figure 7.4: Distributions of important kinematic variables of the $b\bar{b}$ system in HERWIG and AcerMC, after generator level cuts ($p_T(W) > 150$ GeV): invariant mass of the $b\bar{b}$ system (top left), invariant mass of the $b\bar{b}$ system divided by $p_T(b\bar{b})$ (top right), relative $p_T(b, \bar{b})$ (middle left), relative $p_T(b, \bar{b})$ divided by $p_T(b\bar{b})$ (middle right), transverse momentum of the $b\bar{b}$ system (bottom left) and $\Delta R(b\bar{b})$ (bottom right). All distributions are normalised to their relative cross sections in fb.

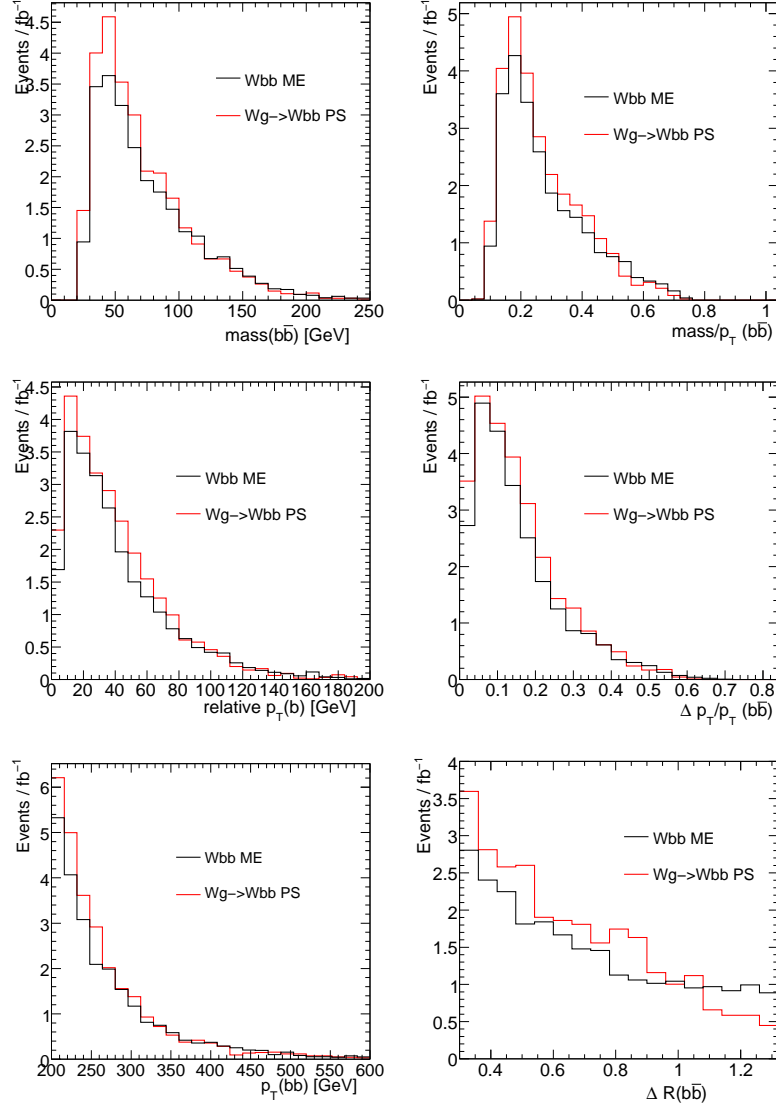


Figure 7.5: Distributions of important kinematic variables of the $b\bar{b}$ system in HERWIG and AcerMC, in the phase space region which is relevant for the WH analysis ($1.3 > \Delta R(b\bar{b}) > 0.3$ and $p_T(b\bar{b}) > 200$ GeV): invariant mass of the $b\bar{b}$ system (top left), invariant mass of the $b\bar{b}$ system divided by $p_T(b\bar{b})$ (top right), relative $p_T(b, \bar{b})$ (middle left), relative $p_T(b, \bar{b})$ divided by $p_T(b\bar{b})$ (middle right), transverse momentum of the $b\bar{b}$ system (bottom left) and $\Delta R(b\bar{b})$ (bottom right). All distributions are normalised to their relative cross sections in fb.

All the above samples are leading order and/or leading logarithmic. At the LHC, a large enhancement in the $Wb\bar{b}$ cross section is expected at NLO, essentially because a new process ($qg \rightarrow Wb\bar{b}j$), which cannot just be described with additional initial state radiation, comes into play at NLO (see Refs. [87, 88]): however, in this case an additional hard jet is expected in the event and, due to the jet veto implemented in the analysis, this additional NLO contribution is expected to be largely reduced, as described in Ref. [89]. In addition, the leading-logarithmic contribution to this correction will again be included in HERWIG.

Single top background

For the single top background, three different channels have to be considered: s and t-channel and Wt production. In the hadron level study performed in Ref. [1], it was noticed that the single top s- and t-channels produce a negligible level of background; while the s channel ($qg \rightarrow tb$) has a very low cross section (7.1pb at LO), the t channel $qg \rightarrow qt(b)$ has a large cross section (251 pb at LO), but in both cases they can hardly fake the signal topology, where a heavy highly boosted object decaying to a b-quark pair is expected together with a high p_T W boson decaying into a lepton on the other side: in fact, in these channels one of the two b -jets is either soft or it goes into a direction which is opposite to the direction of the top, where the other b -jet needs to come from and which needs to be sufficiently boosted in order to produce a high p_T W boson. For this reason no single top events for the s and t-channels were produced in this study. However, the Wt channel needs to be considered, because its topology is extremely close to the signal, with one W boson on one side and a top quark on the other side, which fakes the Higgs boson candidate in case one of the three produced jets from the top is very soft and remains undetected, while the remaining c - and b -quarks fake the $b\bar{b}$ pair. This background was not considered in Ref. [1] and is studied here for the first time.

Associated production of a Z boson with a top quark pair

No event sample was generated and simulated for the $t\bar{t}Z$ background. However a simple parton level study was performed, which confirms that this background does not provide a significant contribution. The result is reported in Section C.1 of the appendix.

QCD multi-jet background

Finally, no background from QCD multi-jet events was considered. In order to reduce the impact of eventual jets faking leptons, an isolation requirement on muons and electrons is applied. There is nothing particular in this analysis which should enhance the QCD fake background compared to more normal W and Z studies, where such backgrounds are found to be very small [44], however, the kinematic configuration and flavour composition selected in this analysis is sufficiently different, so that a more detailed study of such backgrounds would be of value. Unfortunately more detailed studies suffer from the extremely high cross sections and tiny selection acceptance of such processes, so that the generation of such samples would require huge statistics and therefore a very large amount of computing resources, even relying on ATLFEST-II for the simulation. In general, given the high p_T cuts adopted in the analysis

jet contributions), but, in order to avoid double counting the common $b\bar{b}$ contributions, the $qg \rightarrow Wg$ component where a b and a \bar{b} are produced in the parton shower was first removed from the HERWIG sample, keeping the remaining $qg \rightarrow Wg$ component and all events produced in HERWIG by the $qg \rightarrow Wg$ Matrix Element.

both for the W and Higgs boson candidates, it should be more difficult for the QCD dijet background to pass the analysis cuts than in more inclusive analyses.

A bit more in detail, considering the $\ell\nu b\bar{b}$ signature selected by this analysis, the QCD background contribution could in principle arise either due to a fake or due to a real lepton (plus in both cases real or fake E_T^{miss}). While the fake lepton contribution should be small, since it requires the correlated presence of a fake lepton and of a high amount of E_T^{miss} , such that $p_T(E_T^{\text{miss}} + \ell) > 200$ GeV, the contribution from a real lepton from the semi-leptonic decay of a b - or c -hadron for example in $gb \rightarrow b\bar{b}$ or $gc \rightarrow b\bar{b}c$ could be potentially dangerous; in the case of these processes, the corner of phase space where the fragmentation of a b - or c -quark with $p_T > 200$ GeV into the corresponding leptonically decaying b - or c -hadrons does not produce additional particles stemming from the primary event is very small, so most of this background should fail the lepton isolation requirement.

List of samples

All generated samples were passed through the ATLFast-II simulation by using release 13.0.40.5 of the ATLAS simulation and reconstruction software ATHENA and rely on the ATLAS-CSC-02-00-00 version of the ATLAS detector geometry: they do not include the simulation of pile-up and cavern background. The WH signal sample at a reference mass of 120 GeV was also passed through the full simulation of the ATLAS Detector, in order to make the comparison between full simulation and ATLFast-II possible.

The details of the event samples used in this study are given in Table 7.1. For the signal samples, reference Higgs boson masses of 115, 120 and 130 GeV were chosen. For technical reasons, the $t\bar{t}$ sample is divided into two, one where the two top quarks have a $p_T > 150$ GeV (before eventual QCD radiation) and the complementary sample with the two top quarks having $p_T < 150$ GeV.

Process	Generator cut	σ (pb)	Additional filter after generation	Filter efficiency
$WH(115)$	none	1.157 pb	$p_T(H) > 150$ GeV, $p_T(W) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$4.04 \pm 0.03\%$
$WH(120)$	none	0.953 pb	$p_T(H) > 150$ GeV, $p_T(W) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$4.38 \pm 0.04\%$
$WH(130)$	none	0.602 pb	$p_T(H) > 150$ GeV, $p_T(W) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$5.19 \pm 0.03\%$
$WH(120)$	none	0.953 pb	$p_T(H) > 150$ GeV, $E_T^{\text{miss}} > 100$ GeV	$4.39 \pm 0.04\%$
$ZH(115)$	none	0.660 pb	$p_T(H) > 150$ GeV, $p_T(Z) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$3.21 \pm 0.02\%$
$ZH(120)$	none	0.545 pb	$p_T(H) > 150$ GeV, $p_T(Z) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$3.51 \pm 0.02\%$
$ZH(130)$	none	0.347 pb	$p_T(H) > 150$ GeV, $p_T(Z) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$4.18 \pm 0.03\%$
WW	$\hat{p}_T^{\text{min}} = 150$ GeV	2.059 pb	$p_T(e, \mu) > 15$ GeV	$40.7 \pm 0.4\%$
ZZ	$\hat{p}_T^{\text{min}} = 150$ GeV	0.440 pb	$p_T(e, \mu) > 15$ GeV	$61.2 \pm 0.2\%$
WZ	$\hat{p}_T^{\text{min}} = 150$ GeV	0.96 pb	$p_T(e, \mu) > 15$ GeV	$33.6 \pm 0.2\%$
$t\bar{t}$	$\hat{p}_T^{\text{min}} = 150$ GeV	112.7 pb	$p_T(e, \mu) > 20$ GeV	$47.5 \pm 0.2\%$
$t\bar{t}$	$\hat{p}_T^{\text{max}} = 150$ GeV	298.7 pb	$p_T(e, \mu) > 20$ GeV	$39.8 \pm 0.5\%$
$Z + jet$	$\hat{p}_T^{\text{min}} = 150$ GeV	160.3 pb	$p_T(e, \mu) > 15$ GeV	$13.2 \pm 0.2\%$
$W + jet$	$\hat{p}_T^{\text{min}} = 150$ GeV	384.5 pb	$p_T(e, \mu) > 15$ GeV	$21.1 \pm 0.1\%$
$Wb\bar{b}$	none	89.96 pb	$p_T(W) > 150$ GeV, $p_T(e, \mu) > 15$ GeV	$0.51 \pm 0.01\%$
Wt	none	57.896 pb	$p_T(W) > 150$ GeV, $p_T(top) > 100$ GeV, $p_T(e, \mu) > 15$ GeV	$9.76 \pm 0.09\%$

Table 7.1: Monte Carlo samples produced for the present study. All samples have been produced with the HERWIG Monte Carlo generator, except for the $Wb\bar{b}$ and single top samples, which were produced with AcerMC. All cross sections are given at LO.

7.4 Higgs Candidate Identification

7.4.1 Jet finding strategy

The jet finding algorithm for this analysis has been implemented according to the procedure described in Ref. [1]. Here the overall strategy for clustering will be recalled and some details about the actual implementation in ATLAS and about its interplay with jet calibration and

b -tagging will be given. In a later section (7.8.1), the WH analysis based on the this new jet clustering approach will be also compared to the same analysis performed with a more traditional cone based jet clustering algorithm, in order to understand the real advantage of the new clustering method, after a realistic detector simulation is considered.

When a highly boosted Higgs boson decays, it produces in most cases a single fat jet containing two b -quarks. The identification strategy proposed in Ref. [1] uses the inclusive, longitudinally invariant Cambridge/Aachen (C/A) algorithm [90,91] to flexibly adapt to the fact that the $b\bar{b}$ angular separation varies significantly with the Higgs boson p_T and decay orientation.

The C/A jet finding algorithm is analogous to the k_\perp algorithm [92,93], where the following iterative procedure is used: a certain distance is defined between all pairs of particles i and j , the closest pair is combined, all distances are updated and the procedure is repeated until all objects are separated by more than a fixed minimum distance. The difference is that in the k_\perp algorithm the distance is defined as $\min(p_{T,i}, p_{T,j})\Delta R(i,j)$, while in the C/A algorithm the distance is simply $\Delta R(i,j)$. Here $\Delta R(i,j)$ is defined as $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$, where y is the pseudo-rapidity and ϕ the azimuthal angle. The full clustering history above a minimum $\Delta R_{i,j} = 0.07$, is stored to be used later for the substructure decomposition procedure.

When applying the jet clustering method on real reconstructed events, in this procedure the final state particles are substituted by *topological clusters* identified in the ATLAS calorimeter: these are three-dimensional clusters of energy in which deposits of energy in the single calorimeter cells are dynamically grouped together, trying, in the limit of what the calorimeter granularity and resolution effects allow, to let topoclusters correspond to single particles entering the calorimeter, as described more in detail in Section 4.3.

All jets j corresponding to the history stage at $\Delta R_{ij} = 1.2$, that is the stage at which all jets are separated from each other by at least $\Delta R = 1.2$, are selected. Those jets with $p_T > 200$ GeV and $|\eta| < 2.5$ are subjected to an iterative decomposition procedure involving two dimensionless parameters, μ and y_{cut} . A full explanation of the procedure can be found in Ref. [1], and the main stages are directly quoted below:

1. Break the jet j into two subjets by undoing its last stage of clustering. Label the two subjets j_1, j_2 such that $m_{j_1} > m_{j_2}$.
2. If there was a significant mass drop (MD), $m_{j_1} < \mu m_j$, and the splitting is not too asymmetric, $y = \frac{\min(p_{tj_1}^2, p_{tj_2}^2)}{m_j^2} \Delta R_{j_1, j_2}^2 > y_{\text{cut}}$, then deem j to be the heavy-particle neighbourhood and exit the loop. Note that $y \simeq \min(p_{tj_1}, p_{tj_2}) / \max(p_{tj_1}, p_{tj_2})$.
3. Otherwise redefine j to be equal to j_1 and go back to step 1.

In the above quoted list, the “heavy-particle neighbourhood” is the region to which angular ordered QCD radiation from the Higgs boson decay products is expected to be confined. The two parameters μ and y_{cut} are chosen independently of the Higgs boson mass and of p_T . A choice of $\mu = 1/\sqrt{3}$ ensures in principle that if, in its rest frame, the Higgs boson decays in such a way that the energy is equally shared between the three partons in a $b\bar{b}g$ configuration, then it will still trigger the mass drop condition. The cut on $y \simeq \min(z_{j_1}, z_{j_2}) / \max(z_{j_1}, z_{j_2})$, where z_{j_1} and z_{j_2} are the momentum fractions of the two quarks, eliminates the asymmetric configurations that most commonly generate significant jet masses in non- b or single- b jets because of the high probability of soft gluon radiation. In this analysis a small dependence was found on the precise value of μ used, while a bit stronger dependence is found on the

value of y , whose impact on the analysis will be evaluated in a later step: the optimal value has been chosen to be $y_{cut} = 0.1$.

The angular distance between j_1 and j_2 , $R_{b\bar{b}}$, defines the distance between the two b -quarks. In order to obtain a good b -tagging performance, a reliable separation and reconstruction of the two b -subjets is needed, so that the direction of the two b -subjets can be considered as a reasonable approximation for the direction of the outgoing b -partons after eventual QCD final state radiation. This is crucial in order to correctly associate to the two subjets their charged-particle tracks as reconstructed in the inner detector, avoiding to a large extent cross-talk between the two subjets.

At this stage, the effective size of jet j will be just sufficient to contain the QCD radiation from the Higgs boson decay, which, because of angular ordering [94–96], will be almost entirely emitted in the two angular cones of size $R_{b\bar{b}}$ around the b -quarks. Since this radius sets the angular scale (candidate-by-candidate) of the Higgs boson decay, it makes sense to re-cluster, or *filter* the candidate using this information. This involves rerunning the C/A algorithm on the jet constituents, using a finer angular scale, $R_{\text{filt}} < R_{b\bar{b}}$, and taking the three hardest objects (sub-jets) that appear — thus one captures the dominant $\mathcal{O}(\alpha_s)$ radiation from the Higgs boson decay, while eliminating much of the contamination from the underlying event. A choice of $R_{\text{filt}} = \min(0.3, R_{b\bar{b}}/2)$ is made. The jet j is accepted as a Higgs boson candidate if the two hardest subjets have b tags. The filtering procedure provides also an effective way to remove some of the contributions arising directly from the showering of the b -quark before hadronisation (i.e. no long-lifetime component) and thus improves the angular resolution of the two hardest subjets with respect to the two b -hadrons arising from the b -quark pair, which is a fundamental ingredient for b -tagging.

Finally, the cut at $p_T > 200$ GeV is re-applied to the momentum of the filtered Higgs boson candidate. The filtered four-momentum, computed from the three highest p_T filtered subjets, is considered in all subsequent steps.

7.4.2 Invariant mass resolution

At any stage of the jet clustering algorithm, the jet mass, m , is defined summing up the four momenta of all constituents considered for the clustering. The *topological clusters* used as basic input for the jet clustering algorithm are treated as massless. The energy of the topological clusters used as input to the jet clustering algorithm needs to be calibrated, to compensate for the different response of particles from the jet starting an electromagnetic or hadronic shower in the calorimeter and for the energy lost in the dead material of the calorimeter. As described in Sec 4.3, the default calibration procedure used in ATLAS relies on the H1-style based calibration, which is applied directly on the final jet and has two separate correction terms: a cell-by-cell correction applied on all cells which belong to the jet considered and an overall correction which depends on the jet transverse momentum and pseudo-rapidity. The second correction is essentially only an overall jet energy scale correction, which accounts also for the average out-of-cone contributions expected for jets of a predefined size. The H1-style calibration is applied here, but only the cell-by-cell correction is considered. An additional overall scale correction is considered later, in Section 7.9.

In order to estimate the achievable mass resolution, an unbinned likelihood fit to the mass distribution is applied, assuming the following functional form:

$$f(x; m, \sigma_{\pm}, \alpha_{\pm}) = \exp \left[-\frac{(x - m)^2}{2\sigma_{\pm}^2 + \alpha_{\pm}(x - m)^2} \right], \quad (7.2)$$

which, thanks to the parameters α_{\pm} , can account correctly also for the non-Gaussian tails of the distribution. The result is shown in Fig. 7.6, together with the fitted parameters. It can

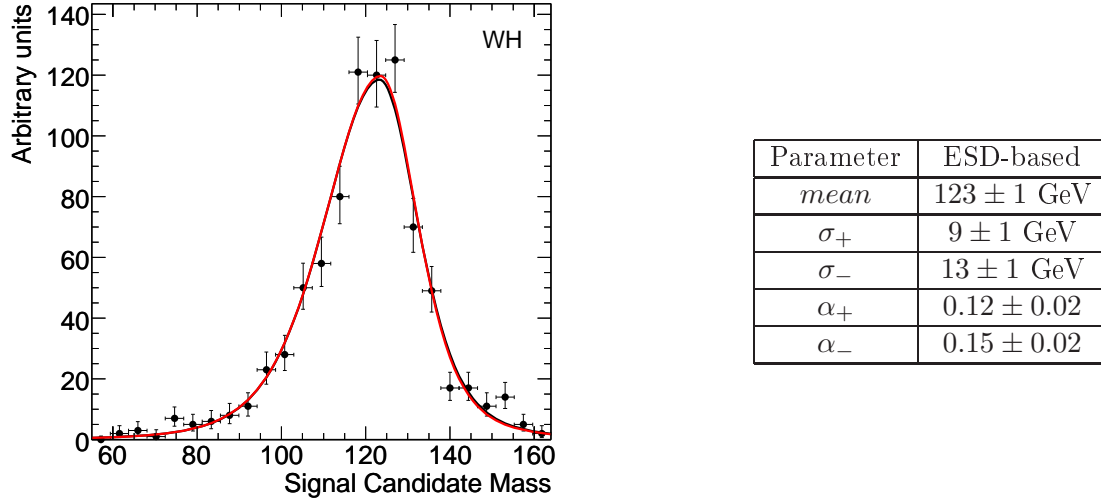


Figure 7.6: Distribution of the invariant mass (in GeV) of the Higgs boson candidate after the jet finding algorithm is applied on the $WH \rightarrow \ell\nu b\bar{b}$ signal sample, with $m_H = 120$ GeV (Monte Carlo data points, with fit superimposed). The full analysis selection, except for the mass window cut, is applied. The fit parameters are illustrated in the table on the right.

be noticed that the jet energy scale is slightly overestimated, resulting in an invariant mass of 123 ± 1 GeV, where it should be considered that a value lower than 120 GeV is expected, due to the frequent presence of neutrinos from the semileptonic decays of b - or c -hadrons in b -jets which remain undetected. The achievable core resolution is around 10%.

7.4.3 Identification of b -subjets

The identification of the two b -quarks originating from the Higgs boson is crucial for separating the signal from the large backgrounds, most of which are dominated by u , d and s -quark jets.

In order to apply b -tagging on them, tracks have to be associated separately to each of the two subjets. The simplest way to perform this task is to take all tracks whose momenta at the point of closest approach to the interaction point in the transverse plane are within a certain ΔR from the subjet direction as determined by the jet clustering algorithm, according to the strategy already defined for conventional jets in Section 6.2 of Chapter 6. However, for the two subjets from the Higgs boson candidate, the jet direction can be defined in different ways, therefore three strategies have been studied:

- Using the directions of the two unbranched subjets before the filtering procedure;
- Using the direction of two or three subjets as determined after filtering the unbranched subjets, applying b -tagging on all of them and choosing the two with highest b -jet probability;
- Using the direction of subjets as determined after filtering the unbranched subjets, choosing the two among them with the highest transverse momentum.

The subjet direction after filtering yields a better estimate of the direction of the outgoing b -hadron, since it filters out contributions from QCD radiation, and thus improves the b -tagging performance. At the same time, it turns out that applying b -tagging to all the subjets after filtering, and then choosing the two with highest b -weight, increases considerably the amount of $t\bar{t}$ background. So the last of the three listed strategies is adopted in the following.

For the distance in pseudo-rapidity ΔR between the subjet direction and the track momenta a fixed value of 0.4 is used, without any dependence on $\Delta R(\text{subjet 1, subjet 2})$, p_T or η of the two subjets: such a dependence might be used in future to improve further the b -tagging performance. Every track can be assigned to one only subjet, so in case of overlaps a track is assigned to the nearest in ΔR of the two subjets.

The direction of the subjet momentum has been corrected to take the displacement of the primary vertex in z with respect to the nominal position of the interaction point into account. In fact the direction associated with each calorimeter cell provides a good estimate of the momentum of a particle depositing its energy in the calorimeter only under the hypothesis that the particle originated in $(0,0,0)$. While the transverse displacement of the primary vertex with respect to the nominal position of the beam spot is negligible, the longitudinal displacement is not and can be accounted for with the correction:

$$\theta_{corrected} = \text{arccotan} \left(\cotan(\theta_{subjet}) - \frac{z_{PV}}{r_{cal}} \right)$$

where r_{cal} is the average distance in the transverse plane from $(0,0)$ of the energy deposits in the calorimeter. Even though for r_{cal} a jet based value could be used, here an average value of 1.750 m is used, which has been chosen by minimising the subjet direction root mean squared residual. This correction provides a small improvement in the subjet direction resolution with respect to the b -quark direction.

No attempt has been made in this study to combine the information coming from the *spatial* b -tagging algorithms (based on the b -hadron lifetime) with the b -tagging algorithms based on the reconstruction of a lepton from a semileptonic b -hadron decay.

After the tracks are associated to the subjets, b -tagging is applied on them, by using the two most powerful b -tagging algorithms available in ATLAS, already introduced in Chapter 6⁴. The two considered algorithms are *IP3D+SV1*, which combines the information from the impact parameter based algorithm *IP3D* with the inclusive secondary vertex based algorithm *SV1*, and the neural network based algorithm *IP3D+JetFitter*, which combines the same impact parameter based algorithm with the information provided by the decay chain fit of the *JetFitter* vertex reconstruction algorithm.

In order to improve the rejection against c -jets, the ability of the *JetFitter* based b -tagging algorithm to provide a discriminating variable which can be specifically optimised to achieve the best overall combined rejection of c - and light-jets is exploited, as described in Section 6.10.4. The discriminator is based on Eq. 6.39, which is reproduced here for simplicity:

$$\text{discr} = \frac{P(b)}{P(b) + c_l P(l) + (1 - c_l) P(c)} \quad (7.3)$$

Using the b -tagging procedure just defined, the b -tagging performance for the specific topological configuration and background flavour composition of this analysis is analysed in more detail in Section 7.7. As a result of this optimisation, for the final analysis selection the *IP3D+JetFitter* b -tagging algorithm based on the neural network approach, is applied on both subjets from the Higgs boson candidate, by applying a cut based on the discriminator of Eq. 7.3, chosen to correspond to a single b -subjet efficiency of $\approx 63\%$ (which corresponds to a $b\bar{b}$ tagging efficiency of 40%) and with the prior light jet background composition set to $c_l = 0.2$.

⁴The b -tagging algorithms are applied using the ATLAS reconstruction software ATHENA (release 14.0.0); the calibration is based on the release 13 calibration files for “conventional” cone jets with size $R = 0.4$.

7.5 Analysis selection and sensitivity based on event counting

In this section the procedure used to select final states compatible with the contribution expected from the $WH \rightarrow \ell\nu b\bar{b}$ signal is explained in detail. The selection is then applied on all simulated signal and background samples considered in this study and a first evaluation of the discovery sensitivity of this analysis to a Standard Model Higgs boson with a mass of 120 GeV is presented.

The procedure used to identify and separate the basic physics objects expected in a WH event is the following:

- The highest p_T muon or electron is identified.
- The Higgs boson candidate is selected using the procedure described in Section 7.4, including the application of b -tagging.
- Additional jets in the event are identified to be used for the additional jet veto, using a simple inclusive k_\perp algorithm with $R = 0.4$.
- Events with additional electrons or muons in the event are rejected.

For a muon to be accepted as the candidate lepton from the W boson decay, it must be reconstructed as a track segment both in the muon and in the inner detector system, and a good match is required between the two. In addition a loose isolation criterion is applied, so that that:

$$\frac{E_{T,cone}}{E_t(\mu)} < 25\%,$$

where $E_{T,cone}$ is the amount of transverse energy in the calorimeter in a cone of $\Delta R = 0.4$ around the track, extrapolated to the position where the track hits the calorimeter.

For an electron to be accepted as the candidate lepton from the W boson, it is required to pass the standard ATLAS medium quality electron identification (isEM=medium), as described in Section 4.2, and the calorimeter isolation requirement is:

$$\frac{E_{T,cone}(\Delta R = 0.2)}{p_T(e)} < 10\%.$$

For the leptons to be considered for the veto, they must not be in proximity to any of the two Higgs subjects (i.e. they must be further away than $\Delta R = 0.4$ from any of them or they have to pass the complementary isolation requirement); muons do not need to have combined inner detector - muon system tracks, while electrons only need to fulfil a looser electron identification (isEM=loose) requirement. Jets overlapping with the highest p_T lepton in the event ($\Delta R(\text{lepton}, \text{jet}) < 0.2$) are not considered for the jet veto.

After a basic definition of the physics objects, the event selection is applied. In this channel, two selections are defined. A “tight” set of cuts is used to allow a close comparison to the particle-level result of Ref. [1], based on a simple event-counting estimate of the significance. In addition a set of “loose” cuts is also used, in order to provide input to a more sophisticated log-likelihood fit, described in Chapter 8.

The basic selection cuts are listed in Table 7.2 and their impact on the number of expected events in the signal and the main background samples is illustrated. The cut on $\Delta\phi$ is motivated by the fact that the W and the H bosons are expected to be back-to-back in the transverse plane. In order for a b - or light jet to be considered for the veto, they are required not to overlap with any of the two leading filtered subjects from the Higgs candidate (the minimum ΔR allowed between the Higgs subjects and the k_T jets used for the veto has been

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	$W+jet$
After filter cuts	1402.3 \pm 7.2	1252.8 \pm 7.8	946.1 \pm 4.9	9331	1609356	169519	2433885
1 Higgs candidate	646.2 \pm 2.8	569.7 \pm 3.0	429.7 \pm 1.9	3509.7 \pm 8.0	806175	69375	562030
filtered $p_T(H) > 200$ GeV	581.7 \pm 2.9	512.7 \pm 3.2	387.6 \pm 2.0	3108 \pm 10	709271	60241	413406
Missing $E_T > 30$ GeV	413.7 \pm 2.9	362.4 \pm 3.2	273.6 \pm 2.0	2183 \pm 13	552284	46779	318400
$p_T(W) > 200$ GeV	194.3 \pm 2.4	171.0 \pm 2.6	128.0 \pm 1.6	1216 \pm 12	137946	18524	206331
$p_T(e/\mu) > 30$ GeV	166.2 \pm 2.2	145.6 \pm 2.4	108.1 \pm 1.5	996 \pm 11	115053	15724	178004
$p_T(\text{additional } \mu) < 10$ GeV	165.1 \pm 2.2	144.6 \pm 2.4	107.3 \pm 1.5	942 \pm 11	106836	14992	177542
$p_T(\text{additional } e) < 10$ GeV	162.9 \pm 2.2	142.9 \pm 2.4	105.8 \pm 1.5	885 \pm 11	97305	13881	174941
$\Delta\phi(W,H) > \frac{2}{3}\pi$	161.9 \pm 2.2	142.2 \pm 2.4	104.9 \pm 1.5	841 \pm 11	84773	12999	167704
no additional b -jets $p_T > 15$ GeV	150.7 \pm 2.2	130.6 \pm 2.3	95.1 \pm 1.4	790 \pm 10	30605	7805	160608
add. jets on W side $p_T < 60$ GeV	133.2 \pm 2.1	115.7 \pm 2.2	83.6 \pm 1.4	637.2 \pm 9.5	19422	5870	121437
add. jets on H side $p_T < 60$ GeV	119.8 \pm 2.0	102.7 \pm 2.1	75.2 \pm 1.3	525.6 \pm 8.8	13841	4370	94055
one subjet b -tagged	108.2 \pm 1.9	91.4 \pm 2.0	66.8 \pm 1.2	126.1 \pm 4.5	8638	2421	6964
both subjets b -tagged	54.3 \pm 1.4	45.6 \pm 1.4	32.85 \pm 0.89	43.7 \pm 2.7	576	161.4 \pm 7.0	266
loose fit cuts	54.2 \pm 1.4	45.4 \pm 1.4	32.75 \pm 0.89	43.0 \pm 2.7	565	156.3 \pm 6.9	257

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jet
After filter cuts	574.2 \pm 3.3	25140	8079	3566400	649215
1 Higgs candidate	295.9 \pm 1.3	8428.0 \pm 5.0	3372.1 \pm 8.1	160154	225597
filtered $p_T(H) > 200$ GeV	267.8 \pm 1.4	7355 \pm 12	2993.4 \pm 9.4	103170	175061
Missing $E_T > 30$ GeV	141.9 \pm 1.3	6217 \pm 15	1414 \pm 10	75129	62883
$p_T(W) > 200$ GeV	33.86 \pm 0.76	4166	457.5 \pm 6.7	3320	13514
$p_T(e/\mu) > 30$ GeV	9.06 \pm 0.41	3463	264.2 \pm 5.3	2899	10135
$p_T(\text{additional } \mu) < 10$ GeV	5.69 \pm 0.32	3441	177.9 \pm 4.4	2508	7174
$p_T(\text{additional } e) < 10$ GeV	3.56 \pm 0.26	3361	121.7 \pm 3.6	2037	4741
$\Delta\phi(W,H) > \frac{2}{3}\pi$	3.43 \pm 0.25	3257	112.0 \pm 3.5	1638	4081
no additional b -jets $p_T > 15$ GeV	3.04 \pm 0.24	3128	102.4 \pm 3.3	494	3751
add. jets on W side $p_T < 60$ GeV	2.35 \pm 0.21	2525	82.2 \pm 3.0	323	2746
add. jets on H side $p_T < 60$ GeV	2.15 \pm 0.20	2083 \pm 14	68.9 \pm 2.8	233	2146
one subjet b -tagged	1.94 \pm 0.19	269.9 \pm 5.8	16.8 \pm 1.4	134 \pm 14	207 \pm 13
both subjets b -tagged	0.96 \pm 0.13	5.43 \pm 0.84	5.26 \pm 0.77	7.6 \pm 3.4	12.5 \pm 3.1
loose fit cuts b -tagged	0.96 \pm 0.13	5.43 \pm 0.84	5.26 \pm 0.77	7.6 \pm 3.4	12.5 \pm 3.1

Table 7.2: Number of events passing the selection at each stage for the signal processes of different Higgs boson masses and the background processes. The numbers are projected to 30 fb⁻¹ of collected data, based on LO MC cross sections.

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	$W+jet$
add. jets on W side $p_T < 20$ GeV	98.9 \pm 1.8	83.2 \pm 1.9	62.5 \pm 1.2	461.3 \pm 8.3	7227	3343	86087
add. jets on H side $p_T < 20$ GeV	67.0 \pm 1.5	55.8 \pm 1.6	41.17 \pm 0.99	275.6 \pm 6.6	1895	1142	48229
one subjet b -tagged	55.5 \pm 1.4	46.4 \pm 1.5	33.50 \pm 0.90	49.8 \pm 2.9	986	498 \pm 12	1825
both subjets b -tagged	23.05 \pm 0.91	19.51 \pm 0.96	13.98 \pm 0.59	16.5 \pm 1.7	38.9 \pm 4.9	18.2 \pm 2.4	87.3 \pm 9.0
112 GeV < mass(H) < 136 GeV	-	13.25 \pm 0.79	-	1.18 \pm 0.45	5.6 \pm 1.9	4.2 \pm 1.1	8.3 \pm 2.8

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jet
add. jets on W side $p_T < 20$ GeV	1.26 \pm 0.15	1835 \pm 14	46.3 \pm 2.3	87 \pm 12	1571
add. jets on H side $p_T < 20$ GeV	0.93 \pm 0.13	1134 \pm 11	28.4 \pm 1.8	18.3 \pm 5.3	850
one subjet b -tagged	0.78 \pm 0.12	68.9 \pm 3.0	6.26 \pm 0.84	7.6 \pm 3.4	49.9 \pm 6.2
both subjets b -tagged	0.333 \pm 0.079	0.65 \pm 0.29	1.68 \pm 0.43	1.5 \pm 1.5	3.9 \pm 1.7
112 GeV < mass(H) < 136 GeV	0.278 \pm 0.072	0.13 \pm 0.13	0.11 \pm 0.11	0.0 \pm 1.5	0.78 \pm 0.78

Table 7.3: Expected number of events with tighter selection cuts for an integrated luminosity of 30 fb⁻¹.

optimised to a value of 0.32). The b -jet veto is applied at a b -tagging efficiency of $\approx 75\%$, so $\approx 25\%$ of the b -jets are supposed to escape the veto. For convenience the veto on light jets in the table is divided into a veto on either the Higgs boson or W boson side, according to which of the two is nearer to the jet to veto in the transverse plane. Finally the values of Higgs boson mass, $p_T(\text{Higgs})$, $p_T(\text{additional jet})$ and $\Delta\eta(W,H)$ are required to be within the range to be used for the maximum likelihood fit.

The dominant backgrounds are $W+jet$, $t\bar{t}$ and Wt . The $W+jet$ background is strongly reduced by the requirement of two b -tagged subjets, so the remaining background is dominated by the irreducible $Wb\bar{b}$ contribution. The case of $t\bar{t}$ and Wt is more complex, since the presence of b and c quarks in the final state makes b -tagging less powerful in rejecting this background; however, in particular for $t\bar{t}$, vetoing additional b -jets and the remaining well identified light-jets in the event down to a certain p_T can suppress a good part of this contribution.

The *tight* selection is applied on top of the *loose* one and is shown in Table 7.3. In this case a harder jet veto is applied, vetoing any additional jets with $p_T < 20$ GeV and up to $|\eta| < 5$, and a tighter b -tagging requirement, corresponding to a $b\bar{b}$ tagging efficiency of 40%.

Since most of the Monte Carlo samples used were produced with a generator filter cut on the two final state partons produced by the $2 \rightarrow 2$ matrix element at $\hat{p}_T^{\min} = 150$ GeV (except for the $t\bar{t}$ background), it is reasonable to look at the shape of the variables used for the selection only after the bias introduced by this filter is removed. This is most effectively done requiring a single high p_T Higgs boson candidate to be reconstructed with $p_T > 200$ GeV and on the other side a W boson with $p_T > 200$ GeV (first four cuts in Table 7.2). After these first cuts are applied, the distributions for $p_T(W)$ and for E_T^{miss} are as shown in Fig. 7.7. The distributions are not too different between the signal and the background components, since in nearly all cases the dominant contribution is from real W bosons. In all distributions shown here and in the following, the WZ , ZZ and WW background components are summed up together in the di-boson VV background, the W +jet and the Z +jet backgrounds are included in the V +jet background, and the ZH sample passing through the selection is added up to the WH signal. It should, however, be kept in mind that the background is dominated by W +jet, $t\bar{t}$, Wt and WZ at any stage of the analysis.

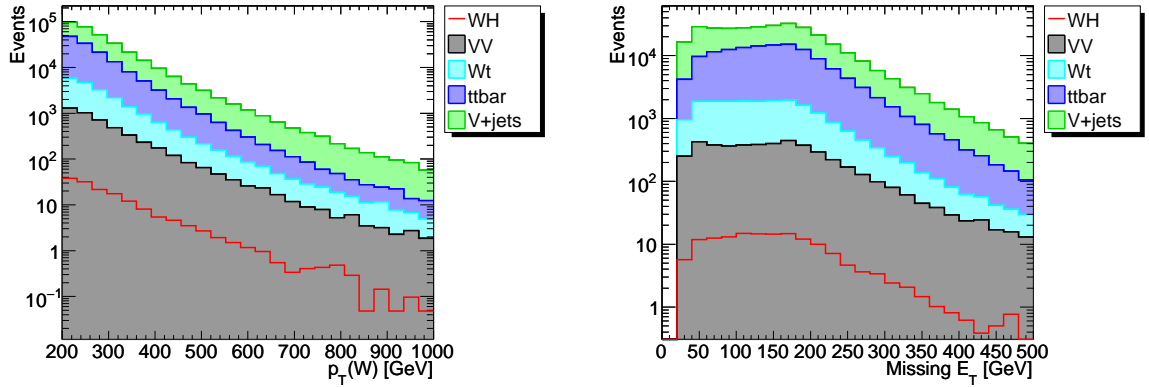


Figure 7.7: Distribution for $p_T(W)$ (left) and E_T^{miss} (right) after the first four cuts of Table 7.2. The backgrounds are added on top of each other, to provide the overall background shape. The signal, with $m_H = 120$ GeV, is just added in foreground.

Fig. 7.8 (left) shows the distribution of the difference in azimuthal angle $\Delta\phi$ between the Higgs and the W boson candidates, after the lepton veto cuts have been applied. As can be seen from the signal distribution, a cut at $\Delta\phi(W, H) > \frac{2}{3}\pi$ is extremely conservative, however, this value has been chosen keeping in mind eventual NLO effects, due to the possible presence of additional hard QCD radiation in the signal which would decrease the rate at which the W and Higgs bosons are emitted exactly back-to-back⁵.

Fig. 7.8 (right) shows the distribution of additional b -jets in the event, after having applied the $\Delta\phi$ cut. Fig. 7.9 shows the b -weight distributions, after having applied the jet veto cut of the *tight* selection (at 20 GeV). The b -weight is defined as $\log(\text{discr})$, where discr is defined in Eq. 7.3 and a value of $c_l = 0.2$ is used: b -tagging is applied in the *tight* selection by requiring the b -weight to be below 1.5. The first selection cut on the highest of the two b -weights of the two subjects from the Higgs boson candidate dramatically reduces the very large backgrounds

⁵A more recent evaluation of possible NLO effects on the signal using the MC@NLO generator seems to exclude this, so in the future a tighter cut could be used to try to increase the analysis significance.

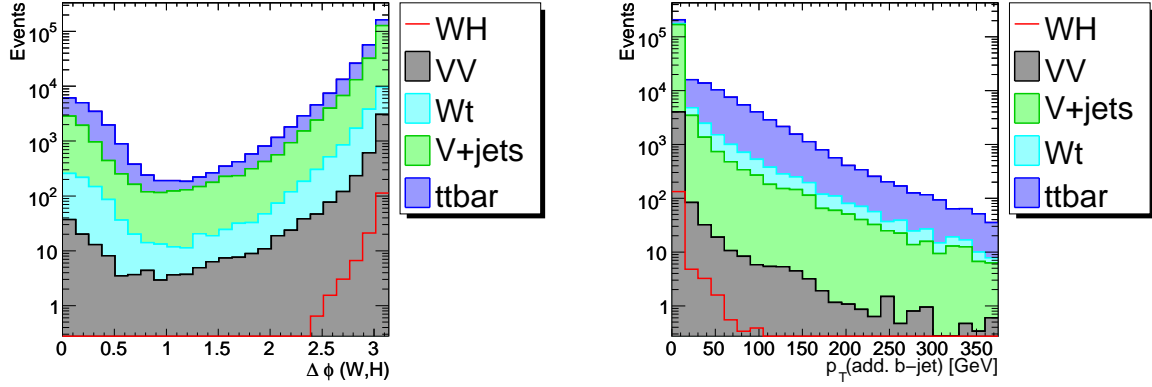


Figure 7.8: Distributions for $\Delta\phi$ between the Higgs boson and the W boson (left) and the distribution of transverse momenta of additional b -jets in the event, on which to apply the b -jet veto (right). For the signal events $m_H = 120$ GeV.

from W +light jet and, in general, backgrounds with two light jets in the final state, while the remaining cut on the lowest of the two b -weights can also help against the $t\bar{t}$ background, where one of the two subjets from the Higgs candidate is typically a real b -jet, but the other one is either a c - or a light-jet.

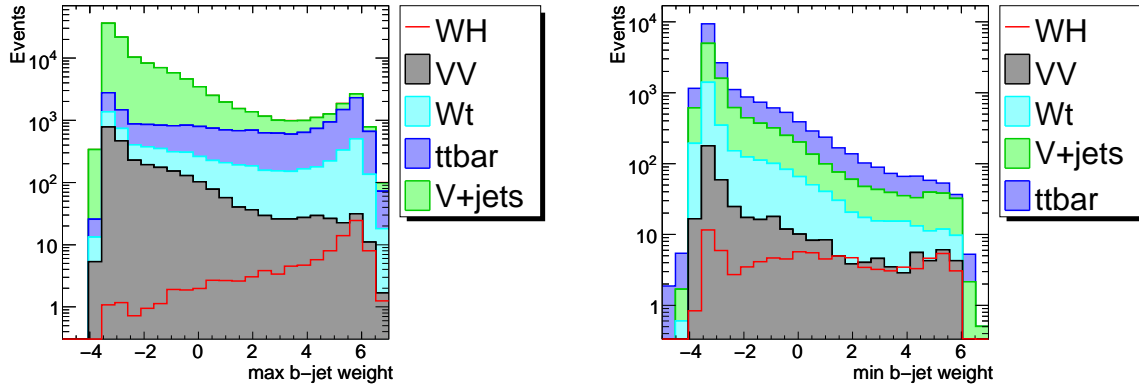


Figure 7.9: Distribution for the highest (left) and the lowest (right) of the b -weights of the two Higgs boson candidate subjets. The right distribution is shown after having applied the *tight* selection cut (at 1.5) on the distribution on the left. For the signal events $m_H = 120$ GeV.

Fig. 7.10 shows how the mass distribution evolves in the four main stages of the analysis selection. It is interesting to note how the jet veto significantly reduces the $t\bar{t}$ background, while the b -tagging requirement is really a crucial ingredient in this analysis to reduce the backgrounds to a manageable level.

The final mass distribution of the Higgs boson candidates is shown in Fig. 7.11, after applying either the loose or the tight analysis selection. In the latter case, both Higgs and Z bosons peaks are visible on top of the $t\bar{t}$, Wt and W +jet backgrounds. Since the statistics for the reference HERWIG sample used for the W +jet background component is rather low, the same distributions are also shown adopting the AcerMC sample for the dominant $Wb\bar{b}$ component, in order to cross-check the mass distribution. A detailed analysis of the differences in terms

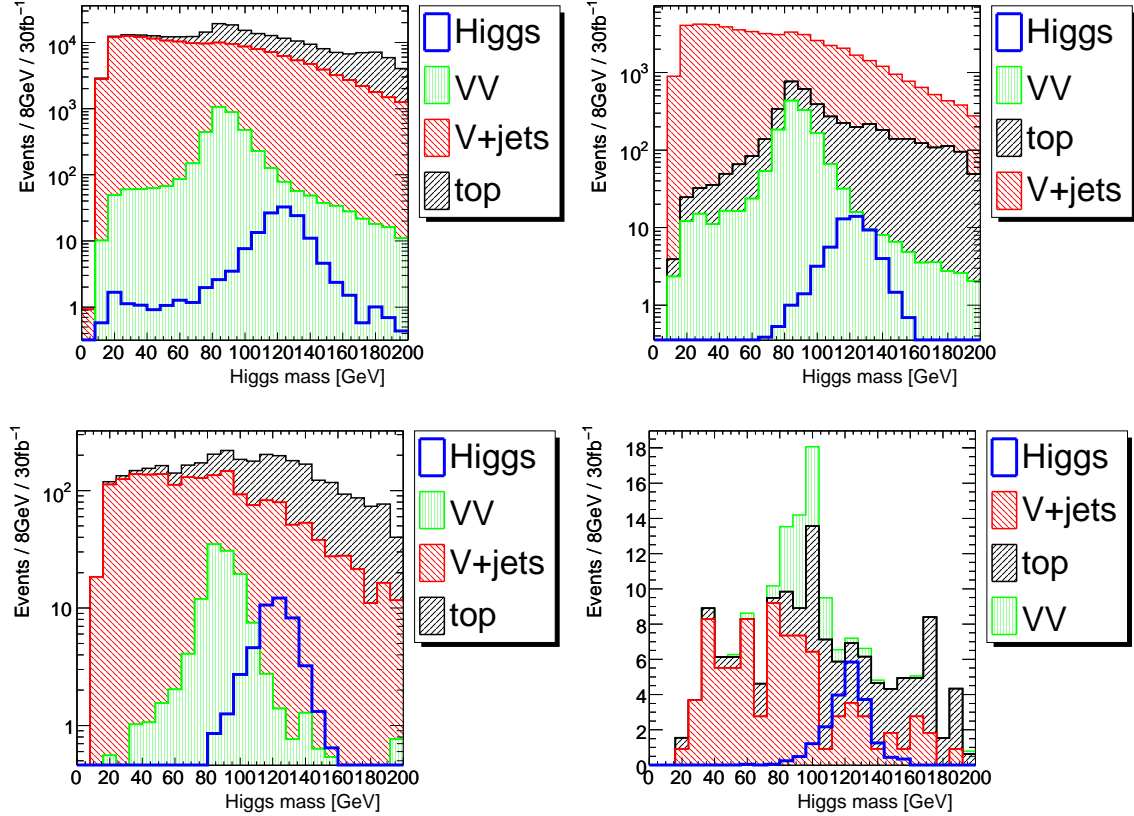


Figure 7.10: Distribution of the invariant mass of the Higgs boson candidate after different stages of the *tight* analysis selection: after the kinematic selection cuts on the W and Z bosons (top left), after the additional jet vetoes (top right), after b -tagging is applied on one subjet (bottom left) and after b -tagging is finally applied on both subjets from the Higgs candidate (bottom right). The backgrounds are stacked on top of each other, the WH signal (for $m_H = 120$ GeV) is shown separately in foreground. The Wt and $t\bar{t}$ backgrounds are labelled as *top* background. All distributions are shown in logarithmic scale, except for the last one, and normalised to an integrated luminosity of 30 fb^{-1} , based on LO cross sections.

of analysis selection between the $Wb\bar{b}$ background simulated with the HERWIG (PS approach) and with the AcerMC (ME approach) samples is illustrated in the appendix (Section C.2).

The resulting analysis significance, considering a mass range of 112-136 GeV, in terms of $\frac{S}{\sqrt{B}}$ is 3.0 ± 0.3 , where the quoted uncertainty comes from the limited available Monte Carlo statistics, and the signal-to-background ratio is $\frac{S}{B} \simeq \frac{2}{3}$.

The impact of trigger efficiency and of systematic uncertainties are not included in this first estimate of the discovery significance of this channel. The first will be estimated in the next section, the second in the next chapter.

7.6 Trigger selection

The WH analysis relies on the identification of a muon or an electron with $p_T > 30$ GeV. These leptons are of sufficiently high transverse momenta that most of the events should pass the online trigger selection for instantaneous luminosities up to $2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. In fact up to this level of instantaneous luminosity the following single-lepton trigger items are expected to stay unrescaled according to Ref. [44]:

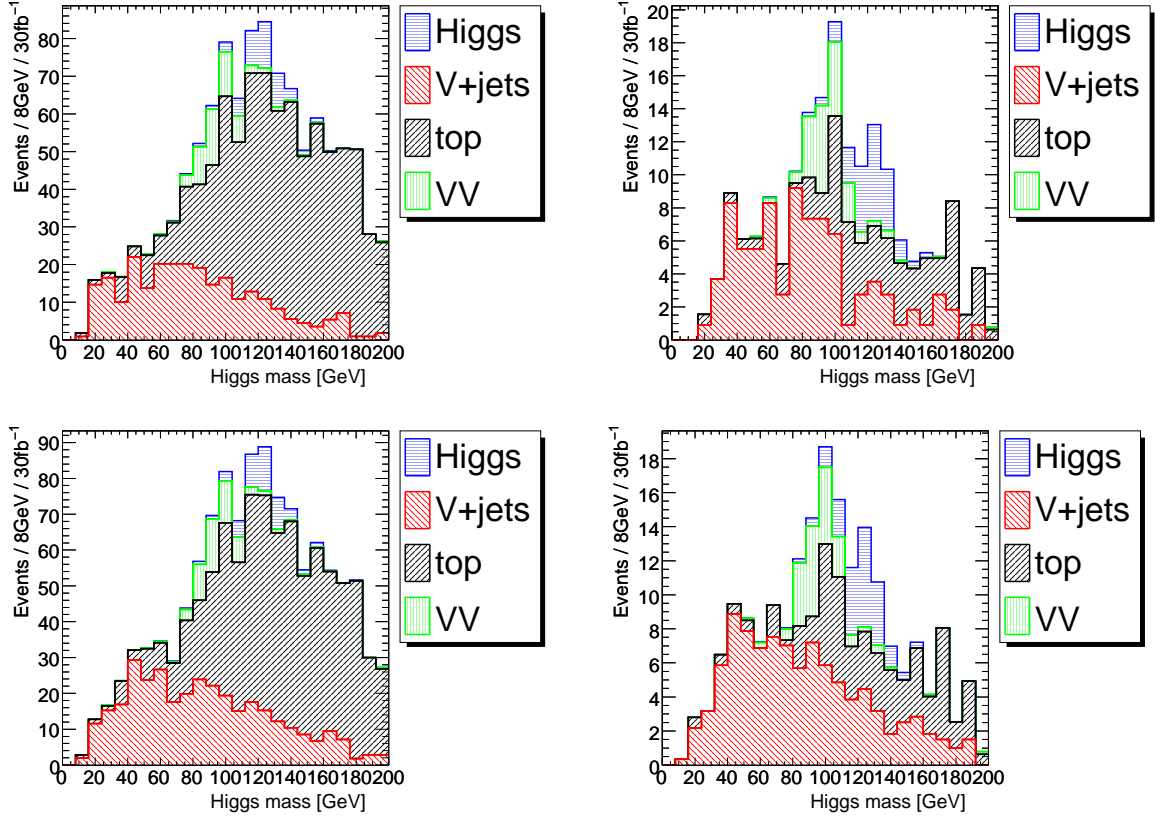


Figure 7.11: Distribution of the invariant mass of the Higgs boson candidate after all selection cuts for the *loose* (left) and *tight* (right) selection have been applied, using the reference HERWIG sample for the $W + jet$ background (top) or, in alternative, using the higher statistics AcerMC sample for the $Wb\bar{b}$ background component (bottom). In the second case the minor contribution from the W +light jet background is not considered in the plot. The WH signal (for $m_H = 120$ GeV) is shown on top of the backgrounds. The Wt and $t\bar{t}$ backgrounds are labelled as *top* background. All distributions are normalised to an integrated luminosity of 30 fb^{-1} , based on LO cross sections.

- $e25i$
- $\mu 20i$
- $e105$
- $\mu 40$

The first two items refer to electrons and muons with a trigger threshold of 25 and 20 GeV, respectively; in both cases, in order for these triggers to stay unscaled, a lepton isolation criterion is applied. In addition, trigger items with higher thresholds for both electrons and muons are considered; they are needed, since the isolation criteria of the trigger items with lower thresholds tend to be too strict for leptons with transverse momenta sufficiently far from the low- p_T thresholds.

Fig. 7.12 shows the efficiency for electrons that satisfy the offline medium electron identification requirement. For electrons with true p_T above 25 GeV, the $e25$ trigger, without isolation requirement, has an inclusive efficiency of $(98.6 \pm 0.2)\%$. As expected, the isolation requirement in $e25i$ tends to remove electrons at transverse momenta higher than 150 GeV, so that

the inclusive efficiency gets reduced to $(94.4 \pm 0.4)\%$. This loss can be addressed by adding electrons selected by the $e105$ trigger, which provides an inclusive efficiency of $(92.7 \pm 0.6)\%$ for electrons with true p_T above 105 GeV and reaches $\approx 100\%$ efficiency for electrons with true p_T above 200 GeV.

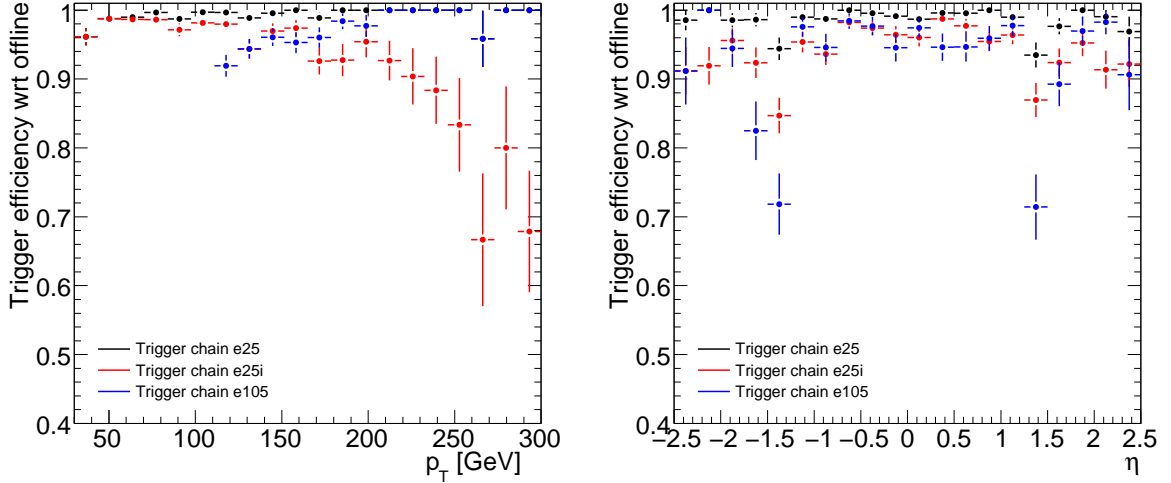


Figure 7.12: Single electron trigger efficiency for offline reconstructed electrons passing the medium electron identification requirement, as a function of p_T (left) and pseudo-rapidity (right). For the pseudo-rapidity distribution a cut on the transverse momentum of the Monte Carlo truth matched electron was done, corresponding to the threshold of the trigger item considered (at 25 and 105 GeV).

The efficiency of different muon triggers is shown in Fig. 7.13 as a function of pseudo-rapidity and transverse momentum, measured for offline-reconstructed muons, successfully combined with inner detector tracks. For muons with true p_T above 20 GeV, the overall efficiency of the $mu20$ trigger is $(81.0 \pm 0.6)\%$, which is not high, but is expected, due to the limited geometric acceptance of the Level 1 (L1) muon trigger. Further losses due to the additional isolation requirement bring the efficiency down to $(52.8 \pm 0.8)\%$ for the $mu20i$ trigger⁶. This loss can also be recovered from quite effectively by using the $mu40$ trigger in addition, which itself provides an efficiency of $(77.6 \pm 0.7)\%$ for muons with true p_T higher than 40 GeV.

The impact of the trigger requirements on the WH analysis can be investigated by analysing the impact of the trigger requirement on the full *loose* offline selection of the analysis. The effects of three possible trigger strategies are shown for the WH signal in Table 7.4.

The lepton trigger combination referred to in the Table is an OR combination of the $mu20i$, $mu40$, $e25i$ and $e105$ trigger items. The efficiency for the same combination, but with the isolation requirements removed, is also shown in order to give an idea on the impact of the isolation on the trigger efficiency: while replacing $e25i$ with $e25$ does not make much of a difference in the overall efficiency, removing the muon isolation brings a small improvement of about 1%. Therefore one can conclude from here that the overall efficiency expected from the lepton triggers is around 90%.

On the other hand, the remaining 10% is unlikely to be recovered with any combination of lepton triggers, since it is caused by the limited acceptance of the L1 trigger muon chambers

⁶This loss in efficiency is most probably due to an inconsistency in the detector geometry used in the trigger simulation and is therefore not expected on real data: the real loss in efficiency expected due only to the isolation requirement on $mu20$ -selected events is around 5%. Since this loss is only marginally affecting the final trigger selection efficiency of the analysis, this problem can be safely ignored.

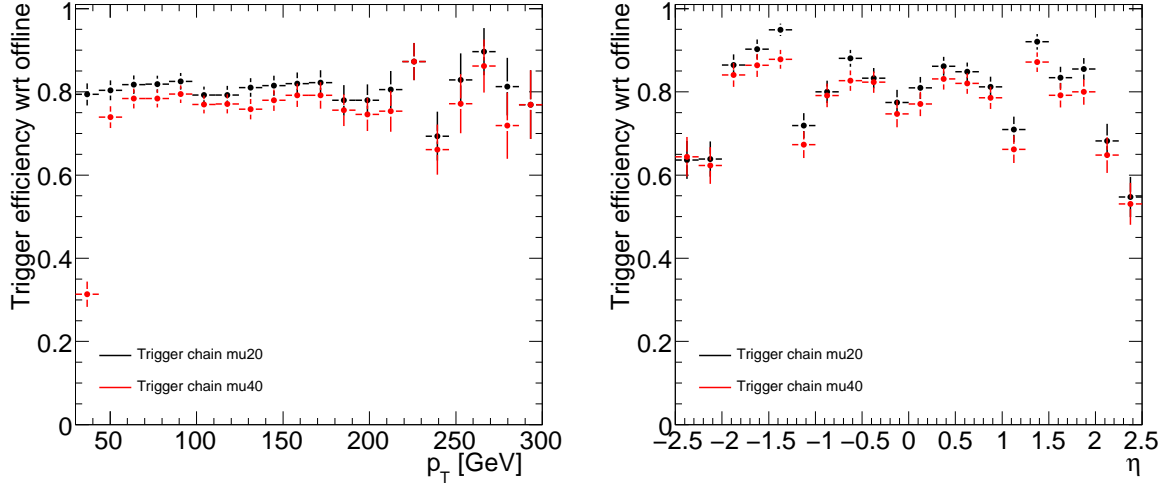


Figure 7.13: Trigger efficiencies for single-muon trigger items, measured for offline combined muons, as a function of p_T (left) and pseudo-rapidity (right). For the pseudo-rapidity plot, a cut on the transverse momentum of the Monte Carlo truth-matched electron was done, corresponding to the nominal threshold used in the trigger item (at 20 and 40 GeV).

	No trigger	Lepton triggers	Lepton triggers w/o isolation	Lepton + MET+jet triggers
After filter cuts	849.2 ± 7.6	849.2 ± 7.6	849.2 ± 7.6	849.2 ± 7.6
1 Higgs candidate	551.6 ± 3.6	232.5 ± 3.4	287.4 ± 3.6	313.5 ± 3.7
filtered $p_T > 200$ GeV	497.7 ± 3.7	210.1 ± 3.3	260.1 ± 3.5	285.0 ± 3.6
$E_T^{\text{miss}} > 30$ GeV	355.5 ± 3.7	183.2 ± 3.1	212.4 ± 3.3	254.3 ± 3.5
$p_T(W) > 200$ GeV	167.8 ± 3.0	130.2 ± 2.7	134.8 ± 2.8	166.0 ± 3.0
$p_T(e/\mu) > 30$ GeV	140.6 ± 2.8	123.9 ± 2.7	125.9 ± 2.7	139.7 ± 2.8
$p_T(\text{additional } \mu) < 10$ GeV	139.8 ± 2.8	123.2 ± 2.7	125.1 ± 2.7	138.8 ± 2.8
$p_T(\text{additional } e) < 10$ GeV	137.5 ± 2.8	121.1 ± 2.7	123.1 ± 2.7	136.5 ± 2.8
$\Delta\phi(W,H) > \frac{2}{3}\pi$	136.6 ± 2.8	120.5 ± 2.6	122.4 ± 2.7	135.6 ± 2.8
no additional b -jets $p_T > 15$ GeV	123.8 ± 2.7	109.4 ± 2.5	111.3 ± 2.6	122.9 ± 2.7
add. jets on W side $p_T < 60$ GeV	109.7 ± 2.5	97.2 ± 2.4	98.7 ± 2.4	109.0 ± 2.5
add. jets on H side $p_T < 60$ GeV	98.9 ± 2.4	87.5 ± 2.3	89.0 ± 2.3	98.3 ± 2.4
one subjet b -tagged	88.8 ± 2.3	78.8 ± 2.2	80.0 ± 2.2	88.2 ± 2.3
both subjects b -tagged	44.6 ± 1.7	39.8 ± 1.6	40.4 ± 1.6	44.2 ± 1.7
loose fit cuts	44.4 ± 1.7	39.7 ± 1.6	40.3 ± 1.6	44.1 ± 1.7
$112 < m_H < 136$ GeV	29.5 ± 1.4	26.1 ± 1.3	26.5 ± 1.3	29.2 ± 1.4
trigger efficiency	-	$(88.1 \pm 0.6)\%$	$(89.5 \pm 0.6)\%$	$(99.4 \pm 0.2)\%$

Table 7.4: Expected number of events through the various analysis selection criteria for the WH signal, compared for different trigger selection strategies. The numbers are projected to 30 fb^{-1} of collected data.

(exclusively TGCs and RPCs, as described in Section 3.2.3). Therefore, trigger items from the *jetTauEtmiss* trigger slice⁷ are investigated. These items are mainly based on the transverse missing energy signature. While most jet triggers are expected to be heavily prescaled up to very high momenta, a jet AND E_T^{miss} combination trigger with relatively low thresholds, $XE70 \cdot J62$, is expected to be among those that will remain unprescaled at $2 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ [44]. This signature is interesting not only when a genuine E_T^{miss} signature is present, but also

⁷A slice is a group of trigger items based on similar signatures.

when muons get lost due to L1 acceptance, creating fake missing momentum. It is not yet clear whether recovering high- p_T muons this way is actually foreseen by the ATLAS High Level Trigger strategy, but, as can be seen in Fig. 7.14, it indeed appears to be quite effective. After combining the lepton triggers ($mu20i + mu40 + e25i + e105$) with the $(E_T^{\text{miss}}\text{-jet})$ -trigger an overall trigger efficiency of $(99.4 \pm 0.2)\%$ is obtained for the offline selected events in the analysis, as reported also in the last column of Table 7.4.

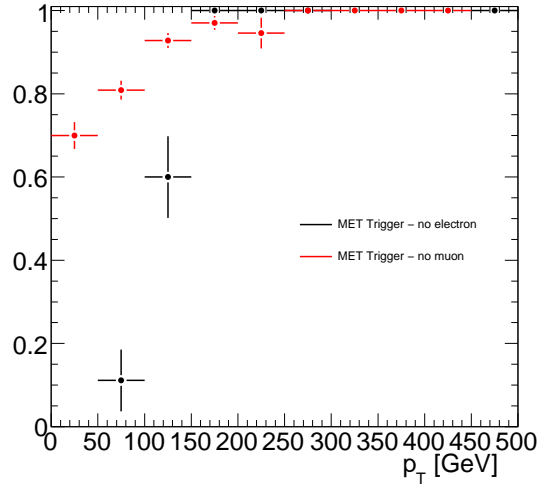


Figure 7.14: Efficiency of the $(E_T^{\text{miss}}\text{-jet})$ -trigger as a function of the true p_T of the lepton from the W boson, for all events which failed to pass the lepton triggers, at the analysis stage immediately after generator filter cuts.

In conclusion, due to the presence of high- p_T leptons and high E_T^{miss} , no significant loss in the discovery sensitivity is expected due to the trigger selection. However, a loss in signal efficiency of $\approx 0.6\%$ will be considered for completeness in the log-likelihood fit based study discussed in Chapter 8.

7.7 Optimisation of b -tagging performance

The b -tagging performance has been analysed by considering the two highest p_T subjects from all Higgs boson candidates present in the three following event samples: $WH(m_H = 120 \text{ GeV})$, $t\bar{t}$ and $W+\text{jet}$. All selection cuts have been applied, apart from the additional jet veto and the b -tagging cuts, and the mass window cut has been loosened to $[100\text{-}140] \text{ GeV}$, to increase the available statistics, still remaining in the kinematic region relevant for the present analysis.

The rejection achievable against light- and c -quark jets is shown in Fig. 7.15 as a function of the b -tagging efficiency on the b -subjett. The performance of three different b -tagging discriminators is analysed:

- $\log(LR)$ (Eq. 6.37) based on the $IP3D+SV1$ algorithm (also defined as COMB)
- $P(b)/[P(b) + P(l)]$ (using the neural network) based on $IP3D+JetFitter$, intended to provide optimal discrimination against light-quark jets
- $P(b)/[P(b) + P(c)]$ (using the neural network) based on $IP3D+JetFitter$, intended to provide optimal discrimination against c -quark jets

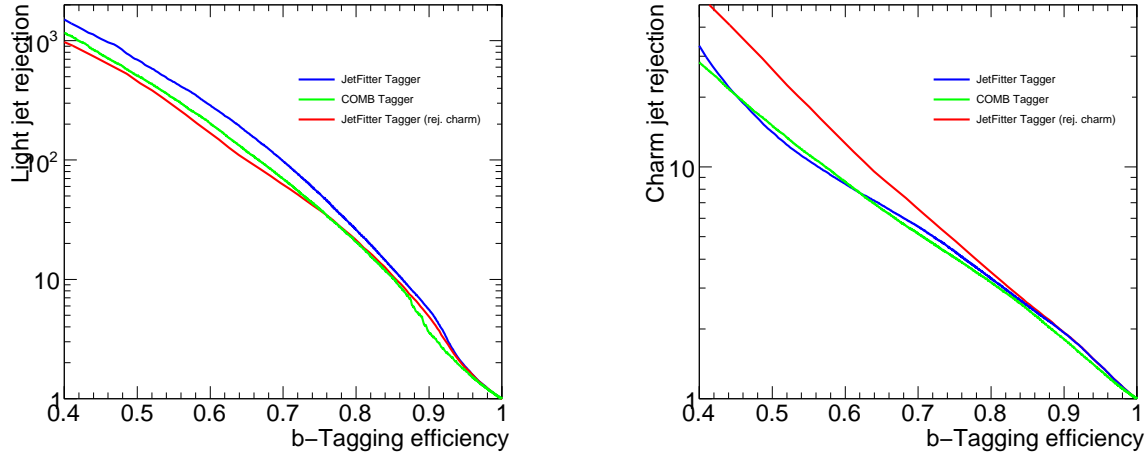


Figure 7.15: Rejection against light-quark jets (left) and against charm-quark jets (right) as a function of the b -tagging efficiency on the single subjet.

Looking at the performance of both $SV1$ and $JetFitter$ based algorithms, it turns out that both perform very well, with $JetFitter$ performing better in rejecting light-quark jets, in particular at b -tagging efficiencies below 80%.

In Table 7.5 the rejection against light-quark jets for few fixed b -tagging efficiency points is listed explicitly. The region of b -tagging efficiencies which is relevant for the present analysis is between 60% and 75% (which would yield a signal efficiency on the Higgs boson candidate, under the hypothesis of uncorrelated b -weights, between 36 and 56%).

b -tagging efficiency	$IP3D+SV1$	$IP3D+JetFitter$ (against light)	$IP3D+JetFitter$ (against charm)
40 %	1162 ± 143	1450 ± 75	906 ± 37
50 %	510 ± 16	691 ± 25	441 ± 13
60 %	203 ± 4	287 ± 7	160 ± 3
70 %	69 ± 1	98 ± 1	60.6 ± 1
80 %	20.6 ± 0.1	26.1 ± 0.2	21.1 ± 0.1

Table 7.5: Rejection against light-quark jets at different subjet b -tagging efficiencies. The various b -tagging algorithms used are described in the text.

At 70% b -tagging efficiency (corresponding to $\approx 50\%$ signal efficiency), a rejection of light quark jets around 100 is expected to be achievable using $JetFitter$, corresponding to 1% light-jet misidentification efficiency. This is exactly the value considered in a first draft of the hadron level study of Ref. [1] in the most optimistic scenario.

Such an efficient rejection of the background is however only valid in the hypothesis that the background is dominated by light-jets, so that charm-jets do not play an important role. However in the WH analysis, apart from the large irreducible $Wb\bar{b}$ component, the background is dominated by combinations of a light- and a charm-jet faking the two subjets representing the decay products of the Higgs boson, so the charm-jets are indeed expected to significantly decrease the non b -jet rejection.

The c -jet rejection values for some fixed efficiency points are also listed in Table 7.6. By us-

b -tagging efficiency	$IP3D+SV1$	$IP3D+JetFitter$ (against light)	$IP3D+JetFitter$ (against charm)
40 %	28 ± 1	31 ± 1	51 ± 1
50 %	15.0 ± 0.2	14.1 ± 0.2	25.6 ± 0.5
60 %	8.6 ± 0.1	8.4 ± 0.1	12.2 ± 0.2
70 %	5.15 ± 0.04	5.52 ± 0.04	6.47 ± 0.06
80 %	3.18 ± 0.02	3.30 ± 0.02	3.48 ± 0.02

Table 7.6: Rejection against c -quark jets at different subset b -tagging efficiencies. The various b -tagging algorithms used are described in the text.

ing the discriminator explicitly optimised to reject c -jets, it is possible to significantly increase the c -jet rejection, at the cost of decreasing the light jet rejection.

The optimal combined rejection against light and c -jets can be obtained by using Eq. 7.3:

$$\text{discr} = \frac{P(b)}{P(b) + c_l P(l) + (1 - c_l) P(c)},$$

where the optimal choice for the coefficient c_l depends on the flavour content of the background in term of fraction of c - and light-jets.

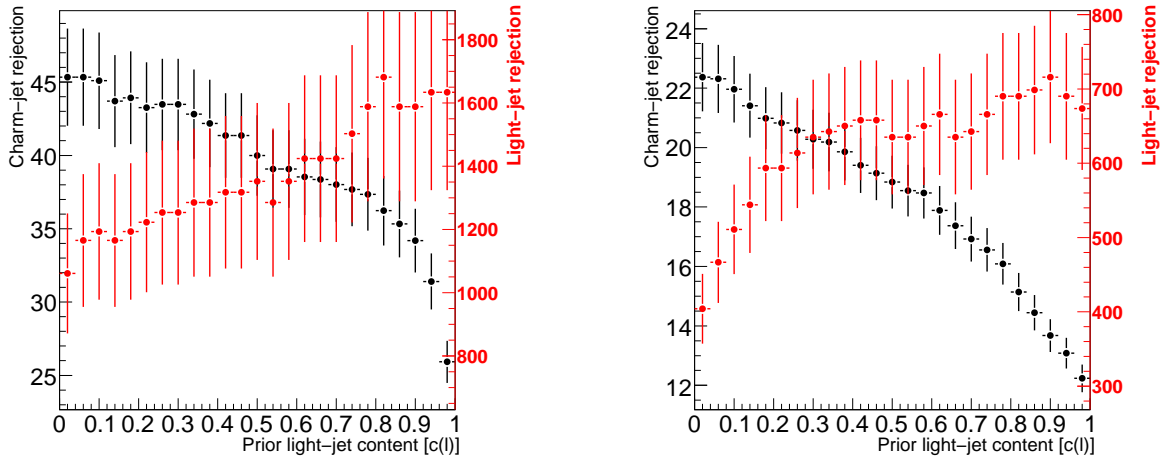


Figure 7.16: Rejection of c -jets (black) and light-jets (red) as a function of the coefficient c_l for b -subject efficiencies of 40 (left) and 50 (right) %.

The performance in terms of light- and c -jet rejection achievable at fixed b -tagging efficiencies of 40, 50, 60 and 70% as a function of the coefficient c_l is shown in Figs. 7.16 and 7.17. At all efficiencies shown except for 40%, it seems that the light-jet rejection saturates at a value of c_l around 50% (0.5). This is mainly due to the fact that the calibration of the algorithm was performed on generic Monte Carlo samples and on “conventional” jets in the presence of quite a different event and jet topology. As a result, the c_l no longer has a clear statistical interpretation and has to be considered as an effective parameter: at each possible b -tagging efficiency, given a certain background composition, there is one only value for c_l which minimises the amount of background.

As was shown in Table 7.3, where the tighter selection cuts were applied, the most important backgrounds are $t\bar{t}$ and W +jet. In order to understand the impact of b -tagging on the analysis,

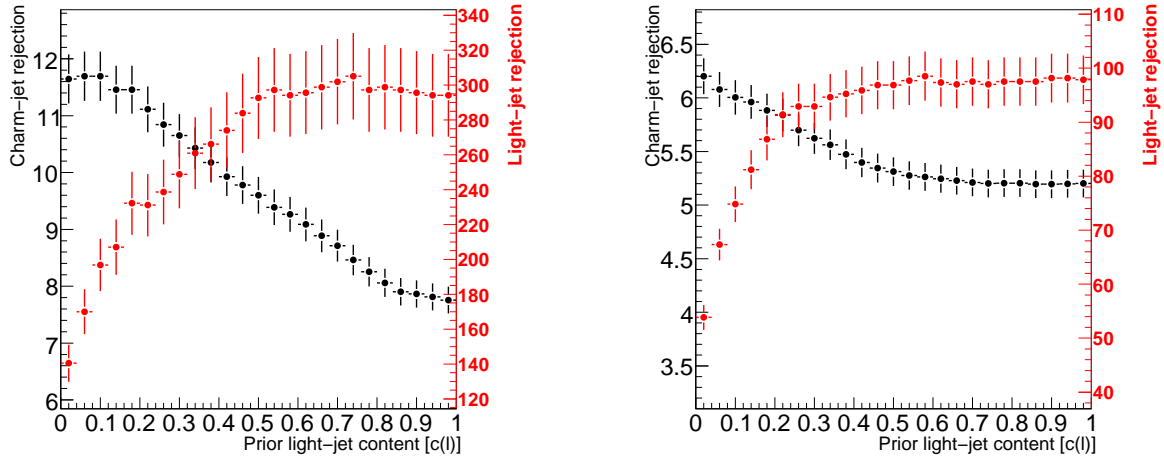


Figure 7.17: Rejection of c -jets (black) and light-jets (red) as a function of the coefficient c_l for b -subject efficiencies of 60 (left) and 70 (right) %.

it is worth analysing the flavour composition of the Higgs boson candidates selected in these two main backgrounds, and the breakdown of the different subjects flavour combinations as a function of the signal efficiency corresponding to a certain b -tagging cut on the two subjects. This is shown in Fig. 7.18 for $t\bar{t}$ and in Fig. 7.19 for W +jet. All remaining selection cuts, in the *tight* version, including the final mass window cut, are already applied, except for b -tagging.

As a reference, the number of signal events is also shown in the plots. In the $t\bar{t}$ background the dominant contribution is given by subject combinations with $b-l$ and $b-c$ subjects faking a Higgs boson candidate.

In general, in order to pass the signal selection, the two top quarks need to be highly boosted. One of the top quarks produces in a very asymmetric way a highly boosted W boson, which is required in the signal selection, together with a low p_T b -jet, which passes through the b - and light jet vetoes. The opposite high p_T top quark produces a b -jet and a W boson which decays hadronically: out of these three jets, the combinations are selected where one jet is lost or has low p_T and the other two fake the Higgs subjects. However the combinations with both subjects originating from the W boson are highly suppressed, since their invariant mass peaks at the W boson mass, so the main remaining contribution comes from the b -jet combined with one of the two jets from the W boson ($b-l$ or $b-c$). The cases where a $b-b$ pair is selected as Higgs boson candidate in $t\bar{t}$ is rare, since in this case one b -quark needs to come from one top and the other b from the other top; so the top quarks need to have a lower p_T , which makes the simultaneous production of a highly boosted W boson very difficult.

While b -tagging can easily reduce the number of $b-l$ subject combinations to an acceptable level, it is much harder to reduce the $b-c$ component. It can be however seen (in the bottom plot of Fig. 7.18) that the explicit rejection of charm-jets by JetFitter can help in rejecting $b-c$ subject combinations, at the cost of letting more $b-l$ combinations go through.

In W +jet events the dominant contribution is given by the $l-l$ and $l-c$ subject combinations, as expected from pure QCD production. The most dangerous contribution comes however from $b\bar{b}$ pairs (e.g. from gluon splitting), which cannot be reduced by applying b -tagging, and up to a certain point also from the $l-c$ subject combinations.

In order to determine the optimal b -tagging strategy for the present analysis, the signifi-

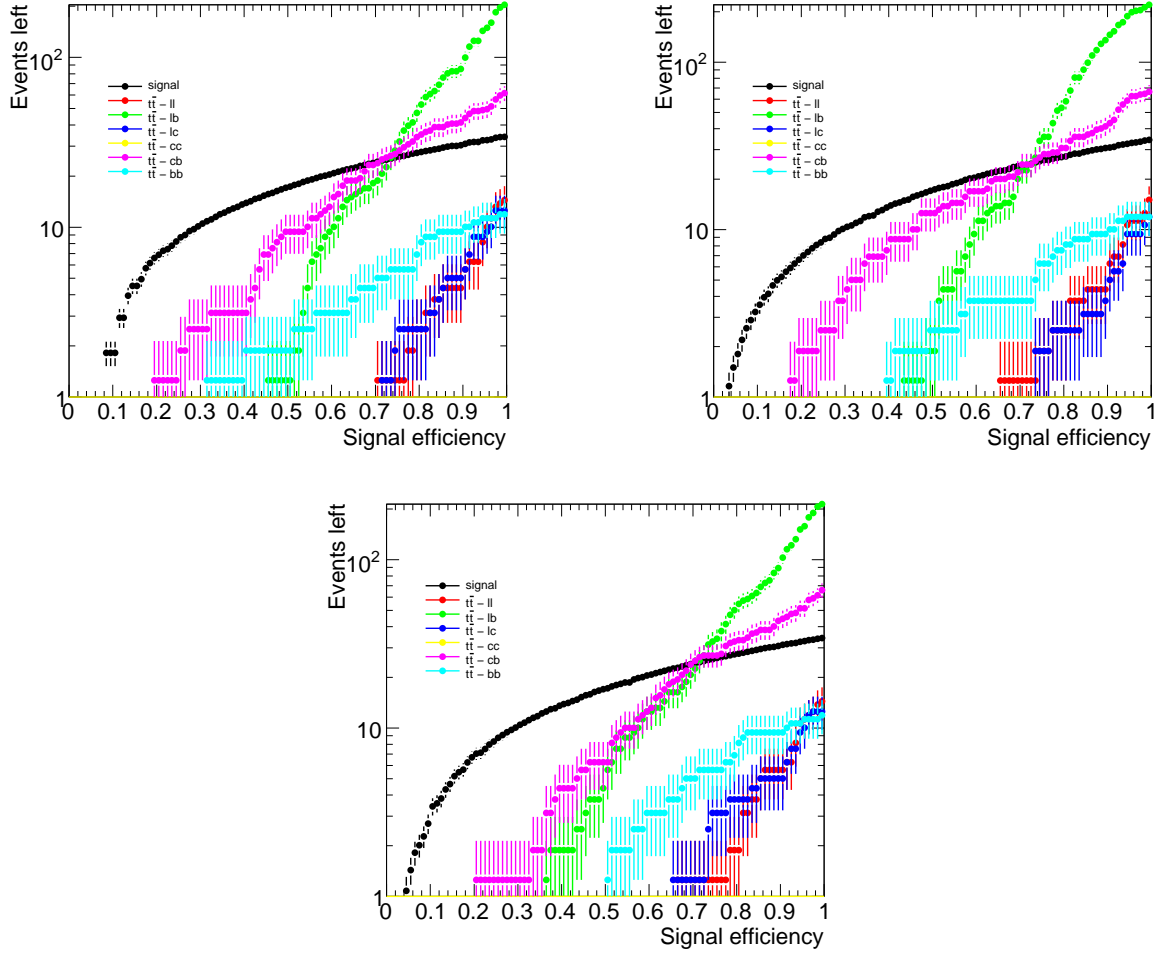


Figure 7.18: Number of events left as a function of the signal efficiency corresponding to applying a certain b -tagging cut for various subjets flavour combinations present in the $t\bar{t}$ background. Different b -tagging algorithms are used: JetFitter (top left), COMB (top right), JetFitter rejecting charm-jets (bottom).

cance, defined as $\frac{S}{\sqrt{B}}$ has been analysed as a function of the signal efficiency given for a certain b -tagging requirement. This is shown in Fig. 7.20, which corresponds to applying the *tight* selection cut flow. For comparison, also the result including the higher statistics $Wb\bar{b}$ sample is shown.

Including a veto on the presence of an eventual third b -subjet only marginally improves the statistical significance, therefore it was not included in the final analysis. The results for two different b -tagging strategies are shown: $IP3D+SV1$ (COMB) and $IP3D+JetFitter$ with a value for c_l of 0.2.

The $IP3D+SV1$ (COMB) b -tagging algorithm provides a maximum significance of around 2.6 ± 0.2 at $\approx 45\%$ signal efficiency, while JetFitter with $c_l = 0.2$ doesn't provide a clear maximum, but is relatively flat between $\approx 35 - 45\%$ signal efficiency, with a significance of $3.0 - 3.1 \pm 0.3$. In addition, with $IP3D+JetFitter$ a significantly better signal-over-background ratio can be obtained, in particular at lower signal efficiencies.

The $IP3D+JetFitter$ algorithm and a working point at 40% signal efficiency⁸ has been

⁸the quoted efficiency corresponds to around 35% $b\bar{b}$ -tagging efficiency before mass window cut

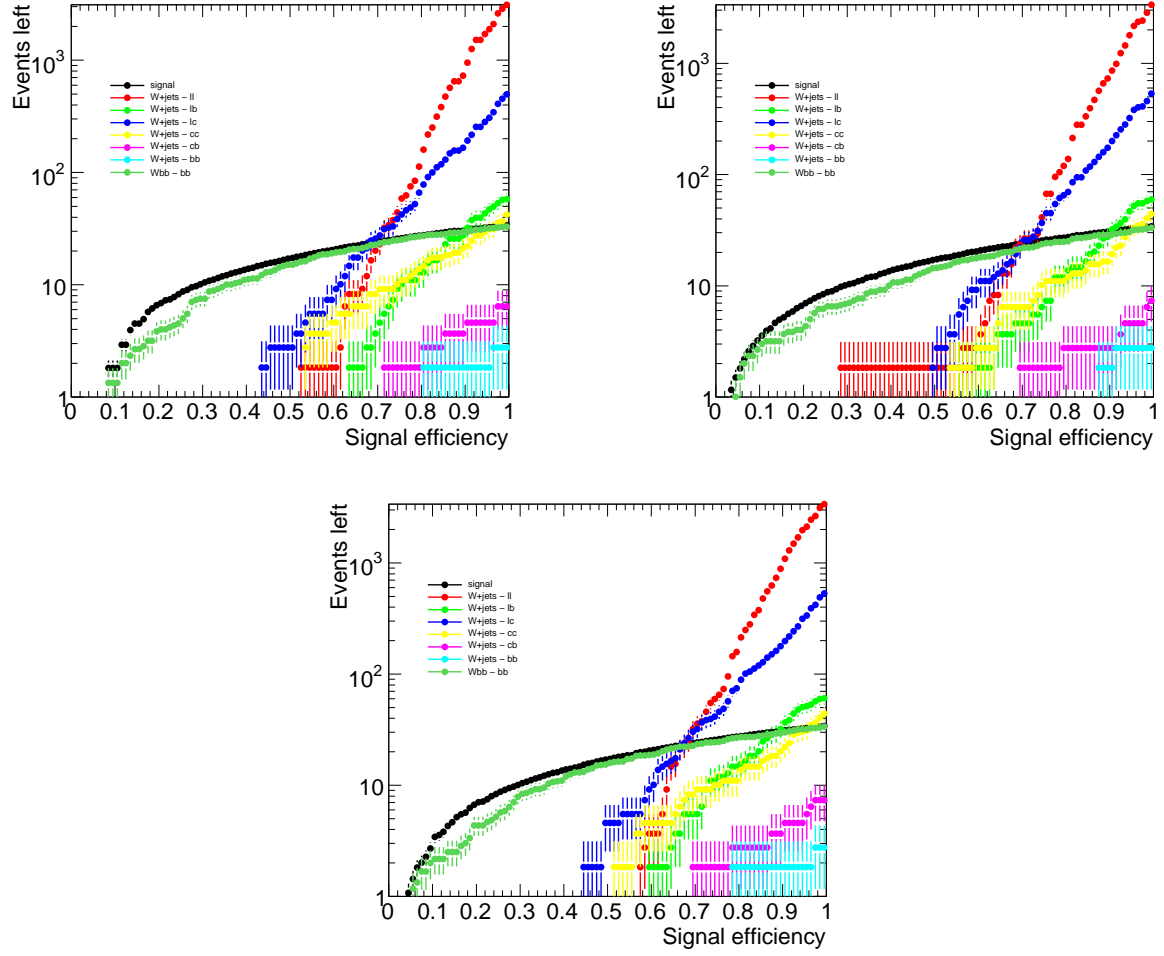


Figure 7.19: Number of events left as a function of the signal efficiency corresponding to applying a certain b -tagging cut for all various subjects flavour combinations present in the W +jet background. Different b -tagging algorithms are used: JetFitter (top left), COMB (top right), JetFitter rejecting charm-jets (bottom).

chosen for the nominal analysis selection in the *tight* version described in Section 7.5.

For the b -jet veto, a b -tagging cut based on the same discriminator, but with $c_l = 1$, is used, corresponding to a b -jet vetoing efficiency of $\approx 75\%$ and a light-jet efficiency of $\approx 2\%$.

7.8 Jet clustering: properties and optimisation

7.8.1 Comparison to traditional jet techniques

The present study is based on the reconstruction of the Higgs boson decaying into a pair of b -jets, by means of the jet clustering procedure described in Section 7.4. In this section this new jet finding method is compared to a more conventional technique based on dijets identified with the k_\perp algorithm (with effective size $R = 0.4$). In order to make the comparison easier, the analysis selection procedure needs to be slightly modified, with the following differences with respect to the default selection:

- The Higgs boson mass is based on the invariant mass of two jets with the highest b -weight, instead of the invariant mass of the highest two or (if present) three subjects of

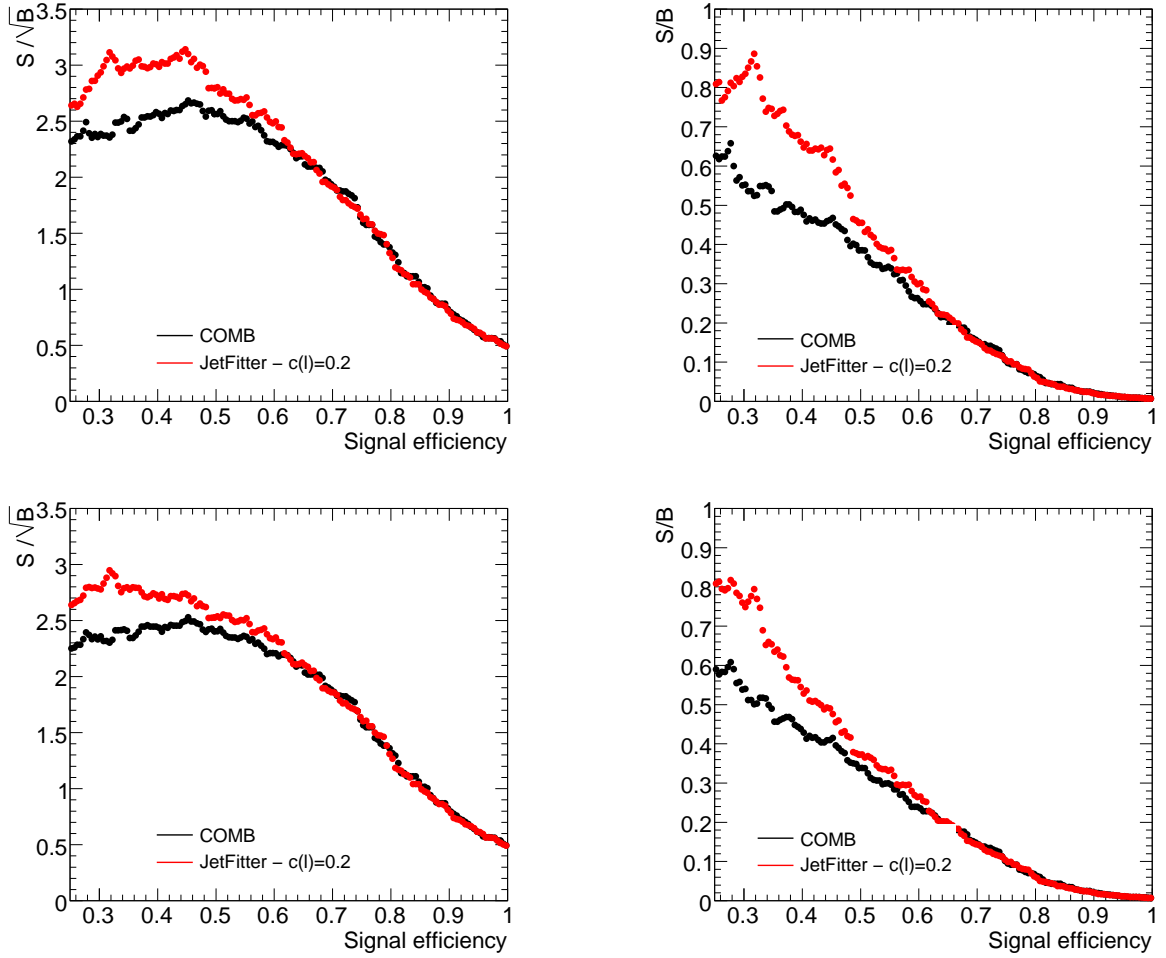


Figure 7.20: Statistical significance (left) and signal-over-background ratio (right) as a function of the signal efficiency corresponding to a certain b -tagging cut, for the reference Monte Carlo samples (top) and for the samples containing the higher statistics $Wb\bar{b}$ background component generated with AcerMC (bottom).

the Higgs boson candidate

- A shift is applied to the value of the dijet invariant mass (+6 GeV), to account for a small difference in the absolute jet energy scale calibration.

No change is applied to the remaining selection cuts.

Applying the new selection, a very large background contribution is seen, yielding a significance $\frac{S}{\sqrt{B}}$ of around 1.5, which can be explained by the fact that on conventional jets there is no limit on the ΔR of the dijet pair, while, since the subjet clustering method starts from C/A jets with $\Delta R = 1.2$, the two subjets aperture is limited to this value of ΔR . Most of the W +jet and $t\bar{t}$ background populate the ΔR distribution up to very large values, a region where only a small amount of signal events can be found.

Therefore a further set of cuts is applied on top of the pair of k_\perp jets representing the Higgs boson candidate, trying to emulate the effective selection induced by the subjet structure based jet finding:

- Di-jet symmetry $\left(\frac{\min(p_T(j_1), p_T(j_2))\Delta R(j_1, j_2)}{\text{mass}(j_1 j_2)}\right)^2 > 0.1$
- Mass drop $\frac{\max(\text{mass}(j_1), \text{mass}(j_2))}{\text{mass}(j_1 j_2)} < \frac{1}{\sqrt{3}}$
- $\Delta R(j_1, j_2) < 1.2$.

The result is shown in Table 7.7, for the signal and the main backgrounds. Since, for purely technical reasons, no specific charm-rejection was implemented for the k_\perp jets, the k_\perp jets based selection needs to be compared with a slightly modified version of the subjets based analysis of Table 7.3, presented in Table 7.8, where the b -tagging requirement does not make use of any specific charm-rejection.

The signal efficiency is 21% lower than in the mainstream analysis and at the same time the top background ($t\bar{t}$ and Wt) is increased by $\approx 20\%$. The b -tagging efficiency, even when applying the same cut value on the b -weight as in the standard analysis, turns out to be higher. However, tightening this cut does not significantly improve the overall analysis significance, which, considering also the remaining (minor) backgrounds, is ≈ 2.1 , against ≈ 2.7 of the subjets based analysis without explicit use of a dedicated charm-jet rejection.

In more detail, Fig. 7.21 shows the invariant mass distribution of the dijet pair representing the Higgs boson candidate, compared to the Higgs boson invariant mass distribution in the subjets based analysis (without dedicated charm-jet rejection).

Fig. 7.22 shows the same comparison for the distance in pseudo-rapidity between the two jets. The k_T jets require a minimum distance between the jets of $\Delta R = 0.4$: this doesn't reduce the signal contribution significantly, but avoids the W +jet contribution to populate the region with low Higgs boson invariant mass, which is very useful in providing a mass sideband to normalise this background component on data and will be exploited in the likelihood fit based analysis.

The subjet clustering method is able, for a fixed signal efficiency, to reject the background more efficiently, in particular when it acts on the $t\bar{t}$ background. A more detailed investigation and a possible explanation for this effect is presented in Section C.3.1.

In addition, in Section C.3.2, the impact of b -tagging on Higgs candidates with two very close b -quarks in $\Delta R(b, \bar{b})$, as potentially reconstructed by the subjet clustering algorithm, is analysed.

	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jet
After filter cuts	858.4 ± 6.4	3935 ± 26	1229506	93155	582364
1 Higgs candidate	848.80 ± 0.68	3832.4 ± 4.1	1221084	92386 ± 15	555598
filtered $p_T > 200$ GeV	320.9 ± 3.1	1225 ± 12	132699	16053	160325
Missing $E_T > 30$ GeV	243.8 ± 2.9	884 ± 11	106361	12789	124545
$p_T(W) > 200$ GeV	148.1 ± 2.4	586.0 ± 9.2	36891	6957 ± 44	89777
$p_T(e/\mu) > 30$ GeV	126.5 ± 2.3	479.7 ± 8.4	30139	5935 ± 41	77987
$p_T(\text{additional } \mu) < 10$ GeV	124.8 ± 2.3	454.5 ± 8.2	27647	5554 ± 40	77740
$p_T(\text{additional } e) < 10$ GeV	123.2 ± 2.3	429.2 ± 8.0	24975	5057 ± 38	76643
$\Delta\phi(W,H) > \frac{2}{3}\pi$	123.0 ± 2.3	416.4 ± 7.9	22798	4814 ± 37	75571
no additional b -jets $p_T > 15$ GeV	121.5 ± 2.2	414.4 ± 7.9	21638	4717 ± 37	75539
jets on W side $p_T < 60$ GeV	83.6 ± 1.9	298.4 ± 6.8	7235	2478 ± 27	55571
cut OnDeltaR2Jets	55.4 ± 1.6	145.7 ± 4.9	1342 ± 29	742 ± 15	12534
jets on H side $p_T < 60$ GeV	39.9 ± 1.4	122.4 ± 4.5	487 ± 17	364 ± 10	10947
one subset b -tagged	34.7 ± 1.3	30.7 ± 2.3	317 ± 14	196.3 ± 7.7	549 ± 22
both subsets b -tagged	17.06 ± 0.90	11.1 ± 1.4	54.6 ± 5.9	24.0 ± 2.7	46.0 ± 6.5
loose fit cuts	17.01 ± 0.90	11.1 ± 1.4	53.9 ± 5.8	22.5 ± 2.6	43.2 ± 6.3
$112 \text{ GeV} < \text{mass}(H) < 136 \text{ GeV}$	11.27 ± 0.73	1.01 ± 0.41	14.4 ± 3.0	3.6 ± 1.1	8.3 ± 2.8

Table 7.7: The table shows the expected events going through the selection at each stage for the signal and the main backgrounds, when using conventional k_T jets instead of the subset clustering procedure. The numbers are projected to 30 fb^{-1} of collected data.

	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jet
add. jets on W side $p_T < 20$ GeV	83.2 ± 1.9	461.3 ± 8.3	7227	3343 ± 31	86087
add. jets on H side $p_T < 20$ GeV	55.8 ± 1.6	275.6 ± 6.6	1895 ± 34	1142 ± 18	48229
one subset b -tagged	46.7 ± 1.5	53.5 ± 3.0	1039 ± 26	529 ± 13	2247 ± 45
both subsets b -tagged	20.82 ± 0.99	17.4 ± 1.7	53.9 ± 5.8	25.8 ± 2.8	104.8 ± 9.8
loose fit cuts	20.77 ± 0.99	17.0 ± 1.7	52.7 ± 5.7	24.3 ± 2.7	103.0 ± 9.7
$112 \text{ GeV} < \text{mass}(H) < 136 \text{ GeV}$	14.26 ± 0.82	1.01 ± 0.41	8.2 ± 2.3	6.7 ± 1.4	9.2 ± 2.9

Table 7.8: Number of events corresponding to the last selection steps of the subset based analysis, but without using any specific charm-jet rejection. The numbers are projected to 30 fb^{-1} of collected data, according to LO cross sections.

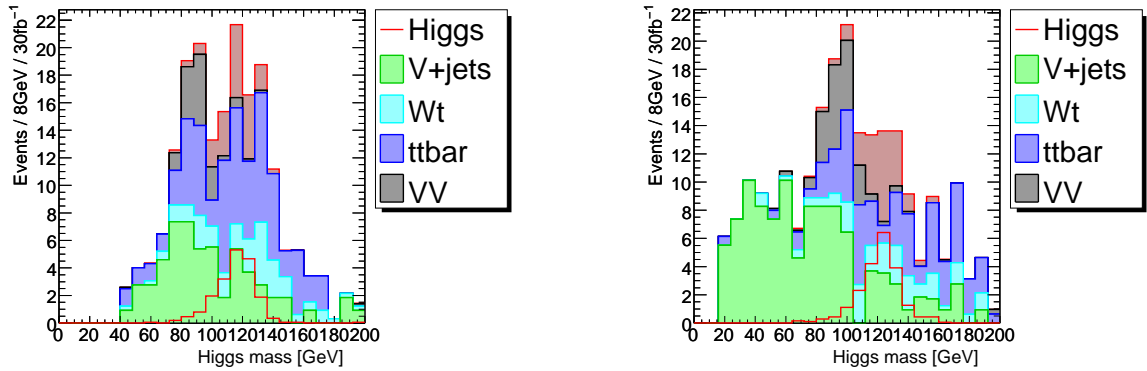


Figure 7.21: Invariant mass of the Higgs boson candidate for the conventional di-jet analysis (left) compared to the Higgs boson invariant mass obtained from the subset based analysis (right), after the full *tight* selection is applied. The numbers are projected to an integrated luminosity of 30 fb^{-1} .

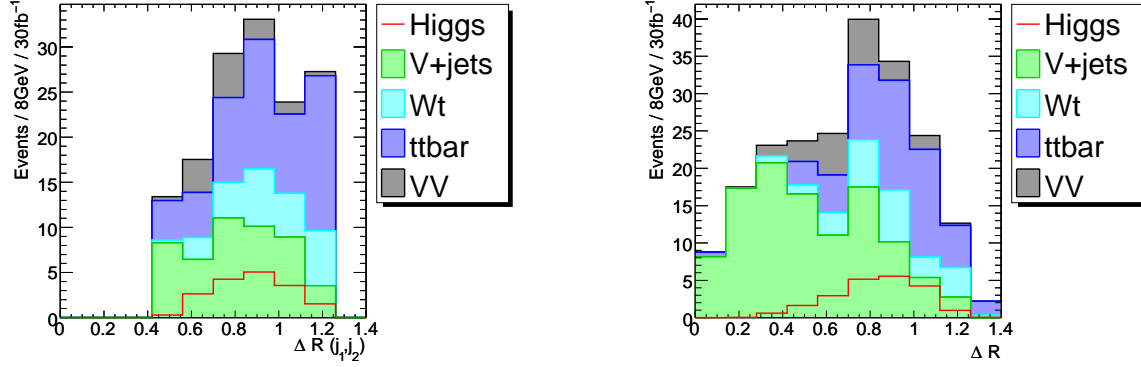


Figure 7.22: $\Delta R(j_1, j_2)$ between the two jets representing the Higgs boson candidate (left) compared to the same quantity referring to the distance in pseudo-rapidity between the two highest p_T subjects in the subjects based analysis (right), after full *tight* selection is applied. The numbers are projected to an integrated luminosity of 30 fb^{-1} .

7.8.2 Optimisation of subjet clustering algorithm

To recap, the parameters used for the subjet clustering algorithm in the present study are $\Delta R_{ij} = 1.2$, $y_{cut} = 0.1$, mass drop = $\frac{1}{\sqrt{3}}$, $p_T > 200 \text{ GeV}$ and filtering $\Delta R = 0.3$.

A value of $\Delta R_{ij} = 1.2$ is nearly optimal, because it contains most of the two body decays of a Higgs boson with a $p_T > 200 \text{ GeV}$. Values of $\Delta R_{ij} = 1.4$ and $\Delta R_{ij} = 1.0$ were also tested, but found to yield a worse performance. The mass drop value and the filtering ΔR were not optimised in the present analysis, while different values were considered for y_{cut} , which provides an efficient way of removing very asymmetric subjects configurations (typical for example for QCD soft gluon emissions). This provides a way to reduce the W +jet background, but is not particularly well suited for rejecting the $t\bar{t}$ background, where the jets are emitted by the decay of a heavy top quark.

In Table 7.9 the signal and main background components are listed for different choices of the parameter y_{cut} (0.075, 0.1 and 0.15), projected to an integrated luminosity of 30 fb^{-1} . Since the statistics is very low, the expected numbers of events without the final mass window cut are also listed. There is no dramatic change when varying the y_{cut} parameter in the chosen range, however the best significance is obtained for $y_{cut} \approx 0.1$. For this study a slightly tighter b -tagging requirement on the two leading subjects was applied with respect to the mainstream analysis.

Parameter y_{cut}	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	Wbb
$y_{cut} = 0.075$	10.79 ± 0.72	1.01 ± 0.41	2.5 ± 1.3	2.12 ± 0.80	8.0 ± 1.2
$y_{cut} = 0.1$	10.79 ± 0.72	1.01 ± 0.41	2.5 ± 1.3	2.12 ± 0.80	7.9 ± 1.1
$y_{cut} = 0.15$	9.44 ± 0.67	1.01 ± 0.41	1.9 ± 1.1	2.12 ± 0.80	7.5 ± 1.1
Parameter y_{cut}	WH(120)	WZ	$t\bar{t}(p_T^{min})$	Wt	Wbb
$y_{cut} = 0.075$	15.66 ± 0.86	12.7 ± 1.5	22.6 ± 3.8	8.5 ± 1.6	71.8 ± 3.4
$y_{cut} = 0.1$	15.42 ± 0.85	12.5 ± 1.4	21.3 ± 3.7	8.8 ± 1.6	69.6 ± 3.4
$y_{cut} = 0.15$	12.96 ± 0.78	12.3 ± 1.4	18.8 ± 3.4	8.2 ± 1.6	59.1 ± 3.1

Table 7.9: The table shows the expected number of events going through the full *tight* selection (top) and through the same selection except for the mass windows cut (bottom), for the signal and the main backgrounds and for different choices of the parameter y_{cut} . The numbers are projected to 30 fb^{-1} of collected data.

Fig. 7.23 shows a comparison between the invariant mass distributions of the same three

working points for y_{cut} . The effect of increasing the y_{cut} parameter on the WH signal is a decrease in efficiency and slight improvement in the mass resolution. This explains also why the expected number of events for the signal scales a bit differently when varying the y_{cut} parameter before and after the mass window cut.

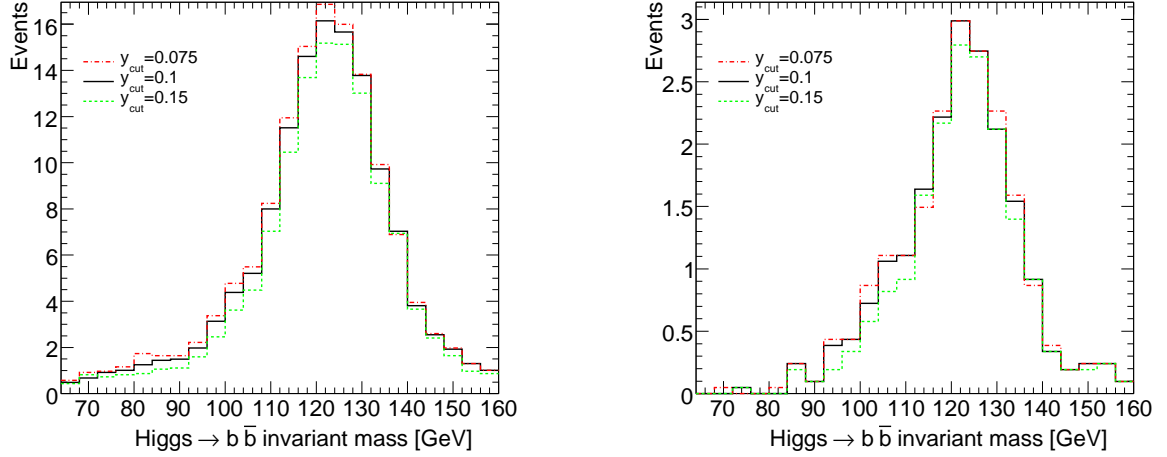


Figure 7.23: Invariant mass distribution of the Higgs boson candidate for various values of the y_{cut} clustering parameter, immediately after the additional lepton veto (left) and after the full selection (right). The numbers are projected to an integrated luminosity of 30 fb^{-1} .

7.9 Optimisation of jet energy scale and mass resolution

The filtered four momentum of the Higgs boson candidate, obtained as described in Section 7.4, can be further refined. In fact, only the cell-by-cell part of the H1-style calibration for conventional cone jets with size $R = 0.4$ was applied, without the final p_T and η dependent energy scale correction. In principle the same kind of additional correction could be derived for the Cambridge/Aachen algorithm and for the special jet clustering method used in the present analysis. In this section some methods are investigated to improve the mass resolution of the Higgs boson candidate and to correct the jet energy scale.

The first correction is specific for b -jets and comes from the fact that the b -quarks produce b -hadrons which decay semi-leptonically into a muon or electron plus neutrino with a branching ratio of around 20%. While correcting for the missing neutrino is more difficult and can be done on a statistical basis only, the muon, when present, can be accounted for by just adding it to the subjet four momentum⁹. Fig. 7.24 (top) shows the Higgs boson candidate mass distribution (after having applied the basic H1-style cell based calibration and after all analysis selection cuts), while in Fig. 7.24 (bottom) the Higgs boson candidate is corrected for the muon energy. In this way the low mass left tail is reduced and an improvement in the RMS of around ≈ 0.6 can be obtained. For computing mean values and RMS, only the mass residuals in a $(-30, 30)\%$ window around the nominal mass value are considered, to reduce the impact of outliers.

A further small improvement could be obtained by correcting the jet energy scale for the missing energy of the neutrinos from the semileptonic b -decays, but this is not studied here.

⁹The energy deposition of a muon in the calorimeter is expected to be small, in average around 2 GeV.

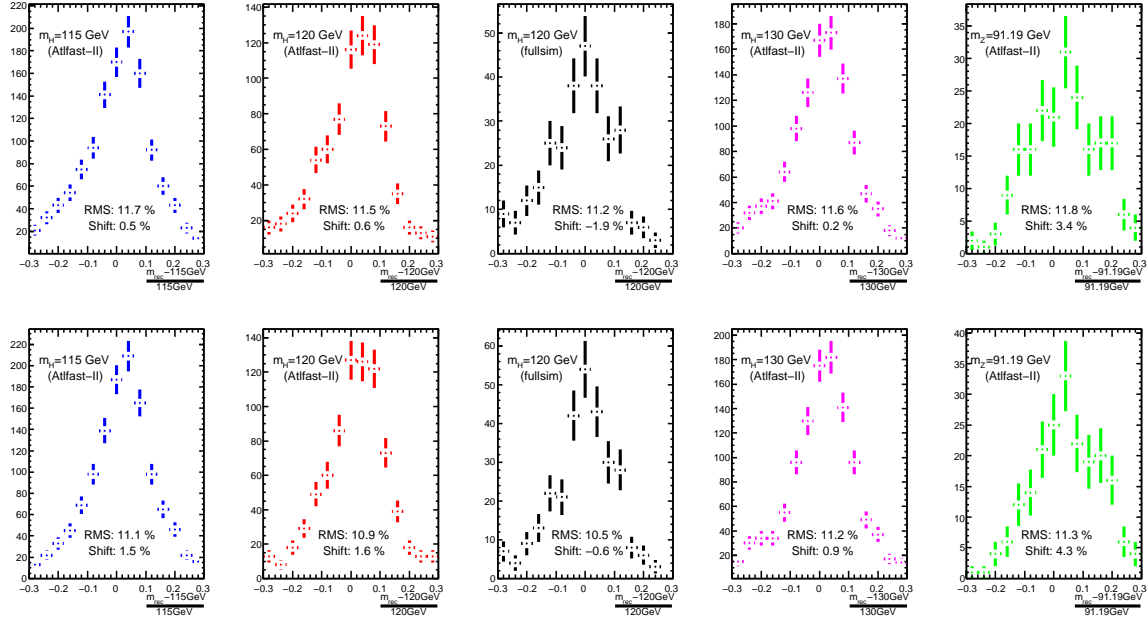


Figure 7.24: (Top) Higgs boson candidate mass distribution as delivered by jet clustering, after having applied the H1-style cell based calibration, for the $WH \rightarrow \ell\nu b\bar{b}$ process and different masses and for the $WZ \rightarrow \ell\nu b\bar{b}$ process. In the case of a 120 GeV Higgs boson mass the mass distribution on a sample produced using the complete full simulation of the ATLAS Detector is also shown for comparison. (Bottom) Same Higgs boson candidate mass distributions, but accounting also for the four momentum of a muon from a semi-leptonic b -decay, when present.

A final p_T and η dependent jet scale correction is not straightforward to perform using the standard calibration tools available in ATLAS, because the two subjects in which the Higgs boson candidate decays (or a similar boosted heavy objects) are strongly correlated: when the first subjet loses energy, this energy is most probably taken by the second subjet. The correlation in energy between the two subjects is illustrated in Fig. 7.25, which displays the average residual in p_T for a subjet with respect to the true p_T of the underlying b -quark as a function of the subjet p_T , after separating the subjects into two categories on an event basis: given the two subjects of the Higgs boson candidate, the one nearest in ΔR to the originating parton goes into one category, the remaining one into the other. The former tends to lose part of its energy, the latter acquires energy from the other jet. This illustrates the difficulty of deriving a calibration for the individual subjects; this is a more acute instance of the fact that calibrating to an initial (coloured) parton is in principle wrong, since QCD radiation implies that any such correction will be highly model dependent and measurements based on it will not in fact be physically well-defined. However, this effect does not strongly feed into the Higgs boson mass resolution, since the hardest gluon radiation from either of the b -quarks is included in the Higgs boson candidate (see Section 7.4). And since the Higgs boson itself is a colour singlet, it undergoes no QCD radiation itself, and a MC-based calibration to the Higgs boson mass itself will be less model-dependent than an attempt to calibrate to leading-order quarks or gluons.

Following the above arguments, a complete calibration strategy for a boosted heavy object reconstructed with this particular jet clustering method could be defined using high statistics Monte Carlo sample with different Higgs boson masses and trying to minimise the deviation of

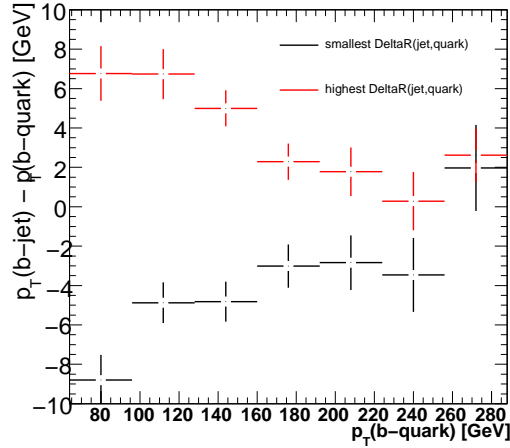


Figure 7.25: The figure shows the difference in p_T of the reconstructed subjet with respect to the b-quark p_T (before *FSR*), as a function of subjet p_T . For each Higgs boson candidate the two subjets are subdivided into two categories, in black the one nearest to the originating b-quark, in red the other.

the reconstructed mass from the Monte Carlo truth mass, applying corrections which depend on the p_T and η of either the subjets or of the whole candidate system. In order to perform the same calibration on data, the decay products of a boosted W bosons for example in $t\bar{t}$ events could be used (the b -scale correction being obtained elsewhere and applied on top of it).

Only a very basic jet scale calibration will be performed here, where a relative invariant mass scale correction is applied as a function of the transverse momentum of the Higgs boson candidate. The average relative residual of the reconstructed invariant mass of the Higgs boson candidate with respect to the true Higgs boson mass is shown in Fig. 7.26 as a function of the Higgs boson candidate transverse momentum; to increase the available statistics all the Monte Carlo samples produced with the ATLFAST-II simulation for Higgs boson masses of 115, 120 and 130 GeV are summed up together.

The dependence of the mass residual on p_T can be described by a simple function of the form:

$$f(p_T) = c_0 - \frac{c_1}{p_T^6},$$

which can be then used to correct for this effect. The fit result is shown in Fig. 7.27, first with the full selection applied (left) and then, to cross-check the validity of the functional choice, at an earlier analysis stage (just after the additional b -jet veto cut), with more statistics. It is not surprising that the b -tagging requirement affects the invariant mass distribution, since it requires the two subjets to be real b -subjets and thus removes much of the the combinatorial background present in the signal. The resulting coefficient and fit quality values are listed in Table 7.10.

After applying the p_T dependent calibration function to the signal Monte Carlo samples, one obtains the mass distributions shown in Fig. 7.28. The resolution improves slightly, the Higgs mass scale tends to be corrected, but some residual dependence on the Higgs boson candidate invariant mass remains. This indicates that a more complex calibration procedure is needed to account for the non-linearity in the reconstructed mass (at lower mass a bias is visible to smaller values, at higher mass to higher values). Given the limited statistics

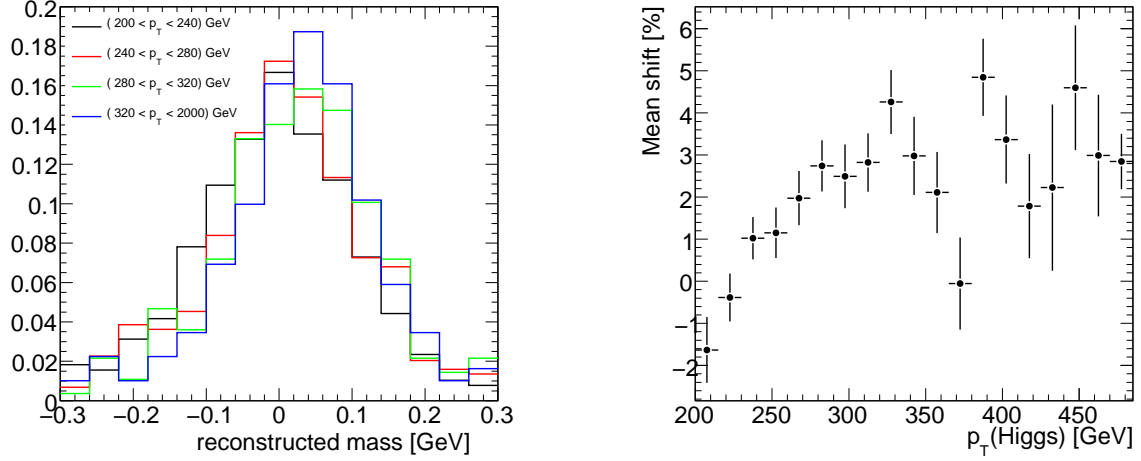


Figure 7.26: Residual of the reconstructed Higgs boson mass with respect to the true mass value in various bins of $p_T(\text{Higgs})$ (left) and average relative residual of the reconstructed Higgs boson mass as a function of $p_T(\text{Higgs})$. The *loose* analysis selection cuts have been applied.

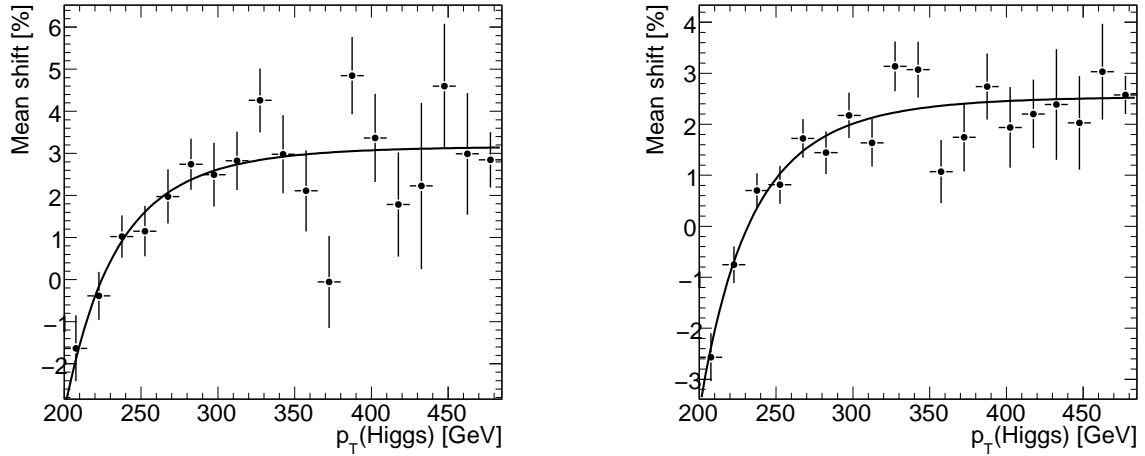


Figure 7.27: Fit of the dependence of the relative mass residual as a function of Higgs boson candidate p_T , after all selection cuts are applied (left) and just after having applied the additional b-jets veto (right). A χ^2 based minimisation procedure is used.

Parameter	value
c_0	3.17 ± 0.25
c_1	$(4.05 \pm 0.53) \times 10^{14} \text{ MeV}$
χ^2/ndf	11.31/17
$\text{Prob}(\chi^2)$	72%

Table 7.10: Result of the fit which fixes the mass scale as a function of Higgs boson candidate p_T .

available for this study, there is no way of setting up a more complex calibration procedure.

In the analysis based on pure event counting presented in Section 7.5 only the muon energy correction was taken into account, whereas the absolute energy scale and p_T (Higgs) dependent calibration is only taken into account in the fit based approach presented in Chapter 8.

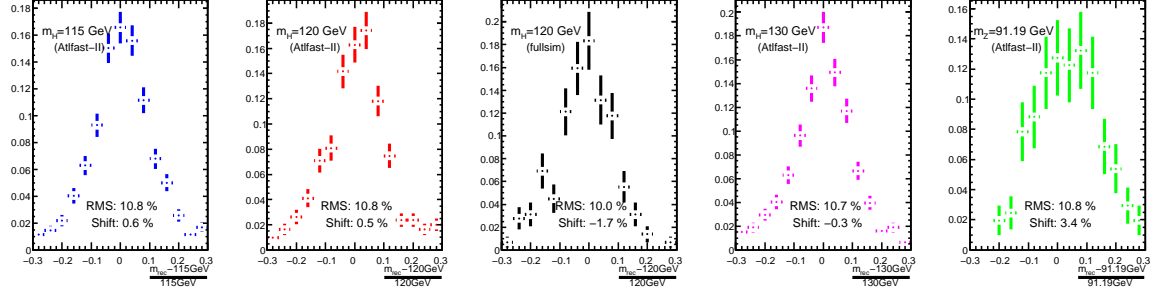


Figure 7.28: Mass distribution after application of simple p_T dependent calibration. The full simulation is shown just for comparison purposes, as well as the $Z \rightarrow b\bar{b}$ peak from the WZ sample.

7.10 Comparison with fast simulation

The present study relies on the ATLFast-II fast simulation of the ATLAS detector. As discussed in Section 7.2, the only difference between ATLFast-II and the full detector simulation based on GEANT4 is in the simulation of the calorimeter.

A general validation of the ATLFast-II fast simulation (and more specifically of the fast simulation of the calorimeter) is presented in Ref. [97]. However, the WH analysis makes use of a jet substructure based jet clustering algorithm, for which no general validation is available, and the final state objects are characterised by very specific kinematic configurations. It is therefore useful to compare the ATLFast-II and complete GEANT4 full detector simulations on a set of signal events, in order to confirm that ATLFast-II provides a realistic simulation of the detector response in this specific scenario and, in case of small discrepancies, to understand them and obtain correction factors. These can be then used in a later step of the analysis, either by applying them on top of the ATLFast-II simulated events or by considering them as a source of systematic uncertainty.

A comparison between the lepton reconstruction efficiency in ATLFast-II and full simulation is presented in Section C.4.1, with the following results:

- The efficiency and energy resolution of muons is well reproduced in ATLFast-II, as expected;
- The electron efficiency in ATLFast-II is overestimated by $\approx 2\text{-}4\%$ (for medium electrons by $3.1 \pm 1.4\%$);
- The electron p_T is underestimated by $\approx 2\%$.

While no correction is therefore necessary for the muons, the differences in the electron efficiency and energy scale is corrected for in the fit based analysis presented in Chapter 8.

In addition, a detailed comparison of the signal acceptance between ATLFast-II and full simulation, corresponding to the selection criteria used for the WH analysis, is presented in Section C.4.2, together with a more detailed comparison of some of the most discriminating variables used in the analysis. The main results of this comparison are:

- The $b\bar{b}$ mass distribution of the Higgs candidate is fairly well reproduced in ATLFast-II: the resolution is comparable, while a small shift is seen due to the slightly different jet energy scale calibration in ATLFast-II and full simulation (this is expected to have no effect on the analysis);
- The hadronic reconstruction efficiency of the $H \rightarrow b\bar{b}$ system in ATLFast-II is overestimated by $5 \pm 1\%$.
- The overall difference in the signal acceptance between ATLFast-II and full simulation is compatible with the differences in the reconstruction efficiency of the $b\bar{b}$ system and of electrons from the W boson decay.

The difference in the hadronic reconstruction efficiency is again taken into account in the fit based analysis presented in the next chapter, by applying a correction factor on the expected signal reconstruction efficiency.

7.11 Combination with the ZH channels

In Ref. [98] the Higgs boson search based on the $\ell\nu b\bar{b}$ signature presented in this chapter was combined with other two search channels, based on the $\ell\ell b\bar{b}$ and $\nu\nu b\bar{b}$ final states, which select mainly events produced via the $ZH \rightarrow \ell\ell b\bar{b}$ and $ZH \rightarrow \nu\nu b\bar{b}$ processes, respectively.

The final distributions for the $b\bar{b}$ invariant mass of the Higgs candidate for the two $\ell\ell b\bar{b}$ and $\nu\nu b\bar{b}$ search channels, after the respective analysis selections are applied and for an integrated luminosity of 30 fb^{-1} , are shown in Fig. 7.29. For both channels the statistical significance was evaluated by counting the number of signal and background events in the 104-136 GeV mass region. The combined statistical significance of all three channels, obtained by simply adding the significances in quadrature is 3.7, based on an integrated luminosity of 30 fb^{-1} .

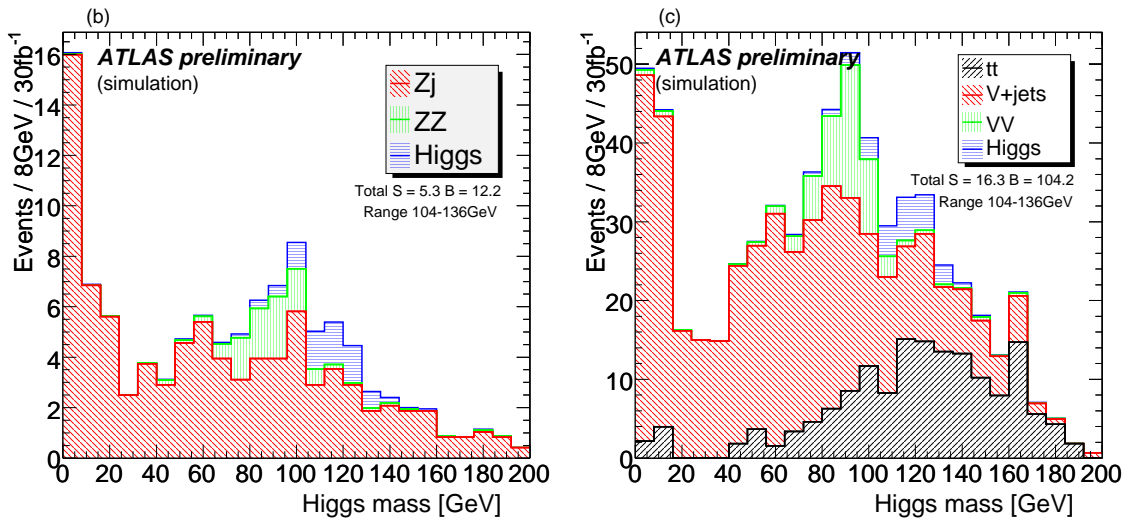


Figure 7.29: Distribution of the invariant mass of the Higgs boson candidate after analysis selection for the $\ell\ell b\bar{b}$ (left) and the $\nu\nu b\bar{b}$ (right) channels. The two distributions are normalised to an integrated luminosity of 30 fb^{-1} , based on LO cross sections. Taken from Ref. [98].

In addition, the influence of the systematic uncertainty on the background levels on the discovery sensitivity was considered. In order to get a rough idea of the influence of correlations of the background levels between different channels, three classes of backgrounds were defined:

- (t): $t\bar{t}$, WW , WZ and ZZ ;
- (w): $W + jets$ and Wt ;
- (z): $Z + jets$,

and the expected number of background events of the same class for different channels were assumed to be completely correlated, while the background levels were assumed to be completely uncorrelated between the different background classes.

Table 7.11 shows the estimated median discovery sensitivity as a function of possible values of the uncertainties σ_t , σ_w and σ_z . Assuming that the different background classes can be controlled with a precision of the order of 10-15%, a median combined discovery sensitivity of 3.0-3.2 σ can be achieved.

σ_t	σ_w	σ_z	Significance
Perfect	Perfect	Perfect	3.7
5%	5%	5%	3.5
10%	10%	10%	3.2
15%	15%	15%	3.0
20%	20%	20%	2.8
30%	30%	30%	2.5
50%	50%	50%	2.2
75%	75%	50%	2.0
50%	10%	10%	2.8

Table 7.11: Significances for different scenarios with differing background uncertainties. A Higgs boson mass of $m_H = 120$ GeV and an integrated luminosity of 30 fb^{-1} is assumed. Taken from Ref. [98].

While this constitutes a first attempt to get a reasonable estimate of the combined discovery significance of the VH ($V = W, Z$) channels, a more detailed study is, however, needed, in order to define a strategy which makes an estimate of the background levels with a precision of 10-15 %, or better, possible. In addition, the correlations between the various backgrounds when performing this estimate need to be more realistically modelled. This is the main aim of the shape-based likelihood fit analysis presented in the next chapter for the $WH \rightarrow \ell\nu b\bar{b}$ search channel. An extension to the other two channels will be needed in order to get also a more realistic estimate of the combined discovery sensitivity.

8 Fit based approach and systematic uncertainties

8.1 Impact of systematic uncertainties

A first evaluation of the discovery sensitivity of the $WH \rightarrow \ell\nu b\bar{b}$ analysis to a Standard Model Higgs boson with a mass around 120 GeV was presented in Section 7.5, based on a first evaluation of the significance based on naïve event-counting. The sensitivity of the analysis can be condensed in the statistical significance $\frac{S}{\sqrt{B}}$.

However, this significance can only be realised under the hypothesis that the expected number of background events is known very precisely. This uncertainty on B , which will be denoted ΔB in the following, can be very high, due to the high uncertainty in the theoretical prediction of the differential cross section of the backgrounds in the very specific selected region of phase space and due to the uncertainty on the background acceptance as predicted by Monte Carlo simulations. In order to understand how this uncertainty propagates into the sensitivity of this channel to the discovery of a Higgs boson with a mass of 120 GeV, the statistical uncertainty on the background \sqrt{B} is summed up in quadrature with the systematic uncertainty ΔB , yielding a corrected significance of the form $\frac{S}{\sqrt{B+\Delta B^2}}$, as shown in Fig.8.1 as a function of $\frac{\Delta B}{B}$.

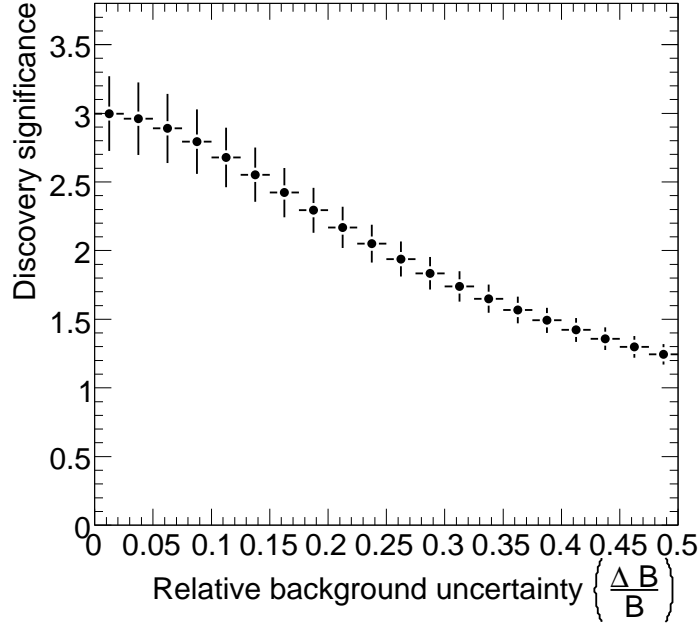


Figure 8.1: Discovery significance for a 120 GeV Higgs boson as a function of the background uncertainty $\frac{\Delta B}{B}$ for 30 fb^{-1} of collected data. The error accounts for the limited available Monte Carlo statistics. No further systematic uncertainties are considered here.

For example, if a relative background uncertainty of 25% is considered (which is not too unrealistic), the discovery significance drops from ≈ 3.0 to ≈ 2.0 . This drop is limited by the relative high purity of the selected signal ($\frac{S}{B} \approx \frac{2}{3}$)¹, but it is still very significant.

In order to avoid such a drop in the discovery sensitivity, a more sophisticated approach, which permits to extract the number of signal and background events from the data, can be adopted, relying on a maximum likelihood fit based on the different shapes expected for few discriminating variables in signal and background.

8.2 Maximum likelihood formalism

The events passing the basic *loose* analysis preselection presented in Section 7.5 are used as input to an (unbinned) extended maximum likelihood fit, which extracts the number of events (*yields*) for the signal and for the different background components, relying on probability density functions (PDFs) for the following discriminating variables:

- Higgs boson candidate invariant mass (*filtered* mass);
- p_T of the hardest additional jet in the event;
- b -weight (b -tagging discriminating variable);
- $\Delta\eta(W,H)$, the distance in pseudo-rapidity between the H and the W bosons candidates.

The b -weight is based on the b -tagging discriminator for the two subjects from the Higgs boson candidate, defined as:

$$w = \log \left(\frac{P(b)}{P(b) + c_l P(l) + (1 - c_l) P(c)} \right), \quad (8.1)$$

where $P(b)$, $P(c)$ and $P(l)$ are the a-posteriori probabilities that respectively a b -, c - or light jet has been selected. A value of $c_l = 0.2$ is used. The b -weight which is fed into the likelihood fit is the smallest b -weight of the two subjects.

Since for the Higgs boson candidate a large dependence of the invariant mass distribution on the Higgs boson transverse momentum is seen in $t\bar{t}$ and Wt events, the invariant mass distribution is described using a PDF which is made conditionally dependent on the transverse momentum of the Higgs boson candidate. The conditional dependence requires the integral over the whole mass range considered to be one regardless of the value of the Higgs boson transverse momentum. The choice of introducing this dependence is motivated by the fact that for the $t\bar{t}$ background the number of events selected for example in a mass window of $(112 < m_H < 136)$ GeV decreases more rapidly as a function of the Higgs boson candidate transverse momentum than for the signal, as shown in Fig. 8.2, so that this behaviour can be exploited to increase the sensitivity of the fit to the presence of a Higgs boson signal.

¹To understand how dramatic the impact of the signal-over-background ratio is, the reader is recommended to compare Fig. 8.1 with the analogous Fig. 7.2 of Chapter 7, which refers to the inclusive (low p_T) WH analysis.

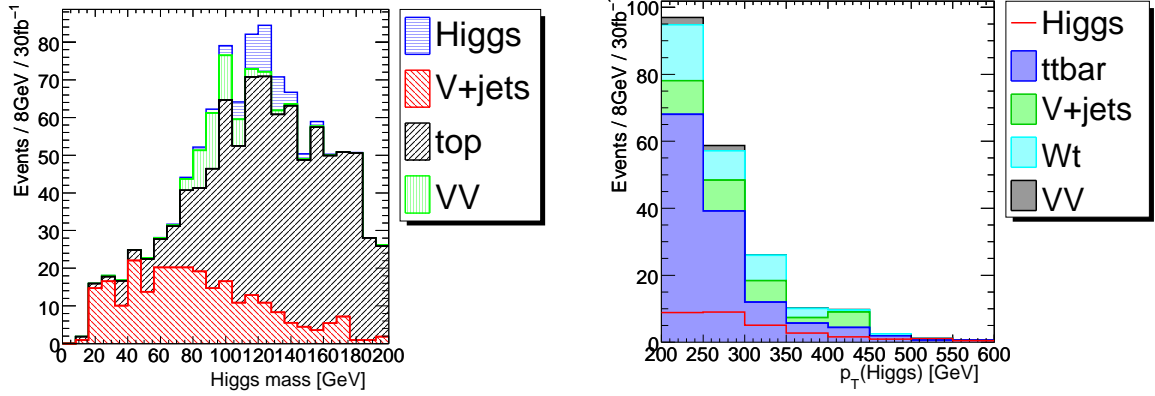


Figure 8.2: (Left) Distribution of the Higgs boson candidate invariant mass, after application of the *loose* analysis preselection cuts. (Right) Distribution of the Higgs boson candidate transverse momentum for all events inside a mass window region of 112-136 GeV, where most of the Higgs boson signal events are expected, after *loose* analysis preselection cuts. Both distributions are normalised to 30 fb^{-1} of collected data. Labels: V+jets includes both W +jets and Z +jets; VV includes ZZ, WZ and WW events and top includes both $t\bar{t}$ and Wt .

The overall likelihood function depends directly on the signal and background yields and, through the PDFs, on the values of the discriminating variables in each event of the data sample on which the fit is applied. Its explicit functional form is:

$$L = \frac{e^{-(N_{WH}+N_{WZ}+N_{t\bar{t}}+N_{Wt}+N_{W+jet})}}{N_{tot}!} \prod_{i=0}^{N_{tot}} \{ \quad (8.2)$$

$$N_{WH}P_{WH}(m_H, i)P_{WH}(p_{T,H}, i)P_{WH}(p_T(\text{add. jet}, i))P_{WH}(b \text{ weight}, i)P_{WH}(\Delta\eta(W, H), i) +$$

$$N_{WZ}P_{WZ}(m_H, i)P_{WZ}(p_{T,H}, i)P_{WZ}(p_T(\text{add. jet}, i))P_{WZ}(b \text{ weight}, i)P_{WZ}(\Delta\eta(W, H), i) +$$

$$N_{t\bar{t}}P_{t\bar{t}}(m_H, i|p_{T,H}, i)P_{t\bar{t}}(p_{T,H}, i)P_{t\bar{t}}(p_T(\text{add. jet}, i))P_{t\bar{t}}(b \text{ weight}(i))P_{t\bar{t}}(\Delta\eta(W, H), i) +$$

$$N_{Wt}P_{Wt}(m_H, i|p_{T,H}, i)P_{Wt}(p_{T,H}, i)P_{Wt}(p_T(\text{add. jet}, i))P_{Wt}(b \text{ weight}, i)P_{Wt}(\Delta\eta(W, H), i) +$$

$$N_{W+jet}P_{W+jet}(m_H, i|p_{T,H}, i)P_{W+jet}(p_{T,H}, i)P_{W+jet}(p_T(\text{add. jet}, i))$$

$$P_{W+jet}(b \text{ weight}, i) + P_{W+jet}(\Delta\eta(W, H), i) \},$$

where:

- the index i runs over all events N_{tot} in the data sample;
- $P_{comp}(var)$ represents the probability density function (PDF) of variable var for component $comp$;
- $N_{WH, WZ, t\bar{t}, Wt, W+jet}$ represent the number of events (yield) for the signal and the various background components.

Further minor backgrounds are included in each of the WZ , $t\bar{t}$, Wt and W +jet background components: WZ also includes the small ZZ and WW components, while W +jet also contains Z +jet. In addition the signal component also includes the small signal contribution coming from ZH , where typically the Z boson decays leptonically, but one of the two leptons is not reconstructed and provides a large E_T^{miss} signature.

8.2.1 Definition of signal and background PDFs

The PDFs used in the fit for the various signal and background components are expressed as a function of the discriminating variables and of few additional parameters. The variation of these parameters make it possible to adapt the shapes of the PDFs. The use of fully parametrised PDFs has several advantages with respect to the use of template PDFs with a fixed shape (e.g. simple smoothed histograms):

- The effect of statistical and systematic uncertainties can be easily handled as variations of the parameters which define the PDFs.
- Whenever the shape of a PDF can be determined on data, the parameters for this PDF can be determined directly in the fit, with the advantage that the resulting uncertainty is intrinsically considered in the fit result.

The functional form chosen for the PDFs, in particular with respect to the number of parameters, is necessarily a compromise between the maximum flexibility in being able to describe any systematic variation of the PDFs and the need to keep the number of degrees of freedom of the model low. A high number of degrees of freedom in fact will increase the amount of data needed to determine the values of the parameters with comparable precision. In some cases, as for the invariant mass of the Higgs boson candidate in the signal, the parameters have a direct physical meaning, as for example the core Gaussian resolution σ . In other cases the functional form and the parametrisation for the PDFs were found heuristically by looking at the distributions and trying to find a suitable functional form.

The filtered mass distribution for both signal and WZ background is parametrised by the use of the following PDF:

$$f(x; m, \sigma_{\pm}, \alpha_{\pm}) = \exp \left[-\frac{(x - m)^2}{2\sigma_{\pm}^2 + \alpha_{\pm}(x - m)^2} \right], \quad (8.3)$$

which represents a bifurcated Gaussian function, with asymmetric left and right Gaussian widths σ_{\pm} and, in addition, with asymmetric left and right non-Gaussian tails described by α_{\pm} . The mass distribution for the $t\bar{t}$ and Wt background is parametrised by the use of a PDF conditionally depending on $p_T(\text{Higgs})$ based on the sum of two bifurcated Gaussian:

$$f(x; m_1, m_2, \sigma_1, \sigma_2, f_1) = (1 - f_1) \cdot \exp \left[-\frac{(x - m_1)^2}{2\sigma_1^2} \right] + f_1 \cdot \exp \left[-\frac{(x - m_2)^2}{2\sigma_2^2} \right], \quad (8.4)$$

where the parameters m_1 , m_2 and f_1 are defined as a function of $p_T(\text{Higgs})$:

$$m_1 = m_{1,0} + m_{1,1} \cdot (p_T - 250 \text{ GeV}) + m_{1,2} \cdot (p_T - 250 \text{ GeV})^2 \quad (8.5)$$

$$m_2 = m_{2,0} \frac{m_1}{m_{1,0}} \quad (8.6)$$

$$f_1 = f_{1,0} + f_{1,1} \cdot (p_T - 200 \text{ GeV}) + f_{1,2} \cdot (p_T - 200 \text{ GeV})^2 \quad (8.7)$$

The form of the dependence on $p_T(\text{Higgs})$ is chosen in such a way that the two peaks of the bifurcated Gaussian move to higher values with increasing $p_T(\text{Higgs})$ values, where this dependence is expressed through a quadratic relation, while the relative fraction of the two bifurcated Gaussians changes also as a quadratic function of $p_T(\text{Higgs})$. The choice of this parametrisation is justified by the fact that two approximately Gaussian peaks are observed in the $t\bar{t}$ and Wt background, the first related to all Higgs boson candidates where one of the three jets originating from the top quark decay does not enter the Higgs boson candidate

selection, the second related to all cases where all three jets originating from the top quark decay enter the selection and are thus considered for the computation of the Higgs boson candidate mass. At higher Higgs boson candidate transverse momenta, the top quark p_T is higher and the decay products are more collimated: this justifies the fact that the fraction of events in the second peak is larger with higher $p_T(\text{Higgs})$.

The W +jet background is parametrised through the use of a Gaussian function modified by a second order polynomial:

$$f(x; x_0, p_2, \sigma) = (x + p_2 \cdot x^2) \exp \left[-\frac{x^2}{2\sigma^2} \right], \quad (8.8)$$

where the σ parameter is also made $p_T(\text{Higgs})$ dependent:

$$\sigma = p_0 + p_1 \cdot (p_T - 250 \text{ GeV}) \quad (8.9)$$

For the signal and the WZ background the shape for the filtered mass distribution does not depend significantly on the Higgs boson transverse momentum, once the correction worked out in Section 7.9 is applied, so no explicit dependence of the mass PDF on $p_T(\text{Higgs})$ needs to be considered.

The $p_T(\text{Higgs})$ distribution is parametrised in terms of a simple step function with only four bins. This PDF is not intended to increase the discrimination power of the fit, because of the theoretical uncertainties connected with its underlying distribution and because it provides a negligible separation power between the signal and background components. In principle, it is possible to setup a likelihood function which depends only conditionally on this variable and the normalisation over $p_T(\text{Higgs})$ can be handled by making explicit use of the values of $p_T(\text{Higgs})$ of the data sample on which the fit is applied. However, this approach relies implicitly on the assumption that the $p_T(\text{Higgs})$ distribution is the same for all signal and backgrounds, which does not hold in this case (Fig. 8.4), and leads to a significant bias on the expected number of signal events². In order to solve this problem, the PDF for the $p_T(\text{Higgs})$ distribution has been introduced explicitly into the likelihood function, however its discrimination power has been limited by adopting four only bins and adding in the fit an additional systematic uncertainty on the content of each bin of 15%³.

The transverse momentum of the additional jets in the event is also parametrised in terms of a simple step function between 0 and 60 GeV. While for transverse momenta above 30 GeV the signal purity is small, extending the fit to higher values is very useful in order to increase the region where the $t\bar{t}$ background is dominant.

The b -weight is also parametrised in terms of a simple step function, with 5 bins, where the intervals were chosen in such a way to possibly increase the discrimination power, while keeping the number of degrees of freedom ($n_{bins} - 1$) low.

In addition to the variables already described, a small additional discriminating power is provided by the variable $\Delta\eta(W, H)$, which is parametrised through the use of a simple Gauss function.

All PDFs need to be normalised in such a way that their integral over the full range defined for the discriminating variable they depend on is one:

$$\hat{f}(var, \vec{\alpha}) = \frac{f(var, \vec{\alpha})}{\int f(var, \vec{\alpha}) d(var)}, \quad (8.10)$$

²This effect is studied more in general with toy models in Ref. [99].

³The influence of the exact value used for this uncertainty onto the fit is found to be small, however the precise value to be used once data will be available will need to be determined by a dedicated study on the influence of theoretical uncertainties on this distribution, for example by varying the renormalisation and factorisation scale used during event generation, in addition to the experimental uncertainties considered later in this study.

where $\vec{\alpha}$ are the additional parameters the PDF depends on. The PDFs defined for the $t\bar{t}$, WZ and the W +jet backgrounds contain a conditional dependence on $p_T(\text{Higgs})$. Their normalisation needs to be computed in a slightly different way:

$$\hat{f}(m(H), \vec{\alpha}|p_T(H)) = \frac{f(m(H), \vec{\alpha}|p_T(H))}{\int f(m(H), \vec{\alpha}|p_T(H)) dm}. \quad (8.11)$$

Even if the second normalisation integral may appear similar to the first at first glance, it contains an important additional dependence on $p_T(H)$. The first normalisation integral depends only on the PDF parameters $\vec{\alpha}$, which change at most once per fit iteration (in case they are left free to be determined in the fit), so it needs to be computed at most once per fit iteration. The second normalisation integral changes as a function of the value of $p_T(H)$ of the event of the data sample the PDF is being evaluated on: since an unbinned fit is used, this normalisation integral needs to be evaluated a number of times equivalent to the number of events in the data sample (≈ 2000 for 30fb^{-1} of integrated luminosity), for every fit iteration. The repeated evaluation of these integrals by using numerical integration is extremely slow. Therefore all conditionally dependent PDFs were first integrated analytically or computed in terms of error function, and the result implemented in the definition of the PDFs. This increased the speed of the fit by more than an order of magnitude.

8.2.2 Determination of PDF shapes using Monte Carlo events

The simplest way to determine the parameters of the PDFs and therefore their exact shape is to run several fits where the likelihood function defined in Eq. 8.2 is maximised with data samples defined as each of the single separate signal and background Monte Carlo samples, with all events passing the *loose* analysis preselection. While performing the fit, the yields of the remaining signal or background components are fixed to 0 and only the parameters of the PDFs involving the signal and background component under consideration are determined in the fit. This operation permits to obtain the shapes of the PDFs from Monte Carlo.

In addition to the uncorrected ATLFAST-II Monte Carlo samples considered in this section, which represent the *perfect* scenario, in Section 8.6 an additional set of Monte Carlo samples will be considered, which represent the *nominal* scenario, where some additional corrections are applied in order to obtain a more realistic estimate of the detector performance. While only the *perfect* scenario is discussed here, the figures presented in the following show also the PDF shapes for the *nominal* scenario.

The shapes obtained for the PDFs are compared with the actual distributions of the discriminating variables after *loose* analysis preselection cuts in the single signal or background Monte Carlo samples, as a first cross-check that the assumed functional forms describe the underlying distributions reasonably well.

The shapes of the mass and b -weight PDFs for the $t\bar{t}$ and Wt backgrounds are determined on a specific control sample derived from the $t\bar{t}$ background, which relies on an alternative analysis selection and which will be described in detail in Section 8.3: in these specific cases the distributions of the discriminating variables after *loose* analysis preselection cuts are therefore compared with the PDFs as determined on the $t\bar{t}$ control sample, which represents a statistically independent set of events. However, a comparison between these PDFs and their underlying distributions in the control sample itself is also shown separately, to permit an easier cross-check of the assumed functional form.

In Fig. 8.3 the projections of the likelihood function on the Higgs boson invariant mass variable are shown for the different signal and background components, together with the distributions of the respective Monte Carlo samples. Such a projection, which consists in

integrating out the dependence on the Higgs boson transverse momentum and the likelihood dependence on all remaining discriminant variables, defines the Higgs mass PDF.

Fig. 8.4 illustrates the PDFs for the transverse momentum of the Higgs boson candidate, for all signal and background components.

The PDFs for $p_T(H)$ as determined on the Monte Carlo samples are shown in Fig. 8.4.

The PDFs based on the p_T of the additional jets as discriminating variable are shown in Fig. 8.5, again for all signal and background components, together with their related distribution in Monte Carlo.

In order to increase the b -tagging discriminating power, in particular against charm-quark jets, the b -weight is introduced explicitly in the fit. The PDFs as determined on the Monte Carlo samples are shown in Fig. 8.6.

Finally the PDFs for the $\Delta\eta(W,H)$ variable are shown in Fig. 8.7.

The conditional dependence of the mass distribution of the Higgs boson candidate on its transverse momentum is illustrated for the $t\bar{t}$ background component in Fig. 8.8 and for the W +jet background in Fig. 8.9. For the $t\bar{t}$ background, since the number of events in the control sample is higher than in the sample of events after signal preselection cuts, the $t\bar{t}$ control sample was used here⁴. It is particularly important that the dependence on $p_T(\text{Higgs})$ is correctly modelled: while this seems really to be the case for the $t\bar{t}$ background, for the W +jet background a higher statistics Monte Carlo sample to cross-check the chosen functional form would be highly desirable.

Since the Monte Carlo statistics is limited, the PDFs determined in this way, even under the hypothesis of a good choice of their functional forms, cannot be considered as a perfect representation of the underlying *true* distributions. The statistical uncertainty connected to the limited statistics of the Monte Carlo samples on which the PDFs shapes are determined will be explicitly considered together with the remaining systematic deformations of the PDFs. It is however worth noticing that this uncertainty can be reduced by increasing the amount of produced Monte Carlo events.

8.2.3 Correlation between fit variables

Apart from the filtered mass and the transverse momentum of the Higgs boson candidate, all other variables are assumed to be uncorrelated in the fit. This is reflected in the likelihood function (Eq. 8.2), where the one-dimensional PDFs of the single signal or background components based on all remaining discriminating variables are just multiplied with each other. In order to avoid any bias in the fit result, this assumption needs to be valid for the signal and all background components.

A first simple cross-check of these correlations is provided by considering the linear correlation factors between all fit variables and is illustrated in the Tables 8.1-8.5, separately for all signal and background components.

The most sizable correlations happen to occur in the $t\bar{t}$ and W +jet background between the filtered mass and the $p_T(\text{Higgs})$ variables, but this correlation is correctly accounted for in the likelihood function. The remaining not modelled correlations are mostly at the few percent level, in all remaining cases smaller than $\approx 12\%$. The effect of these residual correlations is expected to be small.

Since the assumption that the variables used in the fit are uncorrelated is essential for the present study, in Section D.1 a more detailed study of these correlations is presented, which

⁴The equivalence of the dependence of the Higgs mass on $p_T(\text{Higgs})$ between signal and control region will be demonstrated in Section 8.3.

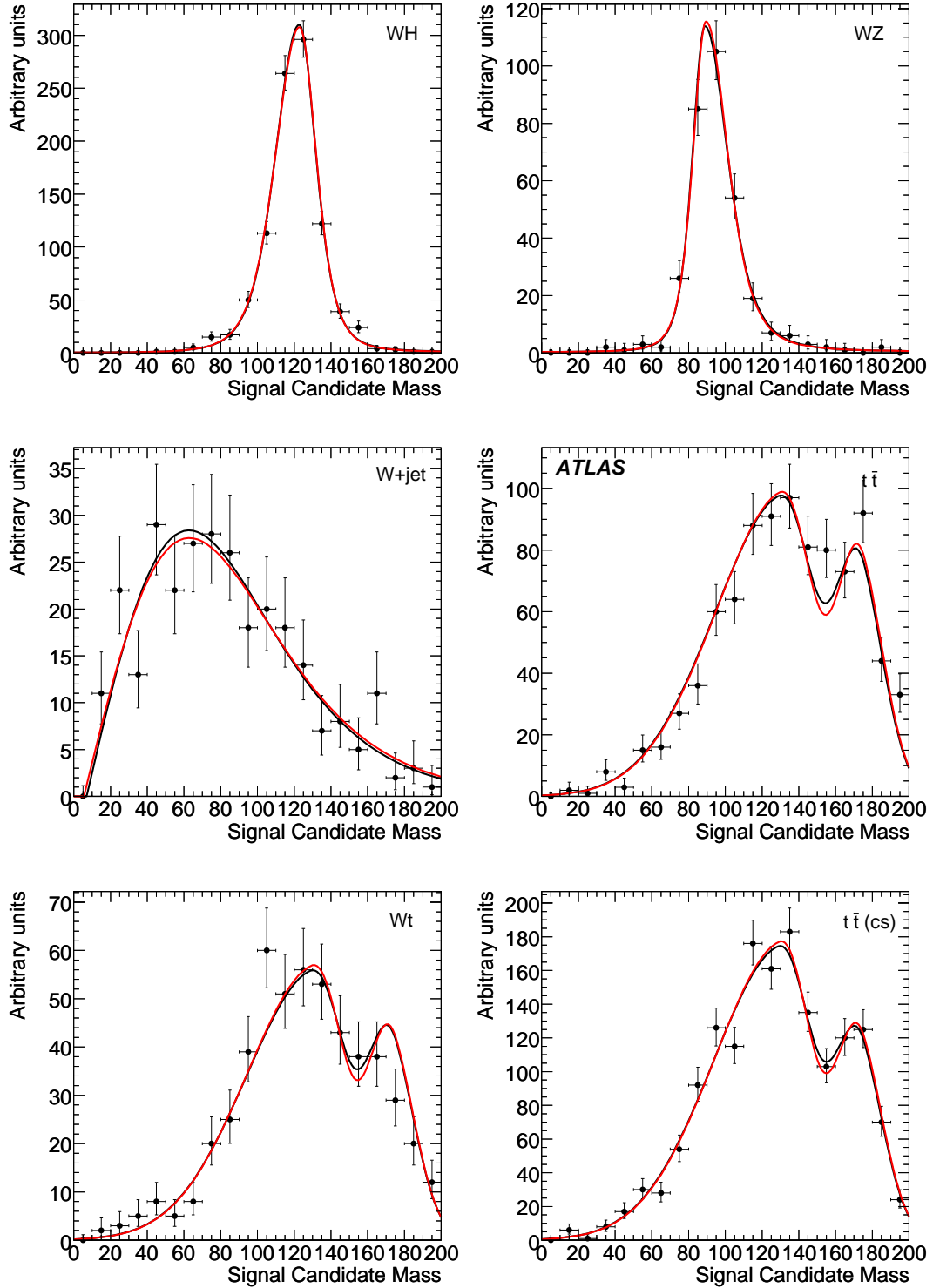


Figure 8.3: PDFs for the filtered mass variable (in GeV) for the signal and the various backgrounds. Overlaid are the respective mass distributions from Monte Carlo. Samples: signal (top left), WZ (top right), W +jet (middle left), $t\bar{t}$ (middle right), Wt (bottom left) and $t\bar{t}$ from the control sample (bottom right); in black the parametrization for the *perfect* case, in red for the *nominal* one.

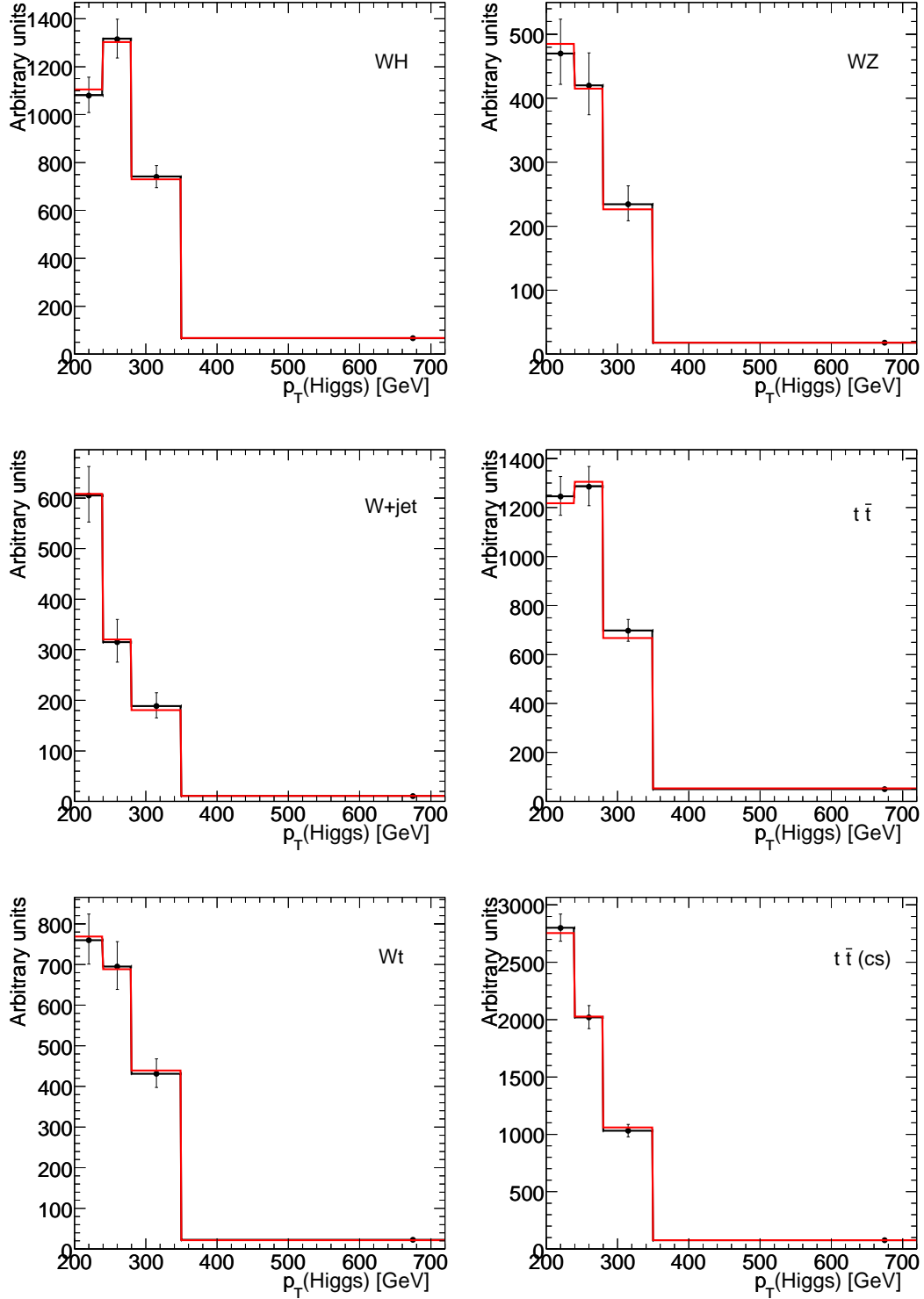


Figure 8.4: PDFs for the transverse momentum of the Higgs boson for the signal and the various backgrounds. Overlaid are their distributions from Monte Carlo. Samples: signal (top left), WZ (top right), $W+\text{jet}$ (middle left), $t\bar{t}$ (middle right), Wt (bottom left) and $t\bar{t}$ from the control sample (bottom right); in black the parametrisation for the *perfect* case, in red for the *nominal* one.

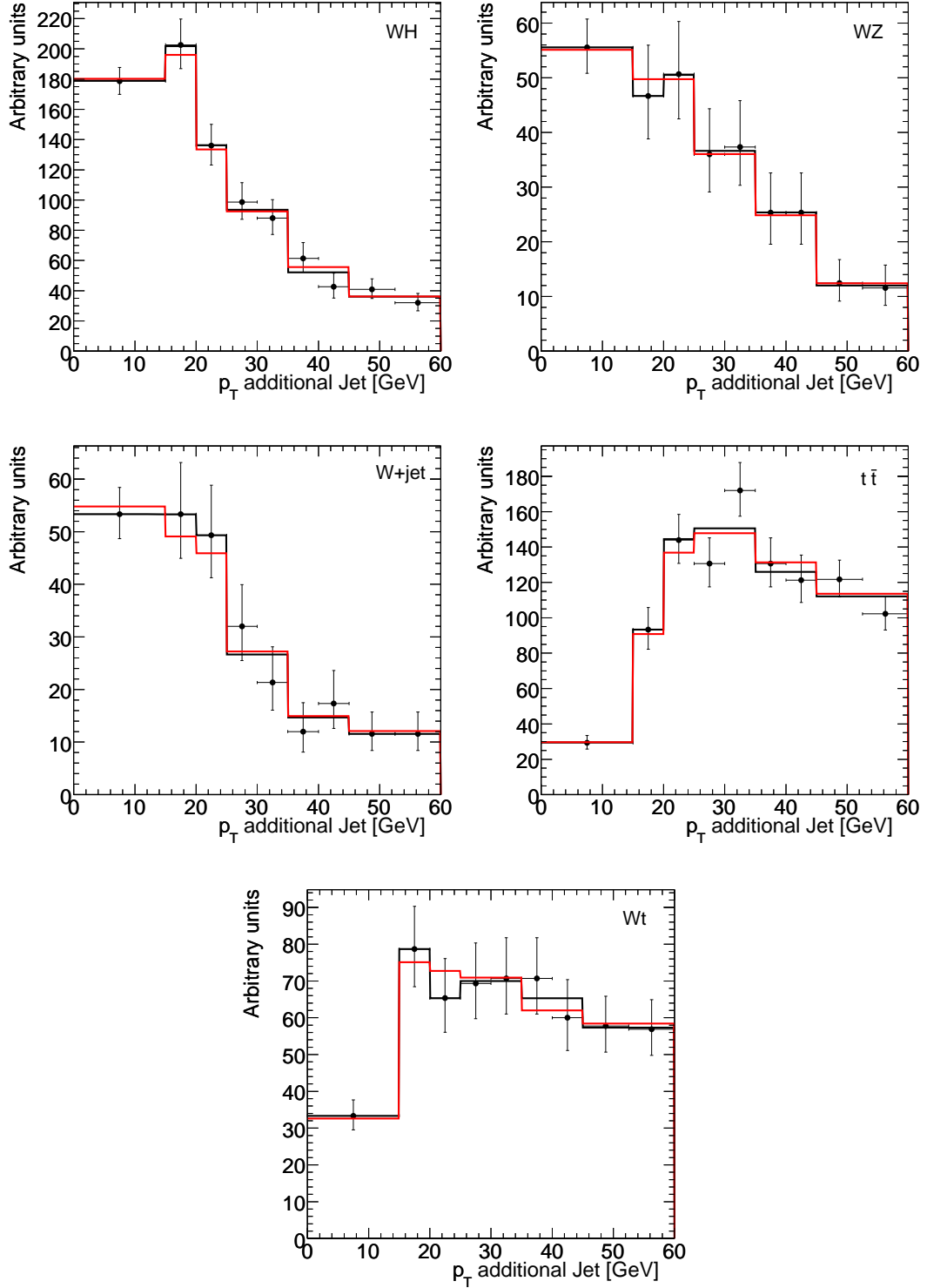


Figure 8.5: PDFs for the transverse momentum of the hardest additional jet for the signal and the various backgrounds. Overlaid are the respective distributions from Monte Carlo. Samples: signal (top left), WZ (top right), W+jet (middle left), $t\bar{t}$ (middle right) and Wt (bottom); in black the parametrisation for the *perfect* case, in red for the *nominal* one.

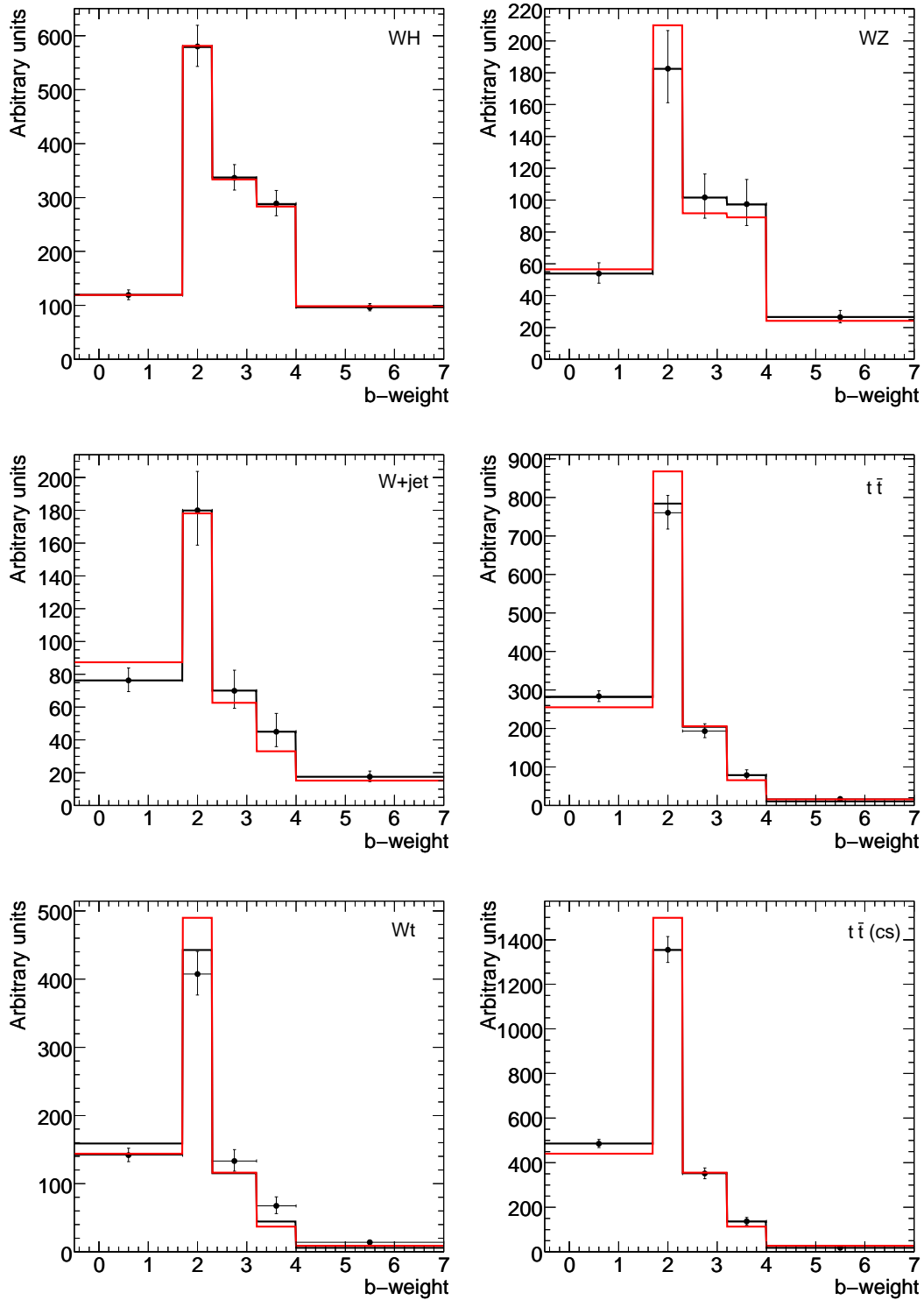


Figure 8.6: PDFs for the b -weight variable for the signal and the various backgrounds. Overlaid are the respective distributions from Monte Carlo. Samples: signal (top left), WZ (top right), W +jet (middle left), $t\bar{t}$ (middle right), Wt (bottom left) and $t\bar{t}$ from the control sample (bottom right); in black the parametrisation for the *perfect* case, in red for the *nominal* one.

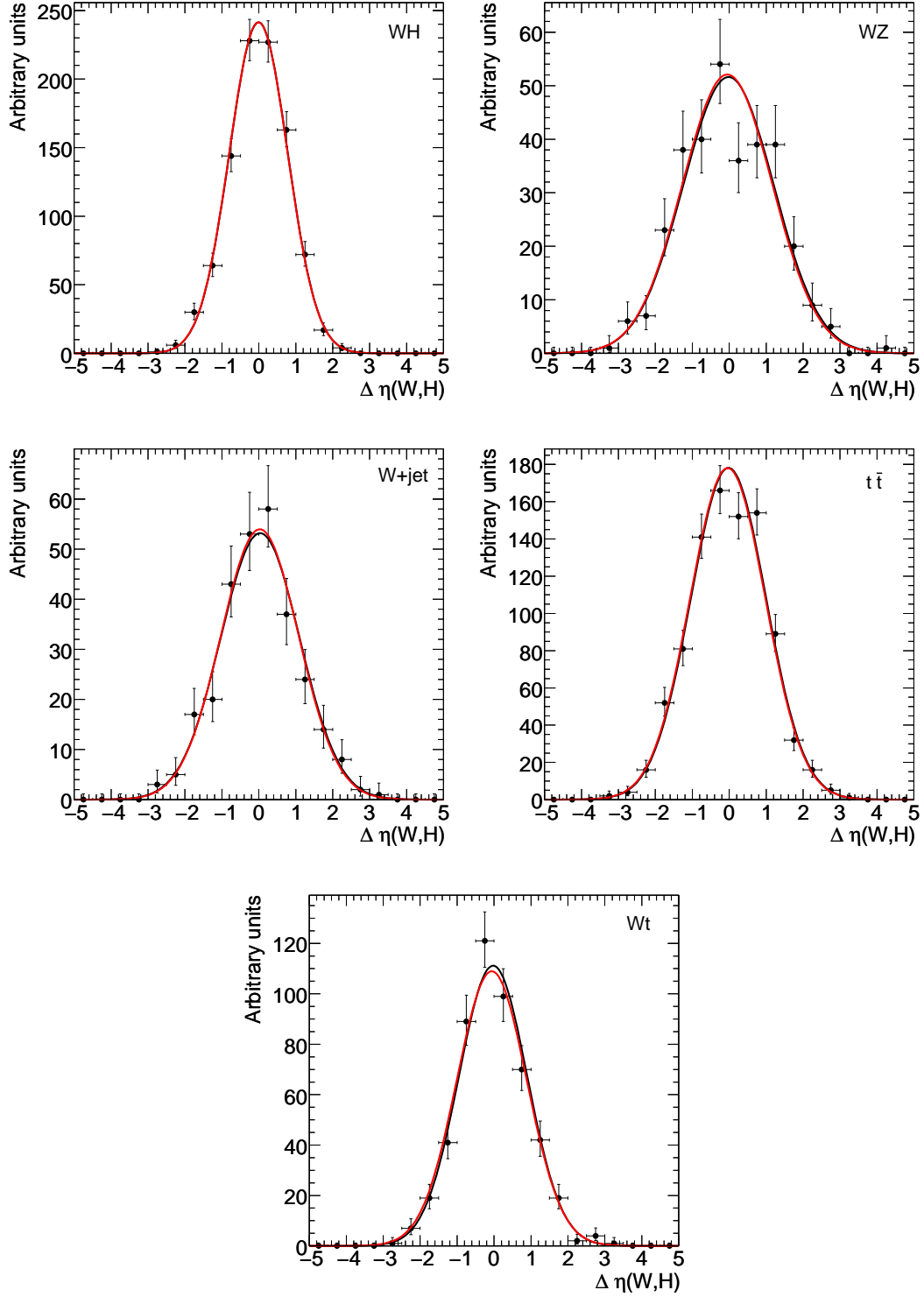


Figure 8.7: PDFs for the $\Delta\eta(W,H)$ variable for the signal and the various backgrounds. Overlaid are the respective distributions from Monte Carlo. Samples: signal (top left), WZ (top right), $W+\text{jet}$ (middle left), $t\bar{t}$ (middle right) and Wt (bottom); in black the parametrisation for the *perfect* case, in red for the *nominal* one.

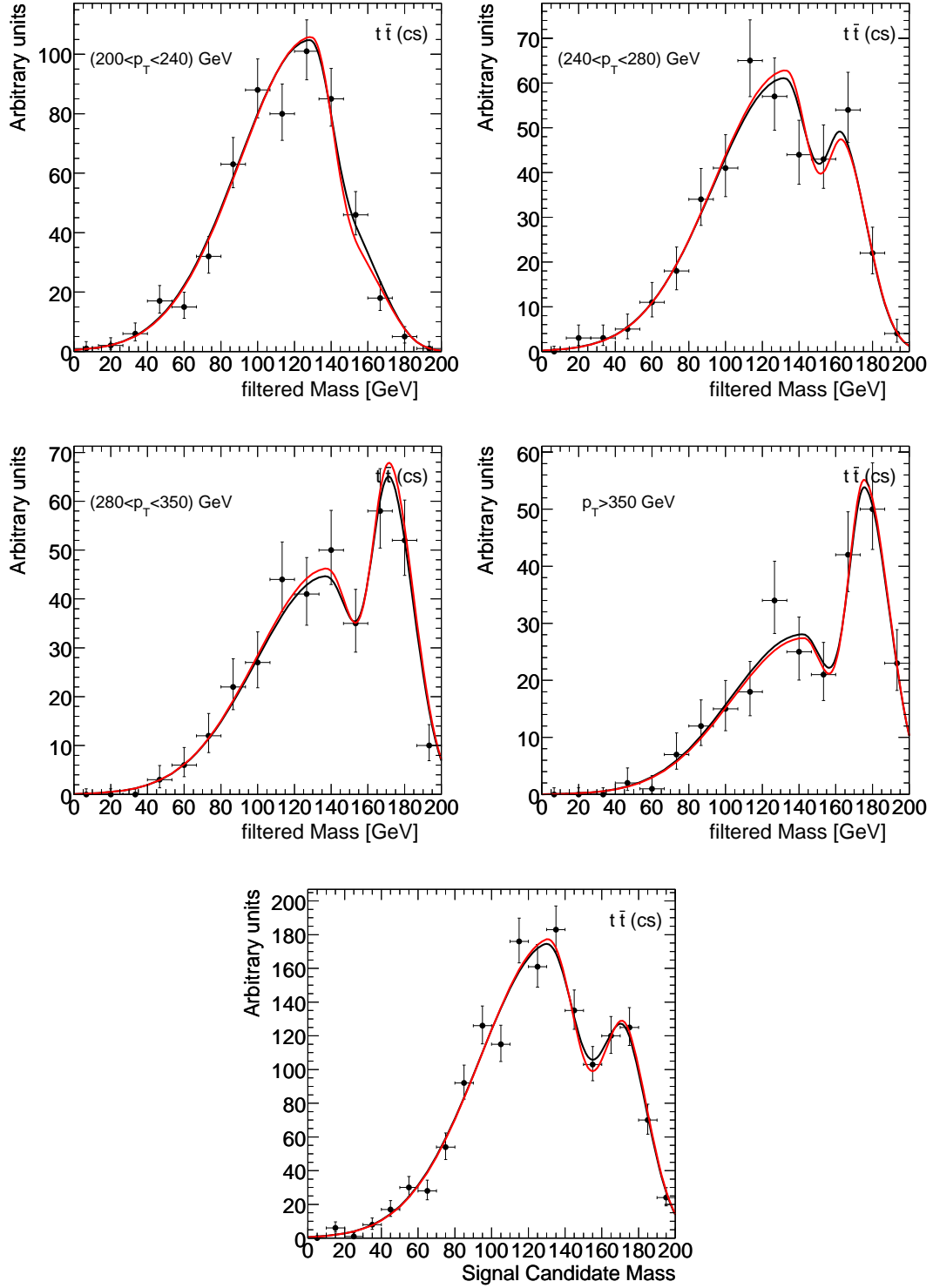


Figure 8.8: PDFs of the filtered mass of the Higgs boson candidate for the $t\bar{t}$ background in the control sample for various $p_T(\text{Higgs})$ bins. Overlaid are the distributions from Monte Carlo. For comparison the inclusive distribution is shown on the bottom. In black the parametrisation for the *perfect* case, in red for the *nominal* one.

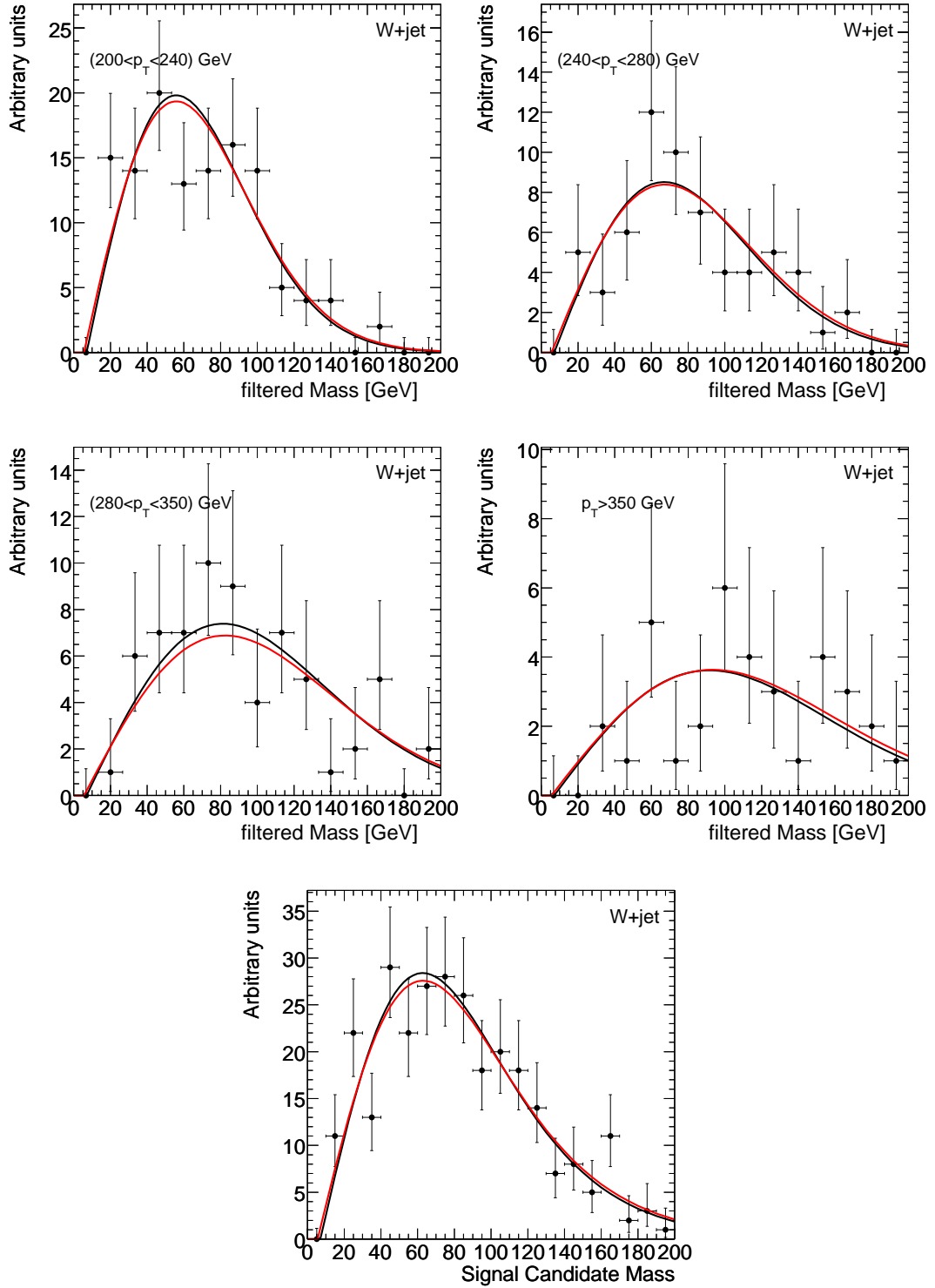


Figure 8.9: PDFs for the filtered mass of the Higgs boson candidate for the W +jet background for various $p_T(\text{Higgs})$ bins, with overlaid the distributions from Monte Carlo. For comparison the inclusive distribution is shown on the bottom. In black the parametrisation for the *perfect* case, in red for the *nominal* one.

accounts also for the presence of eventual non-linear correlations. The result is that these correlations are found to be negligible, in the limit of the available Monte Carlo statistics.

	p_T add. jet	filtered mass	$p_T(\text{Higgs})$	$\Delta\eta(W,H)$	b -weight
p_T add. jet	100%	7%	9%	0%	1%
filtered mass		100%	2%	0%	6%
$p_T(\text{Higgs})$			100%	2%	-12%
$\Delta\eta(W,H)$				100%	1%
b -weight					100%

Table 8.1: Correlation factors between fit variables in the signal.

	p_T add. jet	filtered mass	$p_T(\text{Higgs})$	$\Delta\eta(W,H)$	b -weight
p_T add. jet	100%	-4%	11%	7%	-3%
filtered mass		100%	-7%	-2%	9%
$p_T(\text{Higgs})$			100%	-3%	-5%
$\Delta\eta(W,H)$				100%	1%
b -weight					100%

Table 8.2: Correlation factors between fit variables in the WZ background.

	p_T add. jet	filtered mass	$p_T(\text{Higgs})$	$\Delta\eta(W,H)$	b -weight
p_T add. jet	100%	12%	8%	-10%	2%
filtered mass		100%	34%	-0%	12%
$p_T(\text{Higgs})$			100%	-2%	-0%
$\Delta\eta(W,H)$				100%	10%
b -weight					100%

Table 8.3: Correlation factors between fit variables in the W +jet background.

	p_T add. jet	filtered mass	$p_T(\text{Higgs})$	$\Delta\eta(W,H)$	b -weight
p_T add. jet	100%	5%	-4%	8%	5%
filtered mass		100%	34%	-1%	3%
$p_T(\text{Higgs})$			100%	-2%	-0%
$\Delta\eta(W,H)$				100%	2%
b -weight					100%

Table 8.4: Correlation factors between fit variables in the $t\bar{t}$ background.

	p_T add. jet	filtered mass	$p_T(\text{Higgs})$	$\Delta\eta(W,H)$	b -weight
p_T add. jet	100%	3%	4%	3%	3%
filtered mass		100%	31%	-0%	2%
$p_T(\text{Higgs})$			100%	3%	-5%
$\Delta\eta(W,H)$				100%	-3%
b -weight					100%

Table 8.5: Correlation factors between fit variables in the Wt background.

8.3 Definition of control samples

To reduce the dependence on the Monte Carlo simulation and on the effect of systematics uncertainties, it is desirable to determine the shapes of the PDFs for the main backgrounds directly on data, by defining appropriate control samples. In order for this to work, it is crucial to demonstrate that the shapes of the PDFs in the signal and control regions are the same and, at the same time, that the selection adopted for the control region does prevent events from the signal or any other background to be accepted in the control sample.

In this section, the event selection which permits to define two control samples using $t\bar{t}$ events is described. The first permits to obtain an estimate of the mass and b -weight PDFs for both the $t\bar{t}$ background and the Wt backgrounds, the second permits to obtain the PDF for the p_T of the hardest additional jet in the event for the $t\bar{t}$ background. After a description of the selection procedure and of the physics motivation behind it, the compatibility of the PDFs in the signal and control regions is demonstrated. Finally the inclusion of this information into the fit is discussed.

In order to define a control sample for the $t\bar{t}$ background, a basic understanding of the topological configuration of the top anti-top pair faking the WH signal is needed. After the requirement of a highly boosted back-to-back $b\bar{b}$ and W boson system, and after application of b -tagging, a $t\bar{t}$ event can enter the signal selection essentially only if two boosted back-to-back top quark have been produced. On one side the top quark produces one high p_T W boson, which fulfils the selection requirements, in association with a softer b -jet, which fails the b -tagging veto and has p_T lower than 60 GeV. On the other side, the remaining top quark fakes the Higgs boson candidate, where one of the decay products is a b -jet and the other is typically a charm-jet, coming from the decay of the W boson originating from the same top quark. The third jet (from the W boson) is either low p_T or collinear with the other two jets. This gives rise to the double peak structure already seen in Fig. 8.3, where the second peak is positioned in proximity of the top quark mass. In fewer cases a combination of a b - and a light-jet is selected, where the light-jet can come either from the W boson directly, or through final state radiation: again, a very similar mass distribution is expected.

8.3.1 Control sample for mass and b -weight PDFs

In order to get the filtered mass and b -weight distributions of the top quark mimicking the $b\bar{b}$ system, it is sufficient to reproduce the correct kinematic configuration of one of the two top quarks. The other top quark can be thus fully reconstructed in the following way:

- Require the W boson to have $p_T > 80$ GeV, instead of 200 GeV;
- Revert the b -jet veto, requiring a b -jet with $p_T > 15$ GeV to be present in the event;

- Reconstruct the leptonic top mass, requiring ($148 < m_{top}^{lep} < 188$) GeV and $p_T(\text{top}) > 225$ GeV.

The remaining selection cuts are left unchanged and follow the nominal analysis, including the additional light-jet veto cut. Since the b -jet veto is exactly reverted, there is no overlap between the control sample and the signal region. The complete four momentum of the neutrino is obtained by assuming that the E_T^{miss} in the event originates exclusively from the neutrino from the W boson decay and by solving the equation which requires the invariant mass of the lepton plus neutrino system to be equal to the nominal W boson mass. The solution of this equation is given by:

$$p_{2,z} = p_{1,z} \frac{\sqrt{m_1^2 + p_{1,T}^2}}{\sqrt{m_2^2 + p_{2,T}^2}} \left[1 \pm \sqrt{\frac{m_1^2 + p_{1,T}^2}{m_2^2 + p_{2,T}^2}} \sqrt{1 - \frac{(m_1^2 + p_{1,T}^2)(m_2^2 + p_{2,T}^2)}{\left(\frac{m_W^2 - m_1^2 - m_2^2}{2} + \vec{p}_{1,T} \cdot \vec{p}_{2,T}\right)^2}} \right], \quad (8.12)$$

where p_1 is the four momentum of the lepton, p_2 is the four momentum of the neutrino and m_W is the nominal W boson mass. Of the two possible solutions for p_z the smallest one is taken, since this is in most of the cases the correct one. This equation has one or two real solutions only under the following condition:

$$m_W^2 \geq m_1^2 + m_2^2 + 2 \left[\sqrt{m_1^2 + p_{1,T}^2} \sqrt{m_2^2 + p_{2,T}^2} - \vec{p}_{1,T} \cdot \vec{p}_{2,T} \right], \quad (8.13)$$

which physically expresses the constraint that the transverse mass of the lepton and neutrino system cannot exceed the nominal W mass. This corresponds to the limiting case where the longitudinal components of the momenta of the neutrino and of the lepton are not contributing to the invariant mass of the W boson. Due to resolution effects, however, or, much more rarely, due to the intrinsic width of the W boson resonance, the reconstructed transverse mass can exceed this value. In this case, the hypothesis is made that the limiting condition of Eq. 8.13 is fulfilled, and therefore one only solution is obtained:

$$p_{2,z} = p_{1,z} \frac{\sqrt{m_1^2 + p_{1,T}^2}}{\sqrt{m_2^2 + p_{2,T}^2}} \quad (8.14)$$

The invariant mass of the leptonic top before applying the mass window cut is shown in Fig. 8.10.

The selection just described is applied to the signal and all background samples, as shown in Tables 8.6 and 8.7. In 30 fb^{-1} of data, 982 ± 25 $t\bar{t}$ events are selected, while the amount of signal and remaining backgrounds passing the selection is negligible: the selected control sample can be considered to be background free. The amount of Wt background feeding into the selection is not completely negligible, however, as will be shown in the following, the same Higgs invariant mass distribution is expected as in $t\bar{t}$ events, so that this component can be simply considered as a part of the control sample.

After the selection for the first $t\bar{t}$ control sample is defined, it needs to be demonstrated that the Higgs boson candidate invariant mass and the b -weight distributions are exactly the same as in the signal region. However, in the control sample, due to the way the leptonic top is selected, a kinematic bias will be induced also on the top quark (on the opposite side of the event) which mimics the $b\bar{b}$ system representing the Higgs boson candidate, which will have a slightly different kinematic configuration than in the signal region. This affects the p_T distribution of

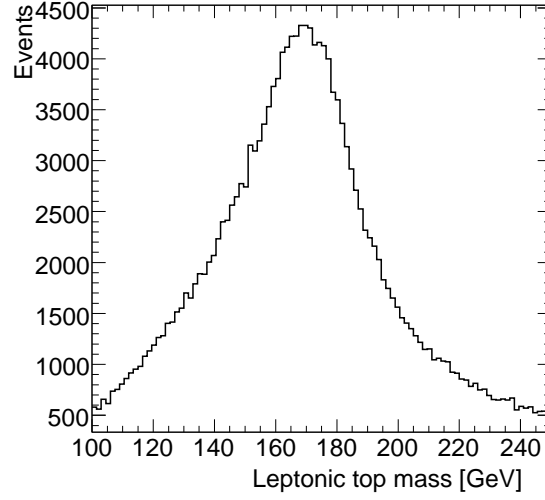


Figure 8.10: Leptonic top mass as reconstructed in the $t\bar{t}$ sample, at the selection stage where usually in the analysis a b -jet veto is required. The normalisation is arbitrary and corresponds to the available Monte Carlo statistics.

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	W+jets	Wt
After filter cuts	962.3 ± 5.9	858.4 ± 6.4	655.2 ± 4.1	3935	1229506	582364	93155
1 Higgs candidate	646.2 ± 2.8	569.7 ± 3.0	429.7 ± 1.9	3509.7 ± 8.0	806175	562030	69375
filtered $p_T > 200$ GeV	581.7 ± 2.9	512.7 ± 3.2	387.6 ± 2.0	3108 ± 10	709271	413406	60241
Missing $E_T > 30$ GeV	413.7 ± 2.9	362.4 ± 3.2	273.6 ± 2.0	2183 ± 13	552284	318400	46779
$p_T(W) > 200$ GeV	283.6 ± 2.7	248.3 ± 2.9	183.3 ± 1.8	1830 ± 13	393385	280223	36616
$p_T(e/\mu) > 30$ GeV	201.8 ± 2.4	173.9 ± 2.6	128.5 ± 1.6	1377 ± 12	282569	223319	26624
$p_T(\text{additional } \mu) < 10$ GeV	200.4 ± 2.4	172.5 ± 2.6	127.6 ± 1.6	1284 ± 12	257044	222648	25047
$p_T(\text{additional } e) < 10$ GeV	197.1 ± 2.4	169.8 ± 2.6	125.4 ± 1.6	1170 ± 12	230888	218839	23085
$\Delta\phi(W,H) < \frac{\pi}{2}$	195.0 ± 2.4	168.4 ± 2.6	123.7 ± 1.6	1099 ± 12	190374	208005	20724
leptonic top selection	3.77 ± 0.37	4.10 ± 0.44	3.00 ± 0.27	18.6 ± 1.8	64740	3607	3388
no additional b -jets $p_T > 15$ GeV	3.41 ± 0.35	3.76 ± 0.42	2.93 ± 0.27	16.9 ± 1.7	50380	3393	2676
jets on W side $p_T < 60$ GeV	2.89 ± 0.33	2.99 ± 0.38	2.30 ± 0.24	13.2 ± 1.5	35887	2139	1961
jets on H side $p_T < 60$ GeV	2.78 ± 0.32	2.70 ± 0.36	2.17 ± 0.23	10.5 ± 1.3	26654	1489	1560
one subject b -tagged	2.20 ± 0.28	2.07 ± 0.32	1.74 ± 0.21	4.22 ± 0.84	16310	118 ± 10	367 ± 11
both subjects b -tagged	0.70 ± 0.16	0.67 ± 0.18	0.50 ± 0.11	0.84 ± 0.38	975	5.5 ± 2.3	15.8 ± 2.2
loose fit cuts	0.66 ± 0.16	0.67 ± 0.18	0.50 ± 0.11	0.84 ± 0.38	956 ± 24	5.5 ± 2.3	15.5 ± 2.2

Table 8.6: Expected number of events passing the $t\bar{t}$ control sample selection criteria for the signal and the main backgrounds, projected to 30 fb^{-1} of integrated luminosity.

the Higgs boson candidate in the control sample, which is compared to the signal region in Fig. 8.11. This difference induces a slightly different overall distribution for the Higgs invariant mass in the signal and control region. However, in the likelihood function, it does not need to be assumed that the overall $PDF(m(H), p_T(H)) = PDF(m(H)|p_T(H))PDF(p_T(H))$ is common for signal and control region, but it can just be assumed that the conditional dependence on $p_T(H)$ is the same in both signal and control region. This is equivalent to assuming that only the conditional PDF ($PDF(m(H)|p_T(H))$) is in common with the control sample, while different distributions for $PDF(p_T(H))$ are assumed.

In order to disentangle the two effects, a correction factor has been worked out as a function of $p_T(\text{Higgs})$: this expresses the ratio between the normalised $p_T(\text{Higgs})$ distribution in the signal region and the same distribution in the control sample as a function of $p_T(\text{Higgs})$, as shown in Fig. 8.11. Applying this correction factor, a re-weighted filtered mass distribution is obtained for the control sample, which can then be directly compared with the filtered mass distribution in the signal region, as shown in Fig. 8.12 (bottom). A more direct test of the equivalence of the conditional dependence on $p_T(H)$ of the conditional PDF $P_{t\bar{t}}(m_{\text{filtered},i}|p_T)$

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jets
After filter cuts	439.6 ± 2.9	8630	4079	1789159	276455
1 Higgs candidate	295.9 ± 1.3	8428.0 ± 5.0	3372.1 ± 8.1	160154	225597
filtered $p_T > 200$ GeV	267.8 ± 1.4	7355 ± 12	2993.4 ± 9.4	103170	175061
Missing $E_T > 30$ GeV	141.9 ± 1.3	6217 ± 15	1414 ± 10	75129	62883
$p_T(W) > 200$ GeV	68.1 ± 1.0	5770	910.4 ± 8.9	37552	29982
$p_T(e/\mu) > 30$ GeV	19.65 ± 0.59	4449	530.2 ± 7.2	27543	17754
$p_T(\text{additional } \mu) < 10$ GeV	12.48 ± 0.47	4414	384.0 ± 6.2	24892	12836
$p_T(\text{additional } e) < 10$ GeV	6.80 ± 0.35	4291	208.6 ± 4.7	21510	7978
$\Delta\phi(W,H) < \frac{2}{3}\pi$	6.35 ± 0.34	4123	185.5 ± 4.4	14888	6587
leptonic top selection	0.296 ± 0.074	59.9 ± 2.8	4.58 ± 0.72	4328	133 ± 10
jets on W side $p_T < 60$ GeV	0.204 ± 0.061	38.4 ± 2.2	2.80 ± 0.56	1470	72.6 ± 7.5
jets on H side $p_T < 60$ GeV	0.185 ± 0.059	29.1 ± 1.9	2.12 ± 0.49	1145	54.6 ± 6.5
one subjet b -tagged	0.167 ± 0.056	2.71 ± 0.59	0.56 ± 0.25	535	7.0 ± 2.3
both subjets b -tagged	0.056 ± 0.032	0.00 ± 0.13	0.22 ± 0.16	25.9 ± 6.3	0.00 ± 0.78
loose fit cuts	0.056 ± 0.032	0.00 ± 0.13	0.22 ± 0.16	25.9 ± 6.3	0.00 ± 0.78

Table 8.7: Expected number of events passing the $t\bar{t}$ control sample selection criteria for the remaining backgrounds, projected to 30 fb^{-1} of integrated luminosity.

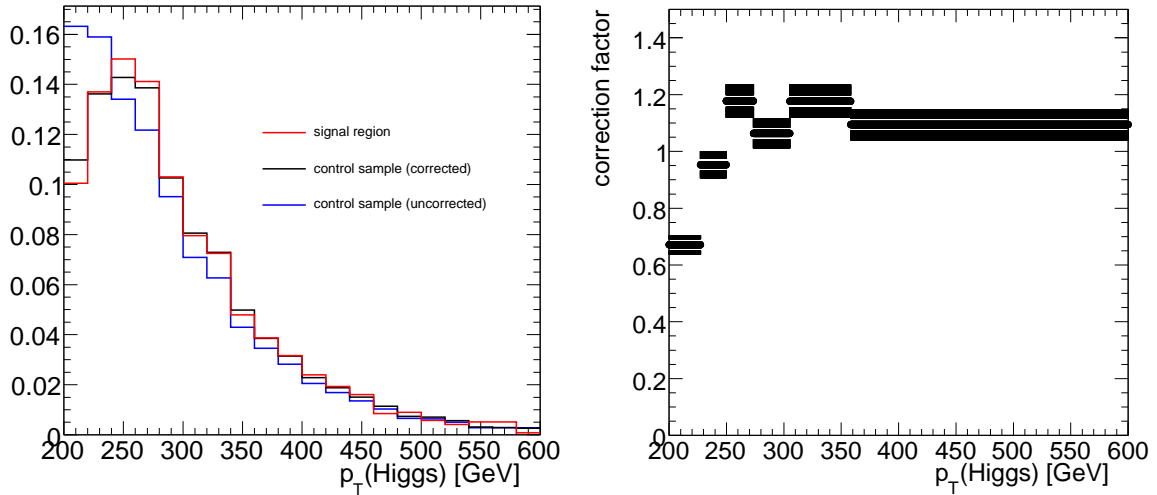


Figure 8.11: (Left) Transverse momentum of the Higgs boson candidates for $t\bar{t}$ background in the control sample compared to those in the signal region, before and after applying the correction, normalised to their respective area. (Right) Correction factor for $t\bar{t}$ background as a function of $p_T(\text{Higgs})$ accounting for the different kinematic configuration in the control sample with respect to signal region.

is provided by comparing the Higgs boson candidate invariant mass in bins of $p_T(H)$. This is also shown in Fig. 8.12.

Due to the limited available Monte Carlo statistics the dependence on $p_T(\text{Higgs})$ of the filtered mass distribution cannot be compared between the signal region and the control sample very precisely. In order to increase the available statistics and make a more detailed comparison possible, the b -tagging requirement in the analysis selection has been applied also directly through the use of truth level information, selecting explicitly b - c quark-jet combinations, which represent the biggest contribution expected in the $t\bar{t}$ background after application of b -tagging, or selecting b -light quark-jet combinations, which represents the remaining minor contribution. The result for both b - c and b -light quark-jet combinations is

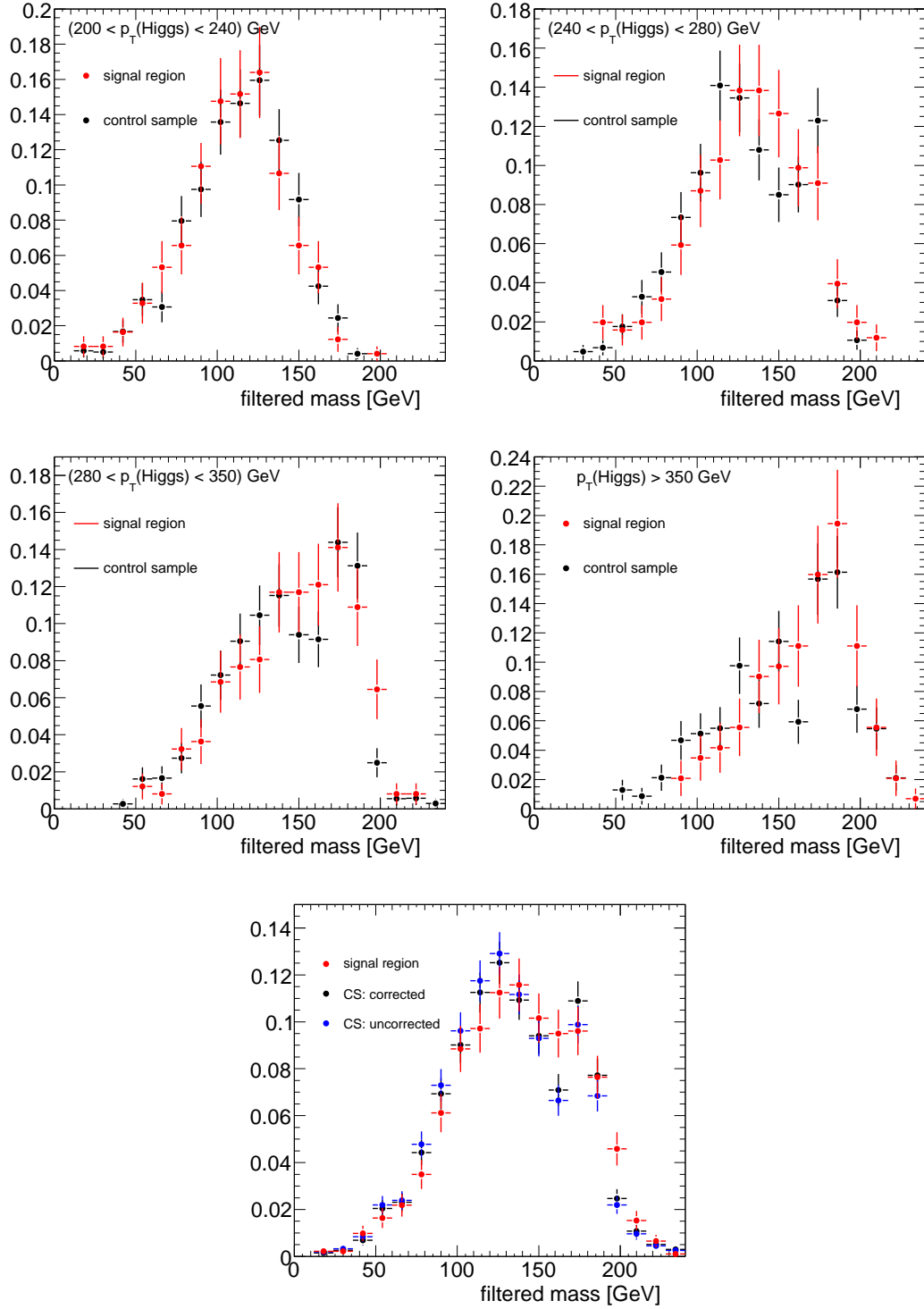


Figure 8.12: Distributions of the filtered mass for the $t\bar{t}$ background for various $p_T(\text{Higgs})$ bins, as obtained from the control sample, compared to the same distribution in the signal region. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

shown respectively in Fig. 8.13 and in Fig. 8.14: the dependence on $p_T(\text{Higgs})$ can be studied in more detail and it can be clearly seen that the shape of the filtered mass distribution is reproduced extremely well in the control sample, in every single chosen $p_T(\text{Higgs})$ bin. A simple Kolmogoroff test shows that all distributions are perfectly compatible with each other, except maybe for the $p_T(\text{Higgs}) > 350$ GeV bin, where a small deviation is seen in the left tail of the mass distribution. The inclusive distribution shows that, after the re-weighting procedure in $p_T(\text{Higgs})$ already mentioned, a perfect agreement between signal region and control sample is obtained.

Since the b - c and b -light jet component in the $t\bar{t}$ background is correctly reproduced, it can be concluded that the kinematic configuration of the top quark mimicking the Higgs boson decaying to $b\bar{b}$ in the control sample is correctly reproduced as well. As an additional check, Fig. 8.15 shows the distribution in pseudo-rapidity of the Higgs boson candidate in the control sample after correction, compared to the same distribution in signal region. Again, the agreement is very good.

Also the b -weight distribution for the $t\bar{t}$ background can be obtained from the same control sample, since this quantity again involves the top quark mimicking the $H \rightarrow b\bar{b}$ system. A comparison is shown in Fig. 8.16. The agreement is again very good. In this case there is no reason to care about an eventual dependence on $p_T(\text{Higgs})$, since this dependence is sufficiently weak not to cause any bias on the b -weight distribution.

8.3.2 Control sample for p_T of the additional jet PDF

In order to obtain the PDF for the transverse momentum of the hardest additional jet in the event for the $t\bar{t}$ background directly from data, the definition of a second control sample is needed. In fact, in this case one needs both top quarks to remain unbiased with respect to the nominal analysis selection, since the hardest additional jet will very often result from a jet stemming directly or indirectly from one of the two top quarks. At the same time one needs to fully reconstruct at least one of the two top quarks in order to select a phase space region which is free from signal and all other background contributions.

In order to do so the following selection is applied:

- $p_T(W) > 200$ GeV (unchanged)
- Revert the b -jet veto, requiring a b -jet with $p_T > 15$ GeV to be present in the event
- Reconstruct the leptonic top mass, requiring $(148 < m_{top}^{lep} < 188)$ GeV.

The remaining cut flow is left unchanged with respect to the nominal signal selection.

While for the $t\bar{t}$ background in the signal region no distinction is made between b - and non b -jets for the jets entering the p_T of the additional jet distribution, in the control sample just defined two separate contributions are obtained, the b -jet from the leptonic top, which is explicitly selected, and the remaining hardest additional jet in the event. In the signal region, the b -jets contribute to the additional jet distribution only if they pass through the b -jet veto, while in the control sample they contribute only if they pass the complementary b -jet identification criterion. Since b -tagging does significantly depend on the jet transverse momentum it is applied on, the kinematic distributions of the b -jets from the leptonic top in the control sample and in the signal region are different. However, since the b -tagging efficiency as a function of jet p_T will be measured in data, this knowledge can be applied to re-weight the transverse momenta distribution of the identified b -jets from the top to obtain the corresponding one for the b -jets passing the b -veto in the signal region. In addition, the two separate b - and non b - contributions derived from the control sample need to be added

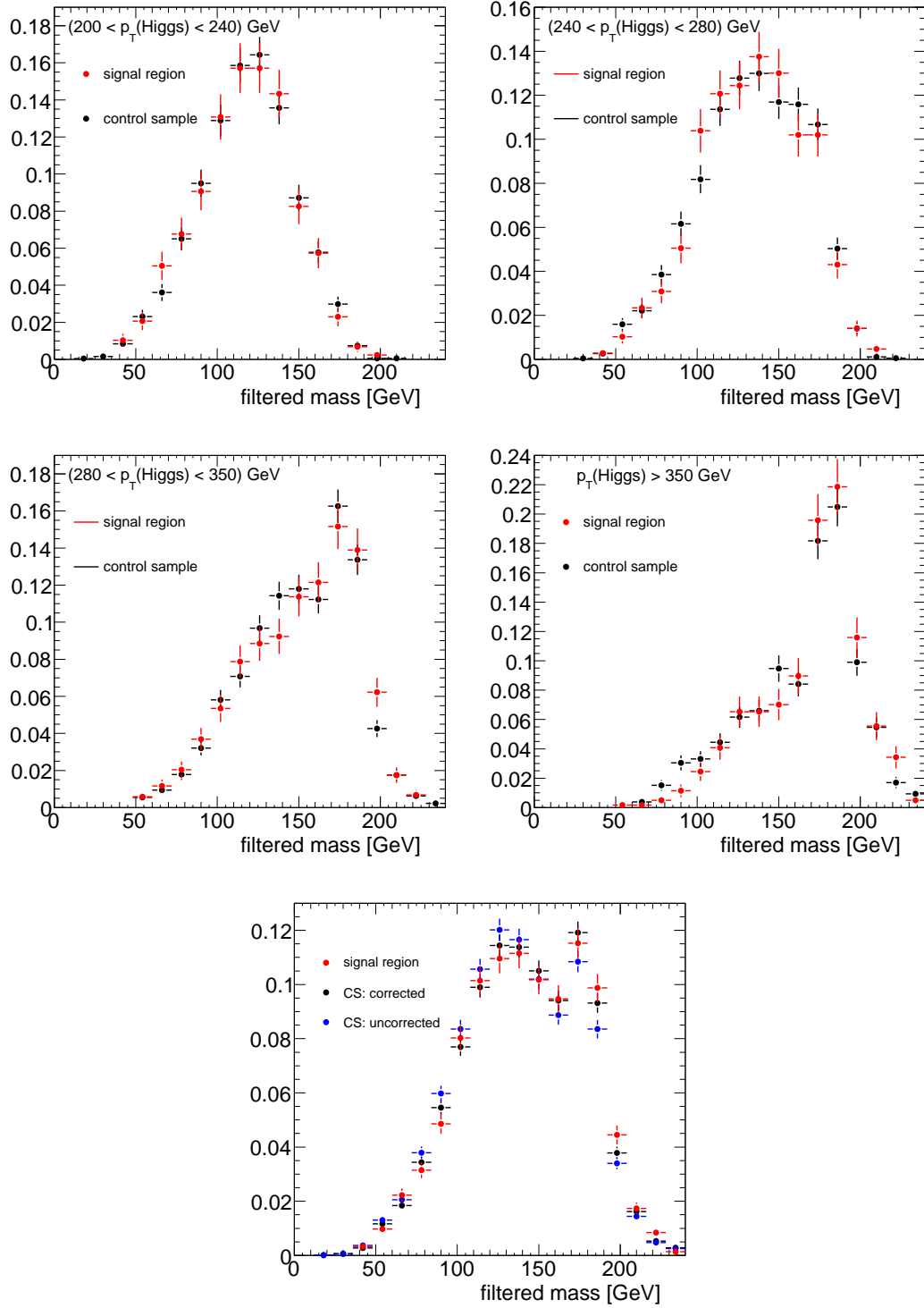


Figure 8.13: Distributions of the filtered mass for the $t\bar{t}$ background for various $p_T(\text{Higgs})$ bins, as obtained from the control sample selecting the b - c jet component through the use of truth level information. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

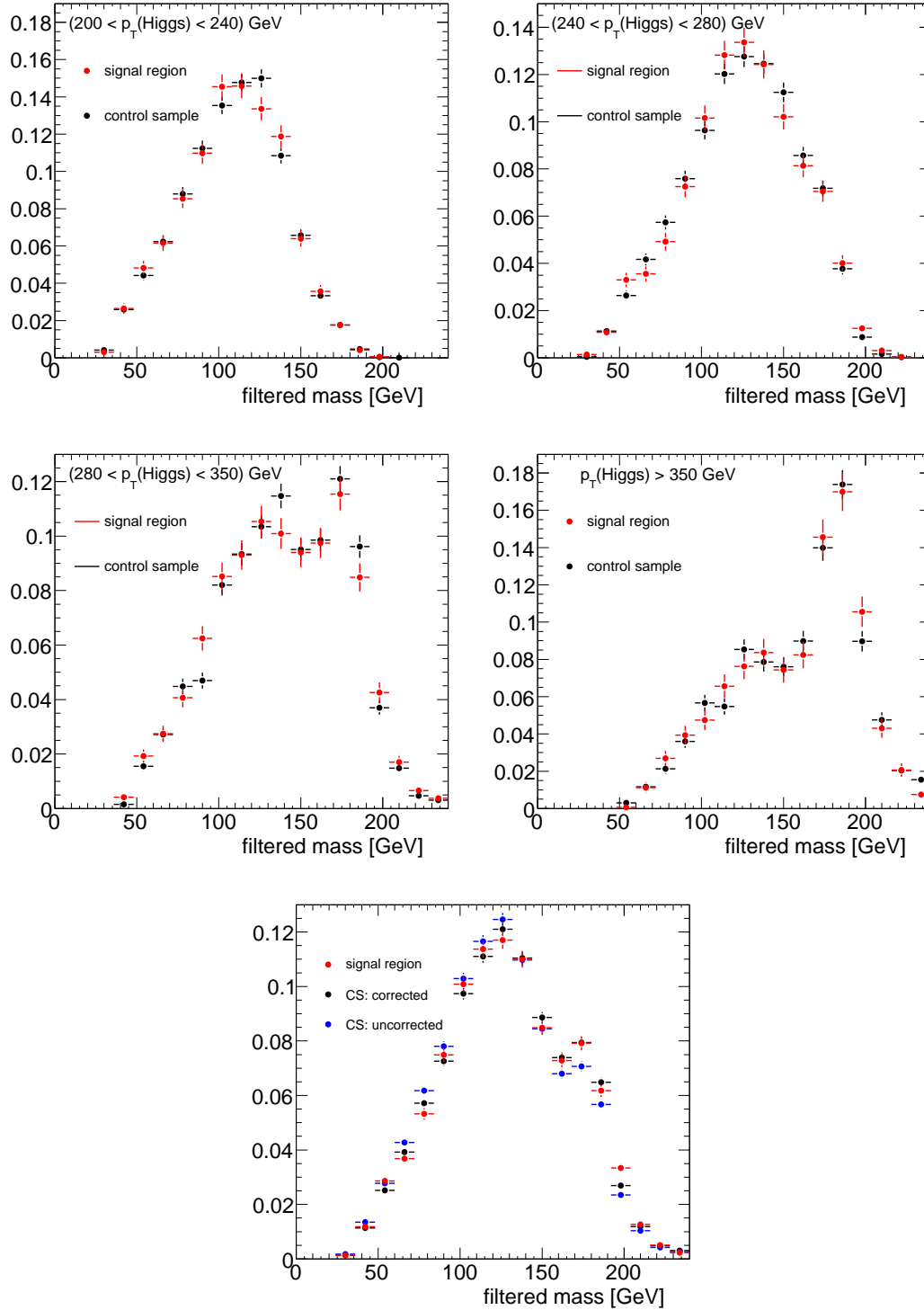


Figure 8.14: Distributions of the filtered mass for the $t\bar{t}$ background for various $p_T(\text{Higgs})$ bins, as obtained from the control sample selecting the b -light jet component through the use of truth level information. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

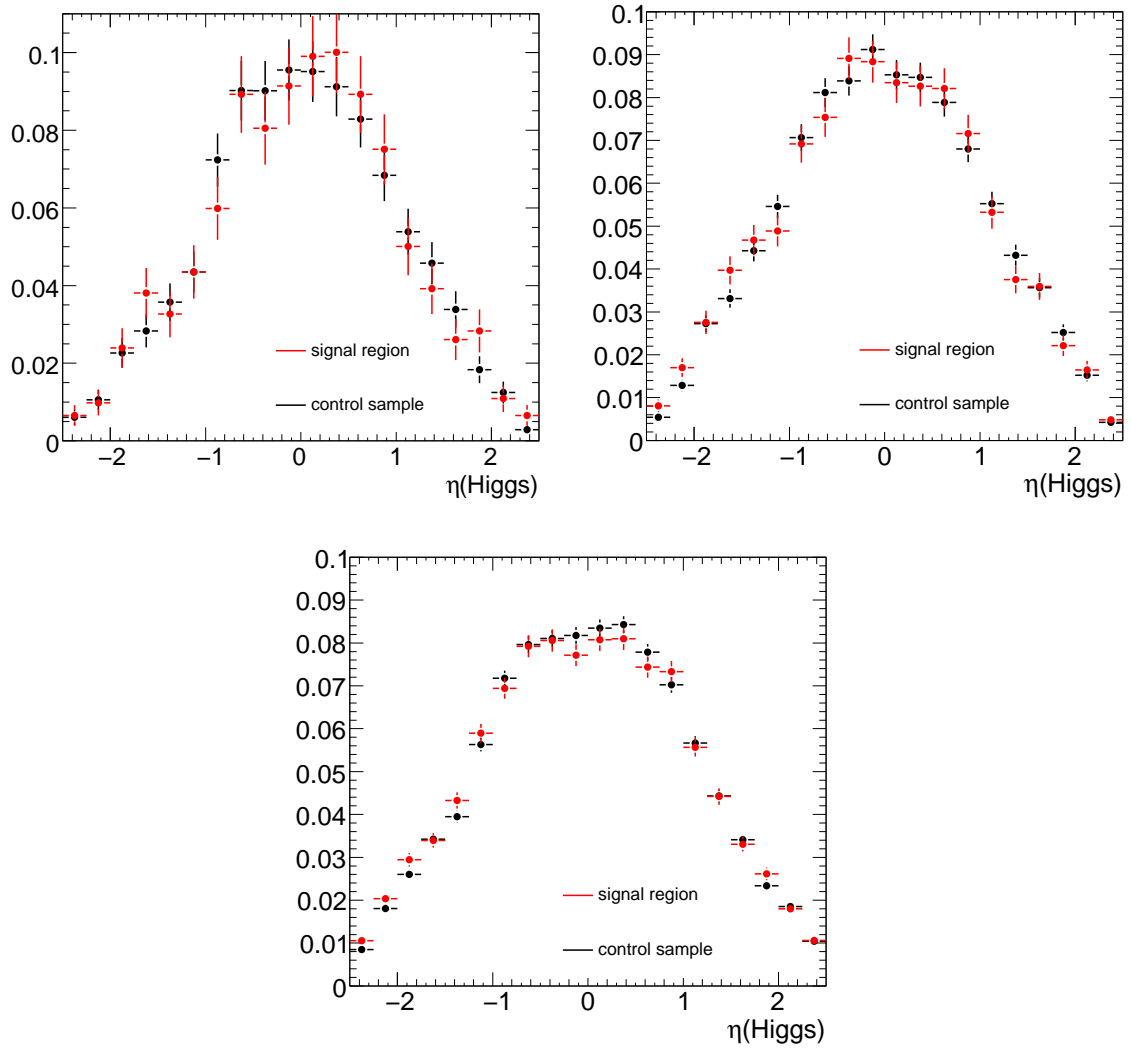


Figure 8.15: Pseudo-rapidity distribution of the Higgs to $b\bar{b}$ candidate in the $t\bar{t}$ background, on the left for the full selection, on the right after selecting only the b - c quark pair component and on the left after selecting only the b -light quark pair component using truth level information. The distributions are normalized to an area of one.

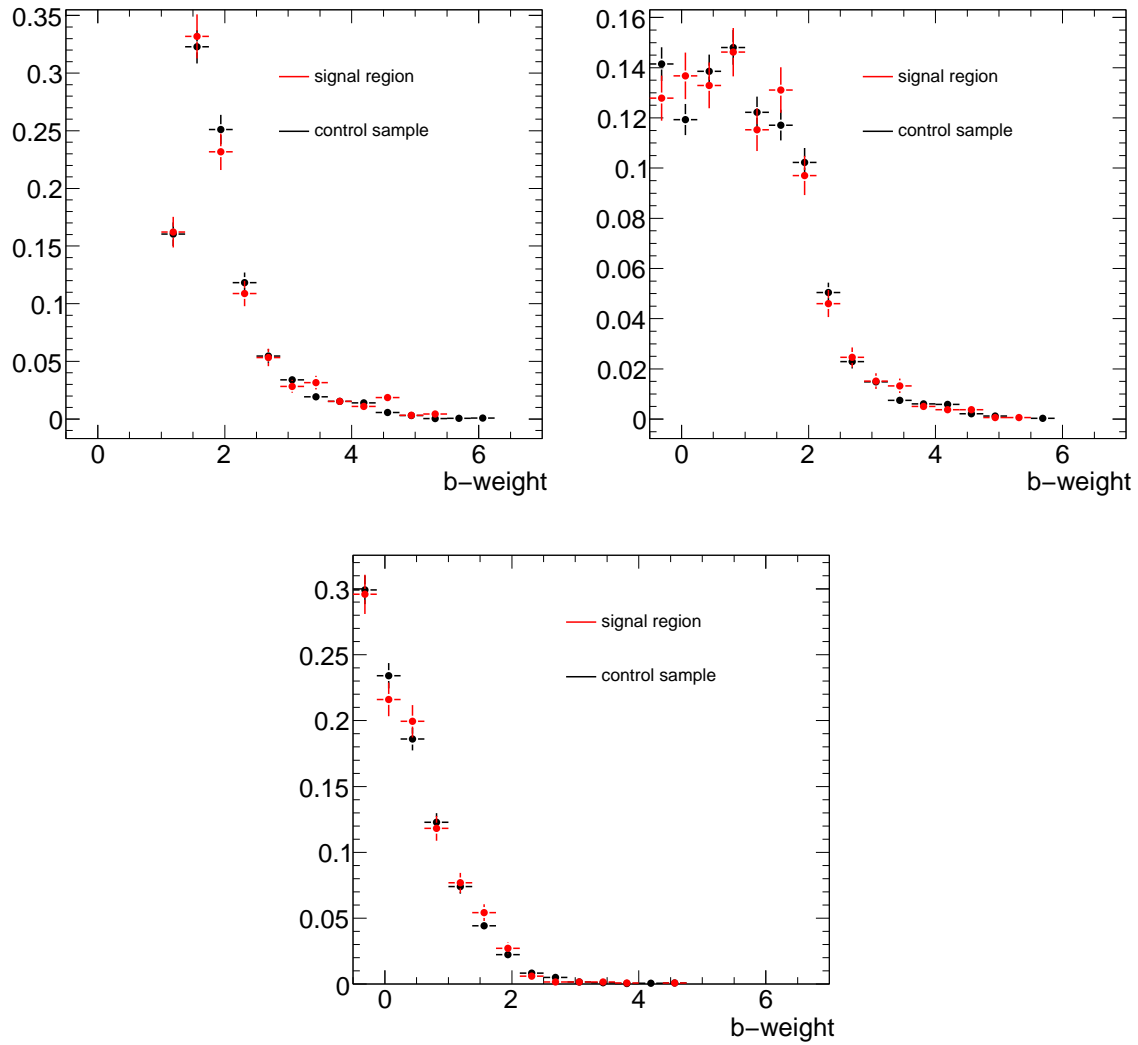


Figure 8.16: b -weight distribution for the Higgs to $b\bar{b}$ candidate in the $t\bar{t}$ background, on the left for the full selection, on the right after selecting directly the b - c quark pair component and on the bottom after selecting directly the b -light quark pair component using truth level information. The distributions are normalized to an area of one.

on top of each other: this is easily done by choosing - on an event basis - the hardest jet between the additional light jet contribution and the one corresponding to the vetoed b -jets, which is the only one which would appear in the p_T of the additional jet distribution in the signal region. The procedure described so far results in two simple steps to be applied on the control sample:

- The p_T of the jet with highest transverse momentum between the $p_T(b\text{-jet})$ from the leptonic top or $p_T(\text{additional jet})$ from the rest of the event is chosen.
- This contribution is re-weighted by $\frac{1-\epsilon_B}{\epsilon_B}$, where ϵ_B is the b -tagging efficiency corresponding to the p_T of the b -jet from the leptonic top.

As a drawback, the distribution obtained in this way misses the region of phase space where the b -jet from the leptonic top has transverse momentum below 15 GeV or is outside the inner detector acceptance⁵, since no event in the control sample can represent such a kinematic configuration. However, under the assumption that the two b - and non b -jet contributions in the control sample are uncorrelated, this will also be corrected for in the following.

The whole procedure relies on the knowledge of the b -tagging efficiency as a function of jet p_T . For the b -jet veto applied in the nominal analysis, corresponding to a cut on $\log(\frac{P_b}{P_l}) < -1$, the mistagging rate turns out to be relatively flat as a function of $p_T(\text{jet})$, with some fluctuations over the whole considered p_T range, while the b -jet efficiency drops at low $p_T(\text{jet})$, as expected (see Fig. 8.17). In the following, a relative precision of 5% on the b -tagging efficiency for each $p_T(\text{jet})$ bin is assumed, which is consistent with the most recent studies performed in ATLAS on the prospects to measure the b -tagging efficiency in data using $t\bar{t}$ or di-jet QCD events [44]. This is presumably a conservative assumption, since these studies rely on few hundreds of pb^{-1} of collected data.

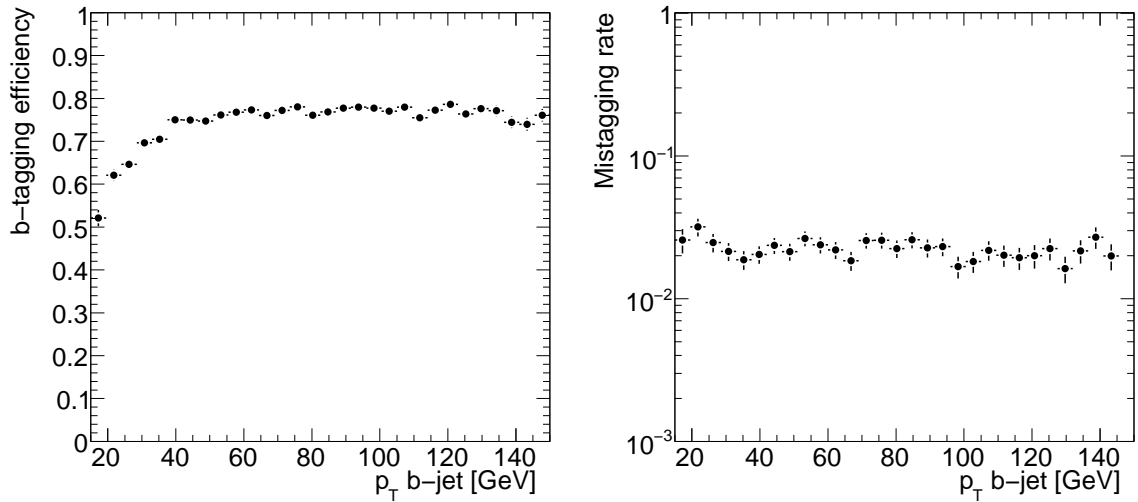


Figure 8.17: b -tagging efficiency and b -tagging mistag rate as a function of jet p_T , corresponding to a cut on the b -tagging weight of $\log\left(\frac{P_b}{P_l}\right) > -1$.

In order to understand if the method performs well, it is convenient to check first separately whether it can correctly reproduce additional jets stemming from the b - and non b -

⁵No b -tagging is possible if the jet pseudo-rapidity is such that the tracks related to the jet are outside the inner detector acceptance of $|\eta| < 2.5$.

jet component. To increase the available statistics, for all intermediate checks no b -tagging requirement is applied on the subjects from the Higgs $\rightarrow b\bar{b}$ candidate, but a truth level requirement is applied, selecting only Higgs boson candidates corresponding to *true* b - c jet pair combinations.

In Fig. 8.18 (left) a comparison is made between the additional non b -jets in the event in the control sample (all additional jets excluding the b -jet used to select the leptonic top), compared to the highest p_T additional jet in the signal region which is not matched at truth level with a b -jet. The plot shows a reasonable agreement, and, most important, it also shows that the additional non b -jet distribution does not depend significantly on the p_T of the b -jet from the leptonic top decay, since the re-weighting based on the p_T of the b -jet (labelled as CS-corrected in the plot) does not have any visible effect on the distribution.

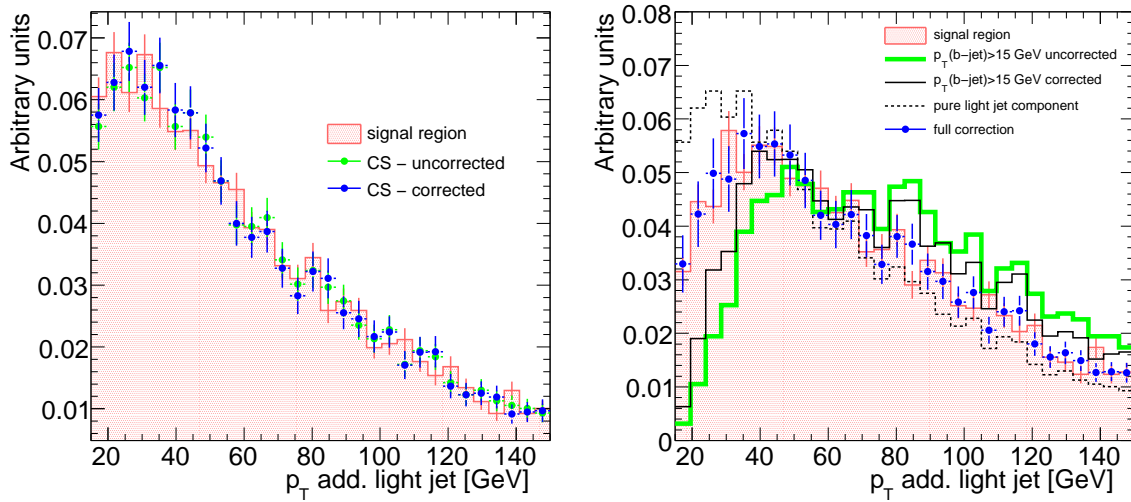


Figure 8.18: (Left) p_T distribution for additional non b -jets in the control sample and in the signal region. (Right) Contribution expected from non b -jets to the final p_T of the additional jet distribution in the control sample and in the signal region.

When combining two b -jet and non b -jet distributions in the control sample to obtain the inclusive p_T of the additional jet distribution expected in the signal region, one is interested in the non b -jet component only for the fraction where $p_T(\text{non } b\text{-jet}) > p_T(b\text{-jet})$. The related distribution one wants to reproduce in the signal region is shown in Fig. 8.18 (right) in red. According to the two-step procedure introduced above, the same contribution, as derived from the control sample, is also shown: it is labelled in the plot as $p_T(b\text{-jet}) > 15$ GeV *uncorrected* before the re-weighting procedure and as $p_T(b\text{-jet}) > 15$ GeV *corrected* afterwards. As already mentioned, in this way one misses the contribution expected from non b -jets in the signal region in events where the b -jet has a transverse momentum below 15 GeV or is outside the inner detector acceptance. The fraction of events where this happens is at the moment estimated by using Monte Carlo information and is used as an external, Monte Carlo dependent, input to the method: this fraction is found to be 19%. A relative error of 20% is assumed on this fraction, to be robust against the influence of systematic uncertainties like in particular shifts in the energy scale⁶. Since the non b -jet component has been found to be fairly independent from the

⁶Further studies to understand the influence of the different experimental and theoretical systematic uncertainties on this assumption would be certainly desirable: the assumed 20% relative error should only be seen as a first assumption. In the future it may be possible to get a handle on the relative fraction of b -jets

selected b -jet from the leptonic top, one can just add on top of the distribution obtained for $p_T(b\text{-jet}) > 15$ GeV, a 19% component with the pure additional non b -jet contribution already illustrated in Fig. 8.18 (left). The final result is shown again in Fig. 8.18 (right) in blue on top of the distribution expected in the signal region. In the limit of the available statistics, the agreement is good.

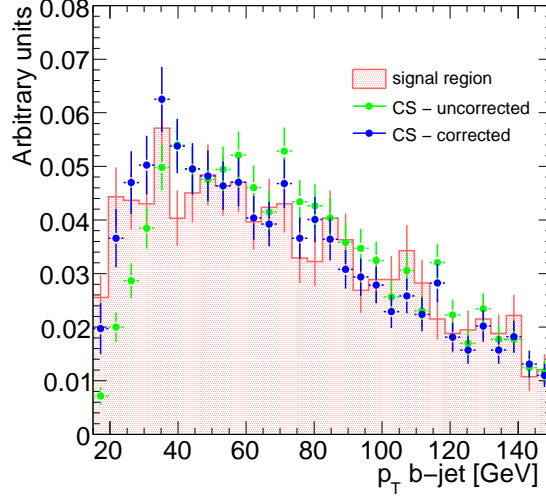


Figure 8.19: Contribution expected from b -jets to the final p_T of the additional jet distribution in the control sample and in the signal region.

In Fig. 8.19 the contribution to the p_T of the additional jet distribution deriving from b -jets is shown, in the control sample (considering the b -jet from the leptonic top) and in the signal region (using truth level information to select only the b -jet component). In this way only events where $p_T(b\text{-jet}) > p_T(\text{non } b\text{-jet})$ are selected. Since the p_T of the additional jet distribution is obtained only for values of p_T above 15 GeV, in this case there is no need to add any additional contribution, since, if $p_T(b\text{-jet}) < 15$ GeV, then, necessarily, either $p_T(b\text{-jet}) < p_T(\text{non } b\text{-jet})$, and thus a light-jet will contribute to the p_T of the additional jet distribution, or there will be no additional jet with p_T above 15 GeV at all. Due to the leptonic top mass requirement, it is not obvious that the kinematic configuration of the b -jets is not biased. The plot shows however that in the limit of the available statistics, the distribution from the control sample can still reproduce the distribution in the signal region reasonably well.

Combining the b - and non b -jet contributions, the distribution of Fig. 8.20 (left) is obtained. The same distribution is shown in Fig. 8.20 (right) applying the full analysis selection with real b -tagging applied. In the limit of the available statistics, the agreement is very good. In the plots, the result of a Kolmogoroff-Smirnov compatibility test between the distributions in the control sample and in the signal region is also shown.

8.3.3 Implementation in the likelihood fit

In this section two control samples were introduced so far for the $t\bar{t}$ background: the first control sample can be used to obtain the PDFs for the filtered mass distribution as a function

from the leptonic top outside the p_T or η acceptance by trying to extrapolate the p_T and η distributions to these non accessible kinematic regions.

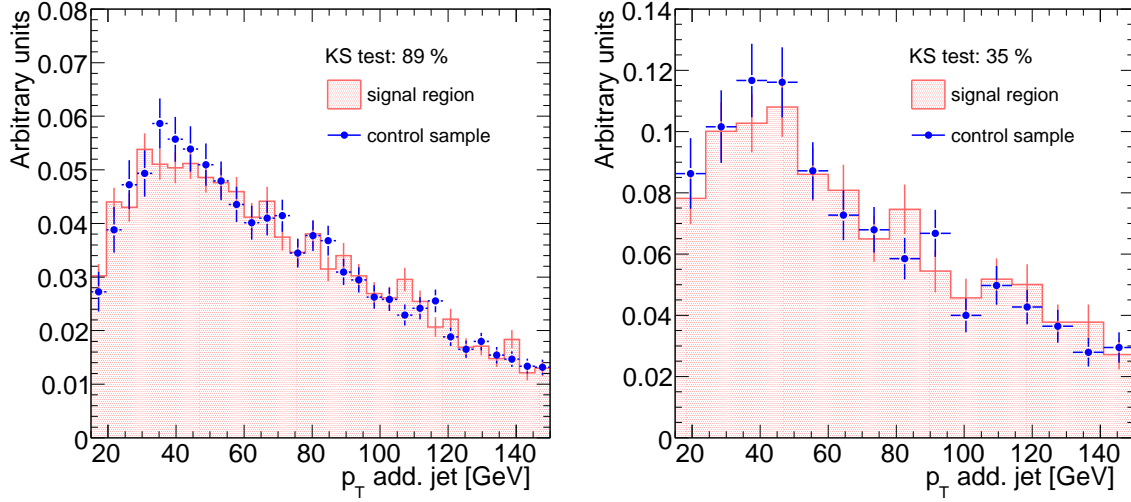


Figure 8.20: (Left) p_T distribution for additional jets as obtained from the control sample, compared to the same distribution in the signal region, for the case where truth level b -tagging is applied to increase the available statistics. (Right) p_T distribution for additional jets as obtained from the control sample, compared to the same distribution in the signal region, corresponding to applying the full analysis selection.

of $p_T(\text{Higgs})$ and the PDF for the b -weight, while the p_T of the additional jet distribution can be obtained from the second control sample.

The first control sample, since it is based on unweighted events, can be integrated directly into the maximum likelihood fit by adding a specific likelihood function for the events in the control sample, of the form:

$$L_{CS} = \Pi_{i=0}^{N_{tot,cs}} P_{t\bar{t}}(m_H, i | p_T, i) P_{t\bar{t}, CS}(p_T(H, i)) P_{t\bar{t}}(b \text{ weight}, i), \quad (8.15)$$

where the parameters for $P_{t\bar{t}}(m_{\text{filtered}, i} | p_T(\text{filtered}, i))$ and $P_{t\bar{t}}(b\text{-weight}(i))$ are constrained to be identical to the parameters for the $t\bar{t}$ background in the signal region, while the parameters for the PDF $P_{t\bar{t}, CS}(p_T(\text{filtered}, i))$ are left free to be determined on the control sample. Instead of first performing a maximum likelihood fit on the control sample to determine these common parameters and then later use them (keeping them fixed) in the nominal fit in the signal region, a simultaneous fit is performed on both control sample and on the signal region, maximising a global likelihood function defined as:

$$\log(L_{TOT}) = \log(L_{CS}) + \log(L_{SR}), \quad (8.16)$$

where CS is the control sample and SR the signal region. In the simultaneous fit the parameters for the PDFs in the signal region are left free to vary, but constrained by the knowledge provided by the simultaneous fit on the control sample. In this way the systematic uncertainties connected to the precision by which the PDFs are determined on the control sample are propagated to the fit in the signal region and are therefore reflected in the statistical uncertainty of the parameters determined in the nominal fit.

The second control sample, used for the determination of the $p_T(\text{additional jet})$ distribution, is more difficult to introduce directly into the fit, because the events used to obtain such a distribution are weighted and weighted events are difficult to feed into an unbinned maximum likelihood fit in a straightforward way. Therefore the parameters for the PDF based on $p_T(\text{additional jet})$ are simply first determined on the control sample and then later used in

the nominal fit. In addition, when considering also the remaining systematic uncertainties, the uncertainties corresponding to the error bars shown in Fig. 8.20 (right distribution) are also propagated to the fit in the signal region.

8.3.4 Control samples for the Wt background

For the Wt background, the definition of a specific control sample appears to be more difficult. However, the Higgs boson candidate mass distribution in the Wt background is expected to be identical to the corresponding one in the $t\bar{t}$ background, since in both cases a high p_T top quark is mimicking the $b\bar{b}$ system representing the Higgs boson candidate. The same is true for the b -weight distribution.

In order to verify these statements, analogously to what was already done for the $t\bar{t}$ background, the mass distribution in the Wt background is compared to the mass distribution as obtained from the first $t\bar{t}$ control sample: this is shown in Fig. 8.21, after applying the full analysis selection, and in Figs. 8.22 and 8.23, after selecting only b - c jet combinations and b -light jet combinations, respectively, through the use of truth level information. For both the b - c and b -light jet components the agreement is very good. The same comparison is shown for the b -weight distribution in Fig. 8.25: applying b -tagging at truth level and selecting separately b - c and b -light jet combinations, the agreement is perfect, while, after applying b -tagging at reconstruction level, some discrepancy is seen, with a more pronounced tail at high b -weight values in the Wt background with respect to the $t\bar{t}$ control sample.

The reason is that, while the relative amount of b - c and b -light jet combinations is compatible in the $t\bar{t}$ and Wt backgrounds, in around 10% of the Wt events, at truth level the b -jet label is assigned to both the subjects representing the decay products of the Higgs boson candidate, while for the $t\bar{t}$ background this component is negligible. The origin of this effect may not be motivated by a real different physical behaviour, but could be caused by the different Monte Carlo model used to generate the Wt and $t\bar{t}$ events, which are produced, respectively, with the the AcerMC and HERWIG Monte Carlo generators, the first relying on the parton shower and hadronisation model as implemented in PYTHIA. The different shower and hadronisation models implemented in HERWIG and PYTHIA could provide subtle differences in the way gluon splitting and jet shapes are modelled and eventually explain a higher rate of b - b jet combinations entering the Higgs boson selection. This will be cross-checked in the future by regenerating the Wt background interfacing AcerMC with the HERWIG Monte Carlo generator. For the moment, since there is no striking reason for the Wt and $t\bar{t}$ backgrounds to behave differently in this respect, it is reasonable to temporarily assume that the disagreement in the b - b jet component is due to this Monte Carlo dependence and will not appear in nature. Fig. 8.24 shows also a comparison of the η distributions of Wt in the signal region and $t\bar{t}$ in the control sample: there is some difference in the tails, but the distributions are still reasonably compatible.

The PDFs for the Wt background involving the conditional dependence of the Higgs boson candidate mass on $p_T(\text{Higgs})$ and the b -weight are therefore derived as in the case of the $t\bar{t}$ background, by just constraining their parameters to be identical to the corresponding parameters of the same PDFs defined for the $t\bar{t}$ control sample. At the same time no strategy has been found yet to get the $p_T(\text{additional jet})$ distribution for the Wt background from data, since the method developed for the $t\bar{t}$ background does not apply to the Wt background.

8.3.5 Control samples for the W +jet background

Defining a control sample for the W +jet background seems to be a very difficult task, since the irreducible W +2 b -jets component which is dominating this background is considerably

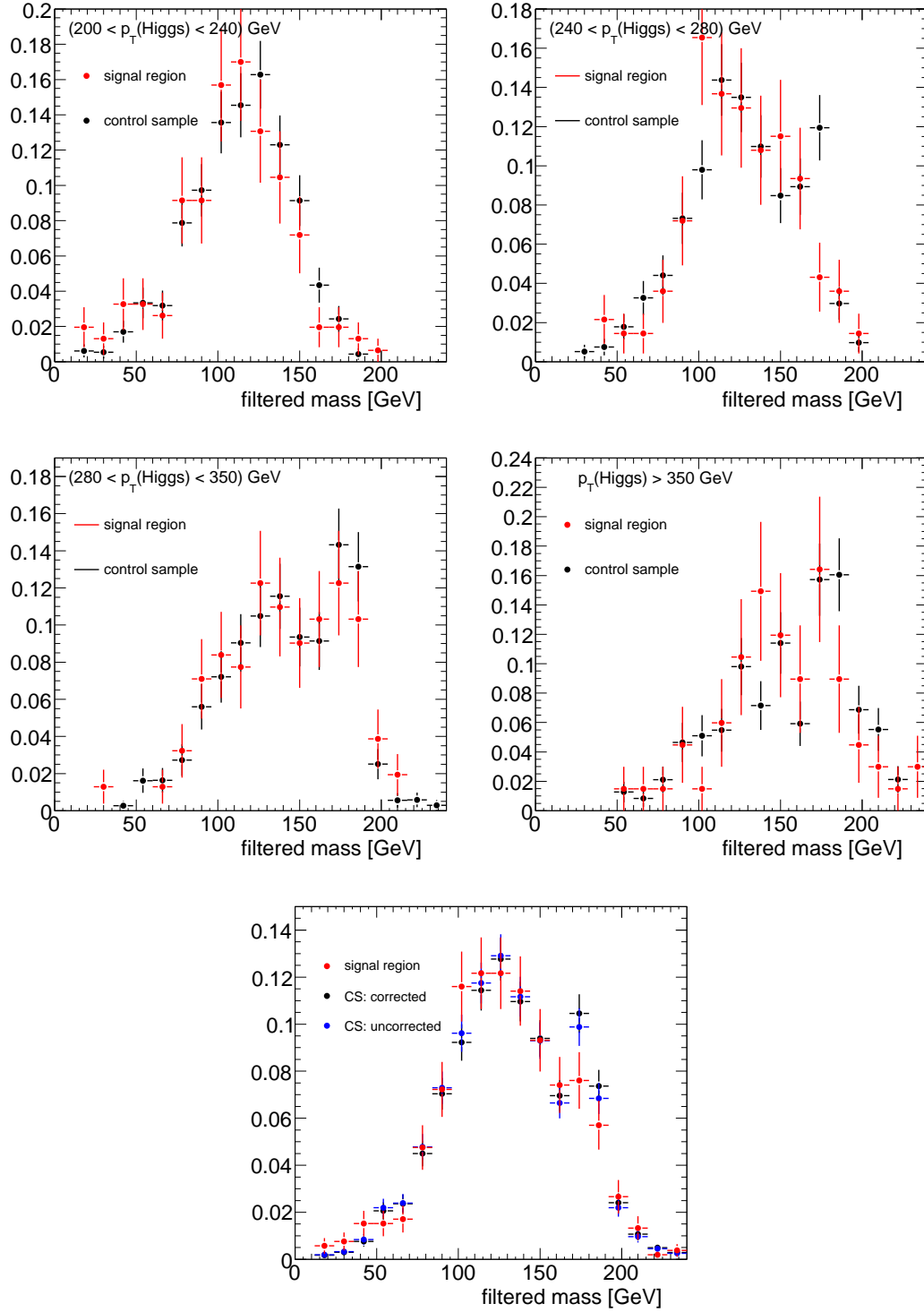


Figure 8.21: Distributions of the filtered mass for the Wt background for various $p_T(\text{Higgs})$ intervals, as obtained from the signal region, compared to the same distribution obtained in the $t\bar{t}$ control sample. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

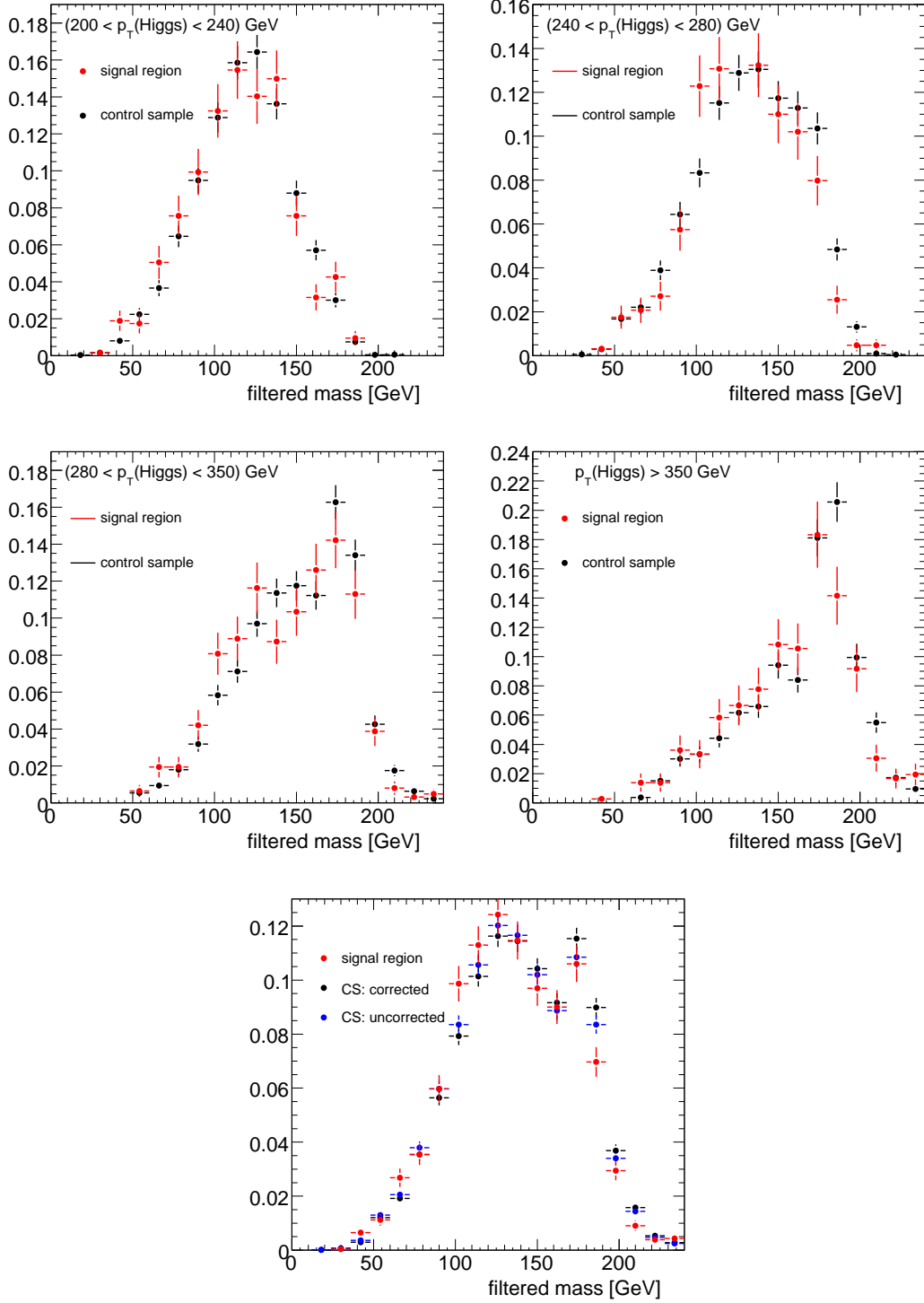


Figure 8.22: Distributions of the filtered mass for the Wt background for various $p_T(\text{Higgs})$ intervals, as obtained from the signal region, compared to the same distribution obtained in the $t\bar{t}$ control sample, selecting in both cases the b - c jet component through the use of truth level information. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

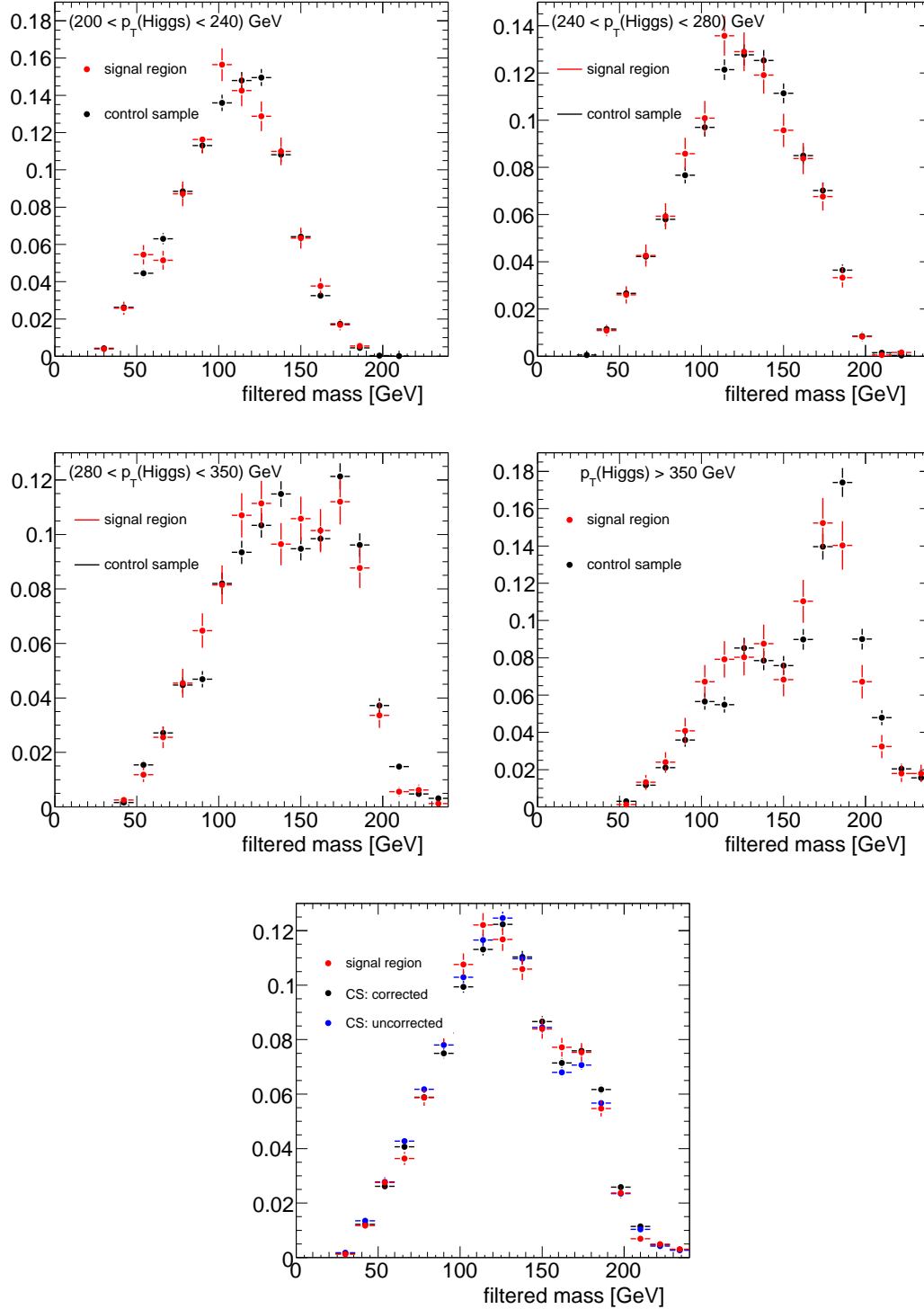


Figure 8.23: Distributions of the filtered mass for the Wt background for various $p_T(\text{Higgs})$ intervals, as obtained from the signal region, compared to the same distribution obtained in the $t\bar{t}$ control sample, selecting in both cases the b -light jet component through the use of truth level information. In addition the inclusive distribution is shown on the bottom, before and after the $p_T(\text{Higgs})$ dependent correction. The distributions are normalized to an area of one.

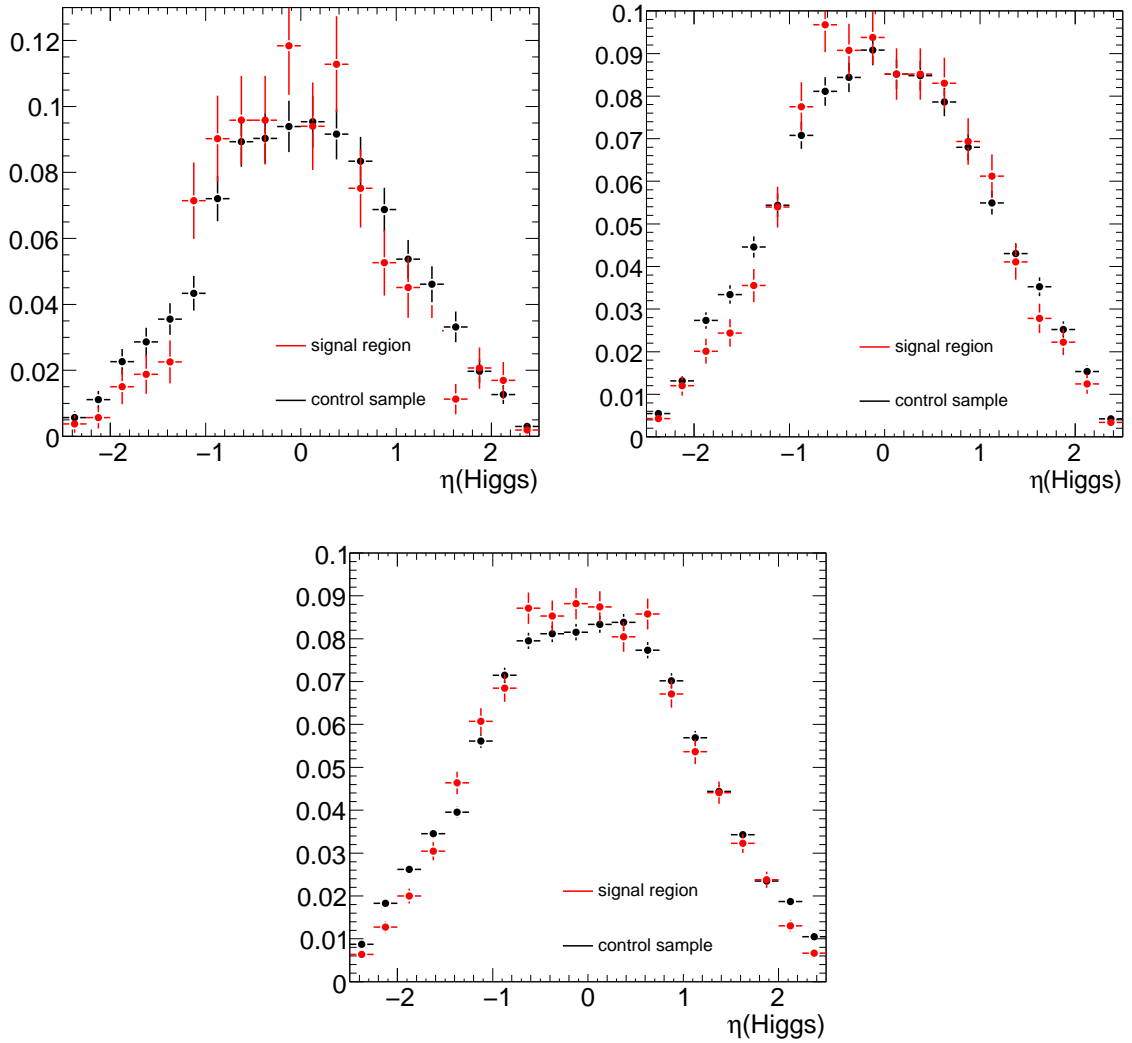


Figure 8.24: Pseudo-rapidity distribution of the Higgs to $b\bar{b}$ candidate in the Wt background, on the left for the full selection, on the right after selecting only the b - c quark pair component and on the bottom after selecting only the b -light quark pair component using truth level information, compared to the same distributions as obtained from the $t\bar{t}$ control sample. The distributions are normalized to an area of one.

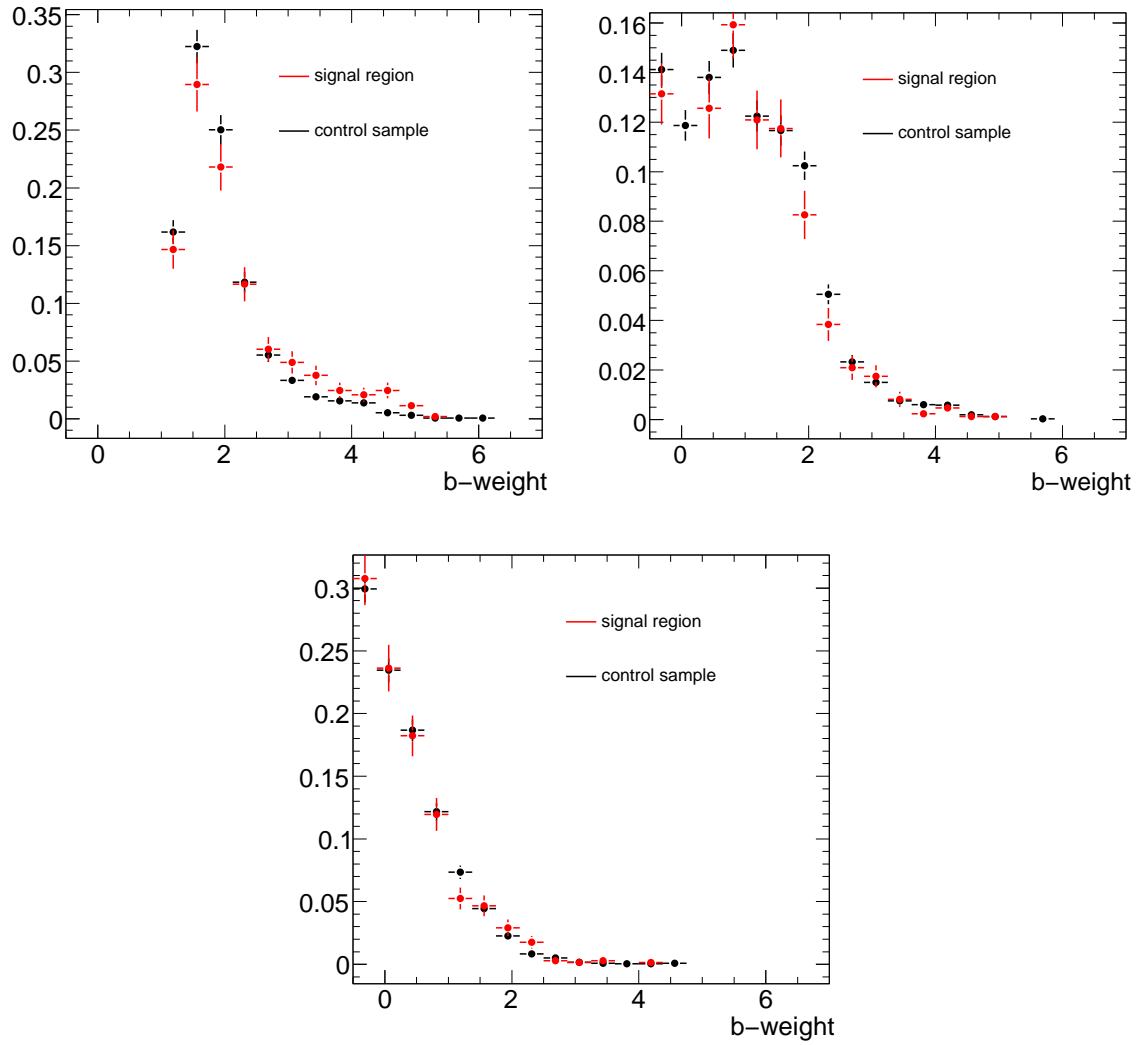


Figure 8.25: b -weight distribution for the Higgs to $b\bar{b}$ candidate in the Wt background, on the left for the full selection, on the right after selecting only the b - c quark pair component and on the bottom after selecting only the b -light quark pair component using truth level information, compared to the same distributions as obtained from the $t\bar{t}$ control sample. The distributions are normalized to an area of one.

different from the $W+2$ light-jets component, which could in principle be obtained from data by reverting the b -tagging requirements on the two Higgs subjects and separating the events selected in this way from the $t\bar{t}$ background in this region of phase space by the use of a separate dedicated maximum likelihood fit. No such a control sample is thus defined in this study. Getting more control over this background contribution and on the uncertainties connected with the estimation of the PDFs by using purely Monte Carlo information will need to rely in the future both on more detailed theoretical studies and on data versus Monte Carlo cross-checks in regions of the phase space where the signal contribution is expected to be negligible.

8.4 Fit validation on mixed Monte Carlo sample

The main aim of the fit is to extract the number of signal and overall background events, in order to define a statistical hypothesis test about the eventual presence of a Higgs boson signal. However, the extraction of the number of events of the single different background contributions is crucial to make the fit insensitive to any assumption about the expected number of events for the single different background contributions. The distributions of the discriminating variables of the signal and single backgrounds are sufficiently different, so that the fit turns out to be able to extract separately the single yields for all these contributions: WH , WZ , $t\bar{t}$, Wt and $W + \text{jet}$. As an exception, the Wt and $t\bar{t}$ backgrounds have very similar shapes for nearly all discriminating variables (the mass and b -weight PDFs are even assumed to be identical in the fit). In this case, in order to obtain a clear maximum of the likelihood function in the space of the yield parameters and to stabilise the fit, a Gaussian constraint is applied on their relative expected number of events:

$$R_{top} = \frac{\hat{N}_{Wt}}{\hat{N}_{Wt} + \hat{N}_{t\bar{t}}} = (22.2 \pm 2.5)\%, \quad (8.17)$$

where \hat{N}_{Wt} and $\hat{N}_{t\bar{t}}$ are the yields as determined from Monte Carlo. This is achieved by adding an additional term to the log likelihood function defined in Eq. 8.16:

$$\log(L_{TOT}) = \log(L_{CS}) + \log(L_{SR}) + \log\left(L\left(\frac{N_{Wt}}{N_{Wt} + N_{t\bar{t}}}\right)\right), \quad (8.18)$$

where the last term contains the Gaussian constraint:

$$L\left(\frac{N_{Wt}}{N_{Wt} + N_{t\bar{t}}}\right) = \exp\left[-\frac{\left(\frac{N_{Wt}}{N_{Wt} + N_{t\bar{t}}} - R_{top}\right)^2}{2\sigma_{R_{top}}^2}\right]. \quad (8.19)$$

The uncertainty $\sigma_{R_{top}}$ used for the ratio reflects the systematic error on the expected fractions of Wt and $t\bar{t}$ events, including both experimental and theoretical uncertainties. The latter include parton density function uncertainties and the effect of QCD renormalisation and factorisation scale variations as derived from the NLO calculation of these processes.

The test statistics, which is used as an estimator of how likely a signal contribution is present in the data sample on which the fit is applied, is based on the profile likelihood ratio defined as:

$$PLR = \frac{L(\hat{N}_{sig}, \hat{\vec{\alpha}})}{L(N_{sig} = 0, \hat{\vec{\alpha}})}, \quad (8.20)$$

which is the value of the likelihood function after it is maximised in the space of all parameters $\vec{\alpha}$ and yields determined in the fit, including the signal yield, divided by the same likelihood function, but in this case maximised with the constraint $N_{sig} = 0$. The parameters $\vec{\alpha}$ at this stage include the background yields and the parameters of the PDFs which are determined on the $t\bar{t}$ control sample, corresponding to the likelihood term of Eq. 8.15, while the parameters of the remaining PDFs (including all PDFs for the signal and $Wb\bar{b}$ background) are left fixed to the values determined on Monte Carlo: their variation will be considered in Section 8.6 to estimate the systematic uncertainty connected with shape variations of these PDFs.

As a first step, the fit is validated directly on the available Monte Carlo samples. All events passing the analysis selection in all Monte Carlo signal and background samples, including

the $t\bar{t}$ control sample, are merged into a single *mixed* data sample, with their respective contribution re-weighted according to the number of events expected in 30 fb^{-1} of collected data. Four different data sample are defined: with no Higgs contribution and with the addition of signal events corresponding to three different values of the Higgs boson mass: 115, 120 and 130 GeV. Three likelihood fits are performed, with the PDFs for the signal corresponding to the hypotheses of Higgs boson masses of 115, 120 and 130 GeV and by applying the fit on the data samples with the signal of corresponding mass. The three likelihood fits are then repeated on the data samples with no signal contribution.

There are two main advantages in this validation procedure with respect to using pseudo-experiments based on the shapes of the PDFs previously determined on the Monte Carlo samples, as will be done in a later step:

- It is not assumed that the distributions of the discriminating variables used in the fit are correctly parametrised in the likelihood function.
- It is not assumed that all discriminating variables, except for the dependence of the Higgs boson candidate mass from the Higgs boson transverse momentum, are uncorrelated.

As a drawback, due to the limited statistics of the Monte Carlo sample, only a single fit can be performed, so it is not possible to study in detail the statistical properties of the fit, which require a large ensemble of experiments.

The results for the three mass points of 115, 120 and 130 GeV are shown in Tables 8.8-8.10, respectively. The number of signal events is constrained to have non-negative values. If not specified the errors on the fitted parameters are computed using *HESSE*, which uses the second order expansion of the likelihood function around the found minimum; as an exception, for the signal yield *MINOS* is used, which computes the errors by doing a scan of the likelihood function over N_{sig} , maximising the likelihood function for every tested value of N_{sig} in the space of all other fit parameters, and provides in addition asymmetric errors.

Even if the fit is applied on weighted events, since an unbinned maximum likelihood fit is used, the fit errors do not reflect the statistics available in the Monte Carlo samples, but the statistics of the samples after rescaling. The errors therefore give a good indication of the average uncertainties expected when applying the fit after collecting an amount of data corresponding to an integrated luminosity of 30 fb^{-1} . Due to the use of the rescaling procedure, the number of signal and background events present in the samples is fixed at exactly the mean Poisson expectation value: if the fit works correctly, the yields provided by the fit should therefore be very close to their expectation values in Monte Carlo. This is, indeed, the case. In addition, the fit, in all mass hypotheses considered, correctly recognises the absence of any signal contribution.

Yield	Monte Carlo	Result without signal	Result with signal
N_{sig}	46	0^{+9}_{-0}	45^{+17}_{-16}
N_{WZ}	54	61 ± 18	61 ± 18
$N_{t\bar{t}}$	573	565 ± 29	565 ± 29
N_{Wt}	156	158 ± 18	158 ± 18
N_{W+jet}	263	260 ± 26	262 ± 26
$N_{t\bar{t},CS}$	1005	1005 ± 31	1005 ± 31

Table 8.8: Results (expected / fitted event numbers) of a fit on the mixed Monte Carlo sample with and without signal, for $m_H = 120 \text{ GeV}$.

Yield	Monte Carlo	Result without signal	Result with signal
N_{sig}	55	0^{+12}_{-0}	56^{+18}_{-17}
N_{WZ}	54	61 ± 16	60 ± 18
$N_{t\bar{t}}$	573	565 ± 29	565 ± 31
N_{Wt}	156	158 ± 19	158 ± 19
N_{W+jet}	263	260 ± 22	261 ± 26
$N_{t\bar{t},CS}$	1005	1005 ± 31	1005 ± 31

Table 8.9: Results (expected / fitted event numbers) of a fit on the mixed Monte Carlo sample with and without signal, for $m_H = 115$ GeV.

Yield	Monte Carlo	Result without signal	Result with signal
N_{sig}	33	0^{+12}_{-0}	27^{+16}_{-14}
N_{WZ}	54	61 ± 16	61 ± 16
$N_{t\bar{t}}$	573	565 ± 29	568 ± 31
N_{Wt}	156	158 ± 19	159 ± 20
N_{W+jet}	263	260 ± 22	263 ± 23
$N_{t\bar{t},CS}$	1005	1005 ± 31	1005 ± 31

Table 8.10: Results (expected / fitted event numbers) of a fit on the mixed Monte Carlo sample with and without signal, for $m_H = 130$ GeV.

The result of the fit on the mixed Monte Carlo samples can be more easily visualised by projecting the maximised likelihood function with the parameters fixed according to the fit result onto the single fit variables, as shown in Figs. 8.26, 8.27 and 8.28 for $m_H = 120, 115$ and 130 GeV. The WH signal, WZ , $W+jet$, $t\bar{t}$ and Wt likelihood components are also shown separately in a cumulative way: their normalisation reflects the yields of the likelihood function as determined during the fit. Together with the projections of the likelihood model, the distributions of the discriminating variables in the mixed Monte Carlo sample are shown, which in the absence of real data represent the *data points*.

Since the likelihood fit relies on more than one variable, it is not possible to deduce the real separation power between signal and background provided by the likelihood function by considering a single variable at a time. Therefore it is useful to look at the same variables, but taking first into account the discrimination power provided by the remaining variables; a way to accomplish this scope is to apply a cut on the likelihood ratio, defined as:

$$LR = \frac{L_{sig}^{oth}}{L_{sig}^{oth} + L_{bkg}^{oth}}, \quad (8.21)$$

where L^{oth} is defined as the likelihood function in which the dependence on the variable to plot has been integrated out. The likelihood function is defined in this case on an event basis, by removing from Eq. 8.2 the extended Poisson term and keeping only the term in the sum corresponding to the single event on which the likelihood function is computed.

The projection of the likelihood model obtained in this way can then again be compared with the distribution of the *data points* in the Monte Carlo mixed sample. The result is shown in Figs. 8.29, 8.30 and 8.31 for $m_H = 120, 115$ and 130 GeV. The signal-to-background ratio is significantly enhanced with respect to the inclusive plots shown in Figs. 8.26-8.28. The likelihood ratio cut has been chosen as a compromise between signal-to-background ratio and amount of events remaining in the mixed Monte Carlo sample after the cut is applied.

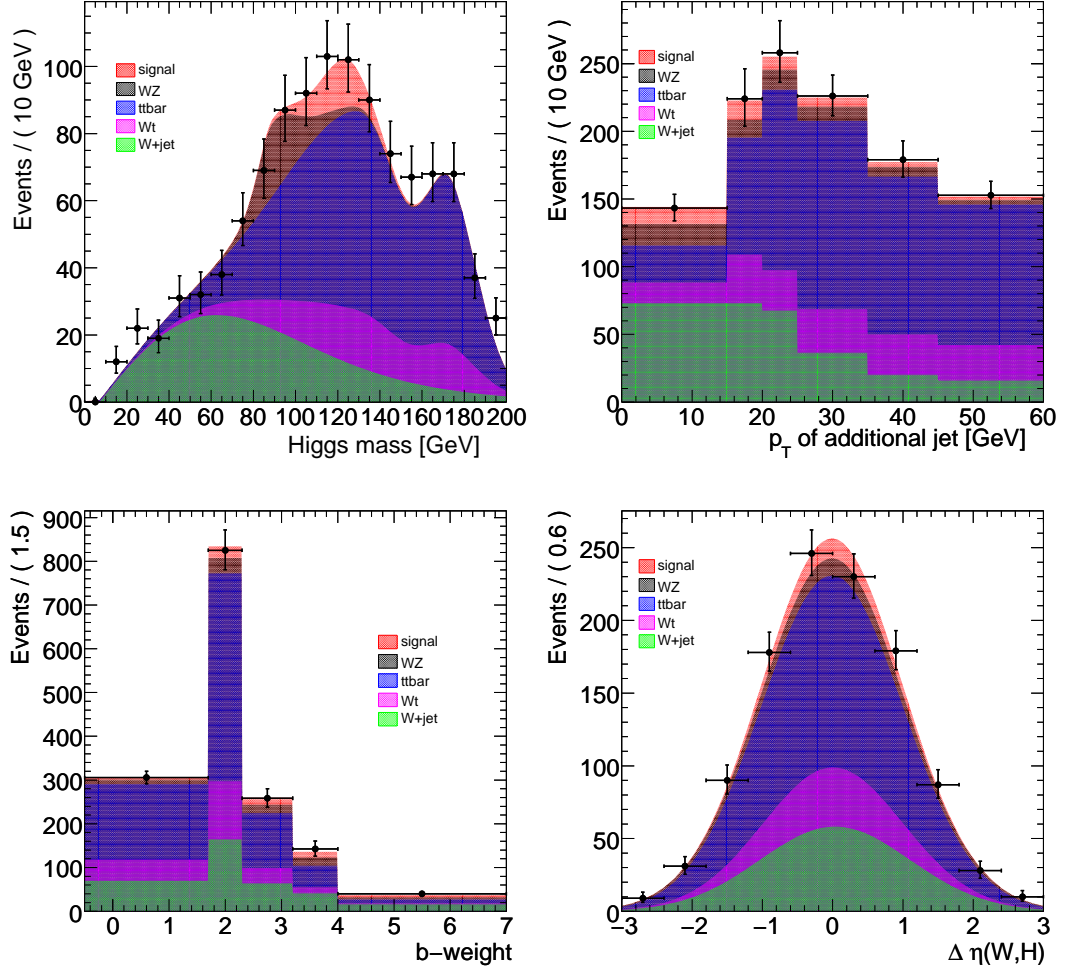


Figure 8.26: Projection of the multi-dimensional likelihood function on the single fit variables. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 120$ GeV are superimposed. The single fit components, scaled to the number of events as determined in the fit itself, are shown with different colours. An integrated luminosity of 30 fb^{-1} has been assumed.

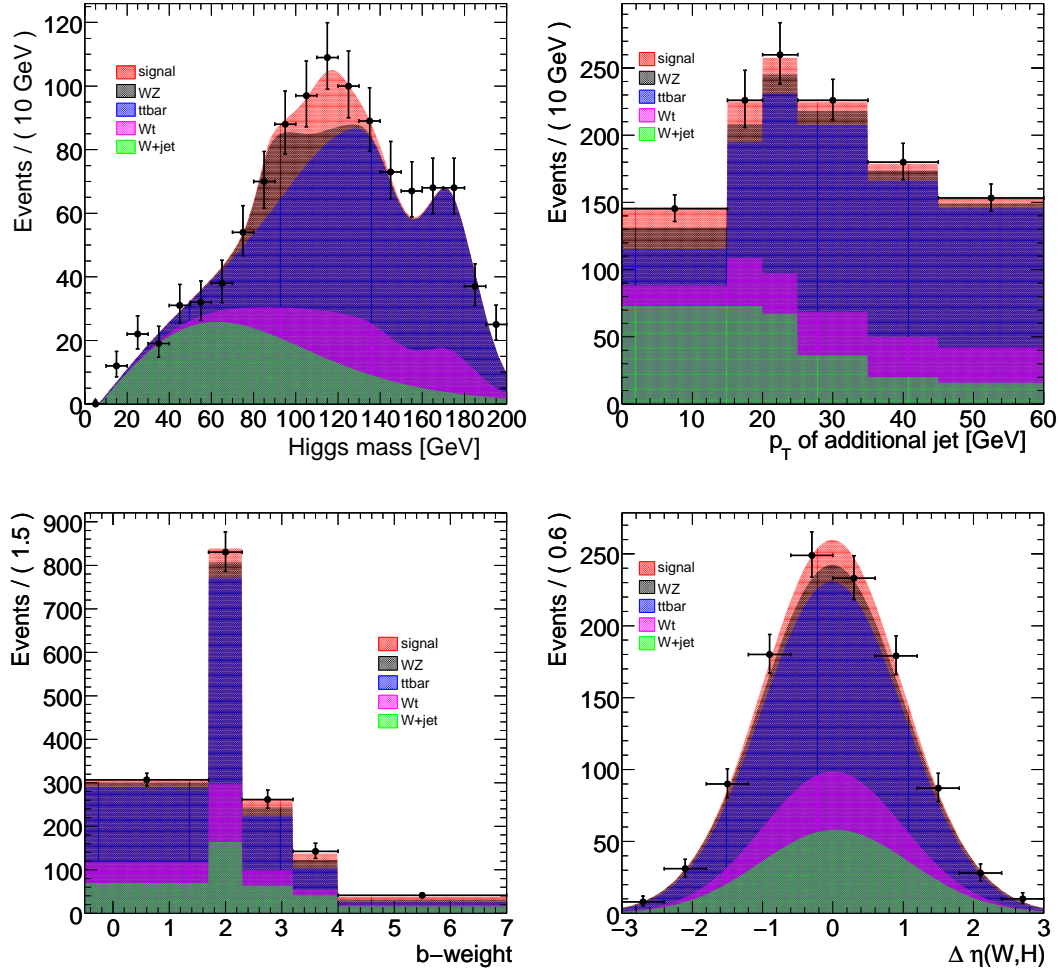


Figure 8.27: Projection of the multi-dimensional likelihood function on the single fit variables. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 115$ GeV are superimposed. The single fit components, scaled to the number of events as determined in the fit itself, are shown with different colours. An integrated luminosity of 30 fb^{-1} has been assumed.

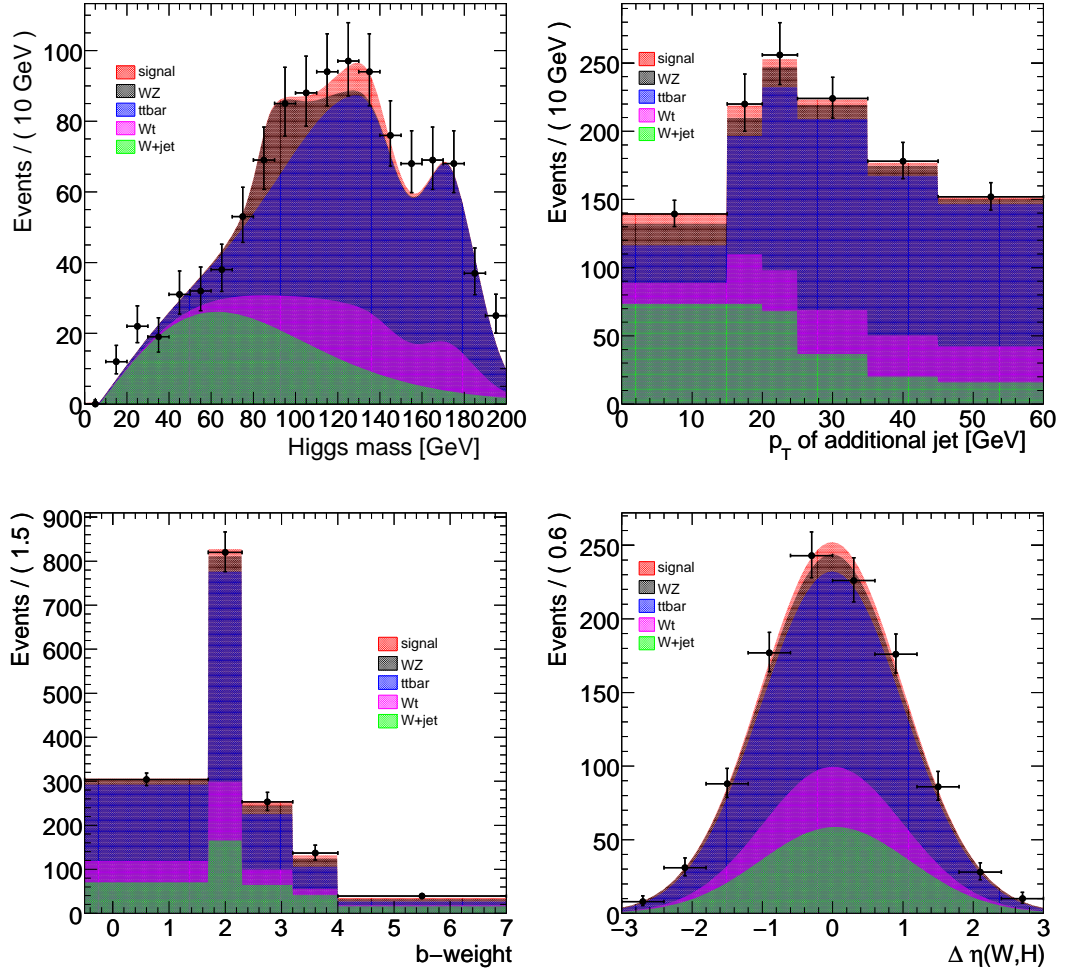


Figure 8.28: Projection of the multi-dimensional likelihood function on the single fit variables. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 130$ GeV are superimposed. The single fit components, scaled to the number of events as determined in the fit itself, are shown with different colours. An integrated luminosity of 30 fb^{-1} has been assumed.

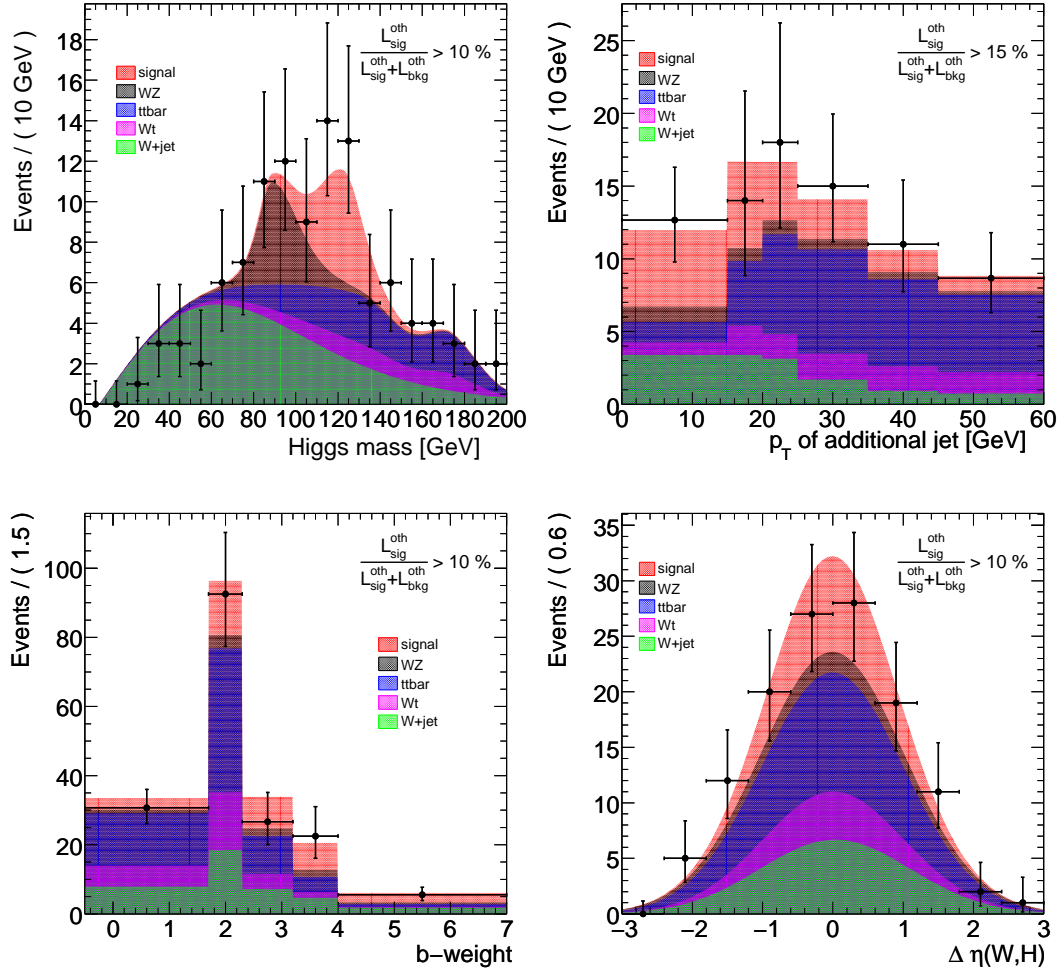


Figure 8.29: Projection of the multi-dimensional likelihood function on the single fit variables, after a likelihood ratio cut on the remaining variables at 10 (15) %. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 120$ GeV are superimposed. An integrated luminosity of 30 fb^{-1} has been assumed.

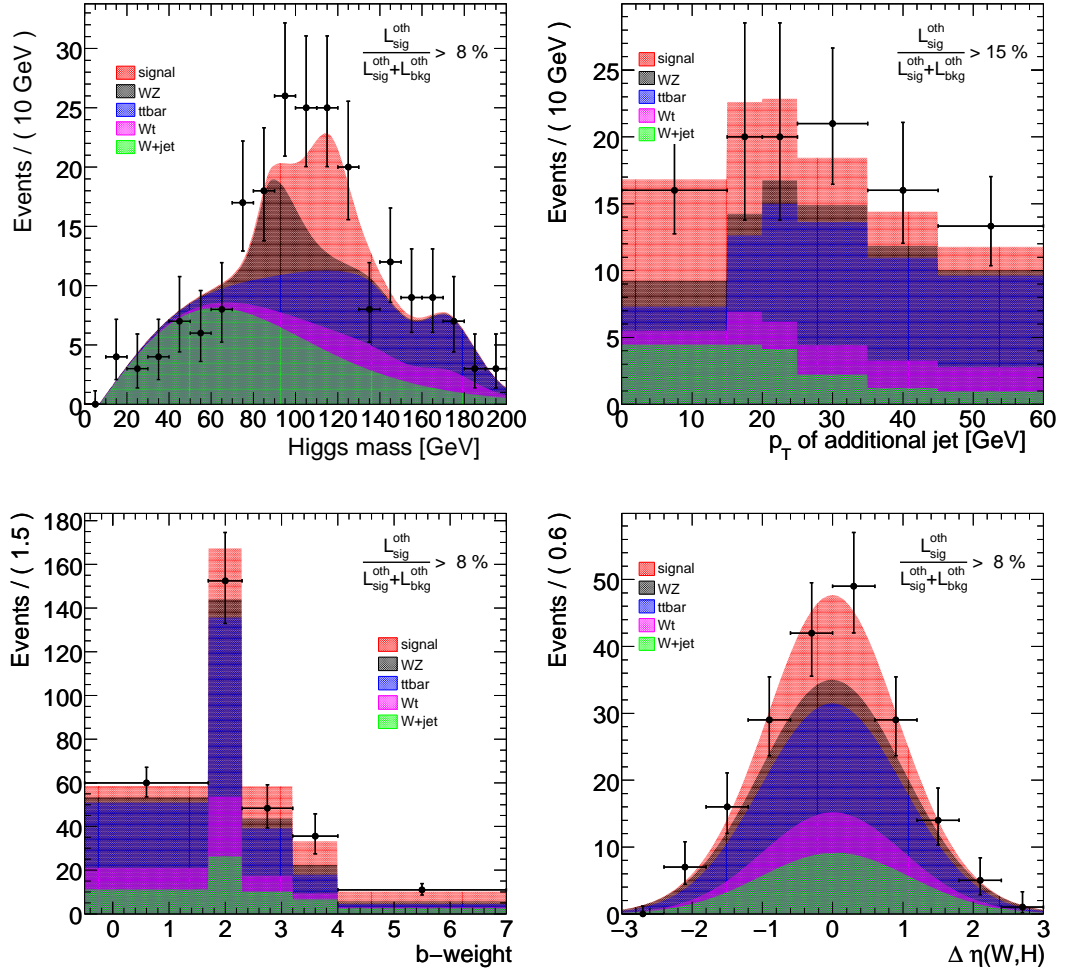


Figure 8.30: Projection of the multi-dimensional likelihood function on the single fit variables, after a likelihood ratio cut on the remaining variables at 8 (15) %. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 115$ GeV are superimposed. An integrated luminosity of 30 fb^{-1} has been assumed.

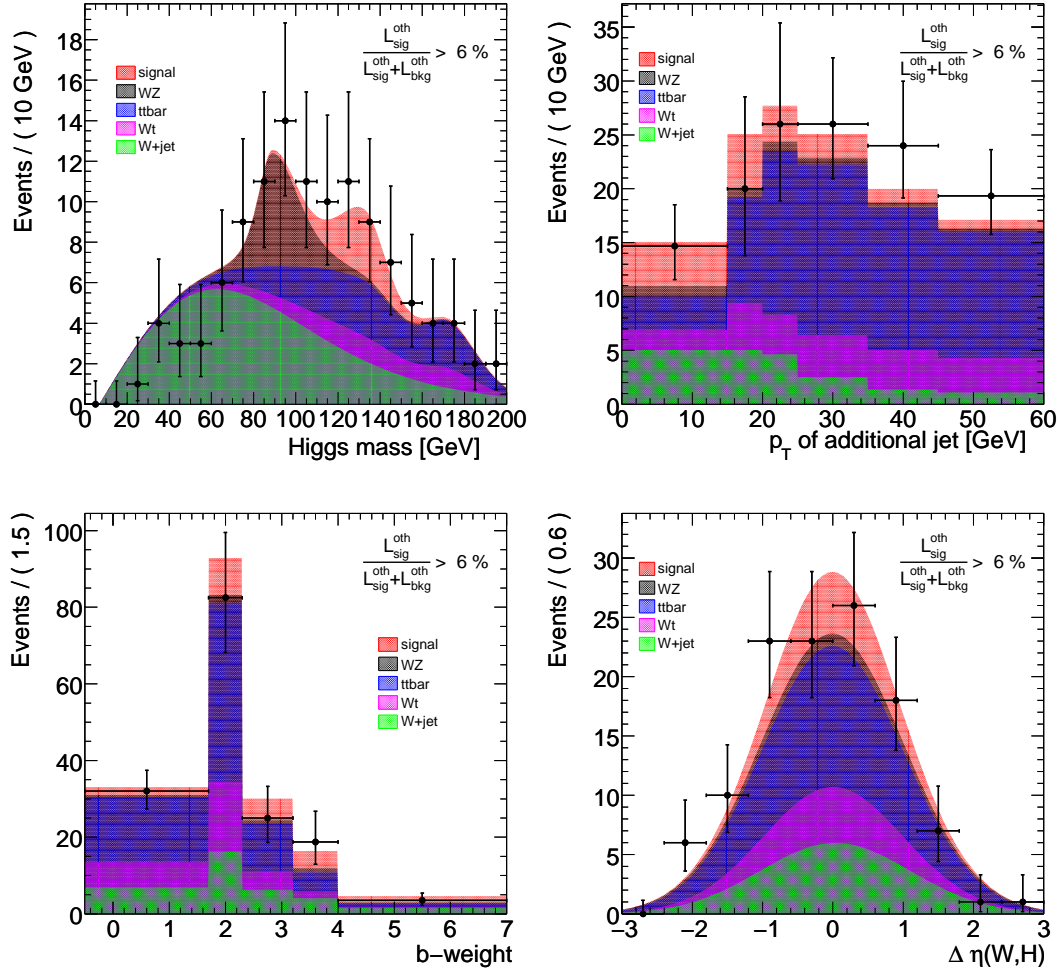


Figure 8.31: Projection of the multi-dimensional likelihood function on the single fit variables, after a likelihood ratio cut on the remaining variables at 6 %. The experimental points corresponding to the mixed Monte Carlo sample with $m_H = 130$ GeV are superimposed. An integrated luminosity of 30 fb^{-1} has been assumed.

8.5 Fit validation and discovery potential using Monte Carlo pseudo-experiments

8.5.1 Fit validation

In order to estimate the median discovery potential of the likelihood fit based analysis, a large set of pseudo-experiments was performed. The yields and the shapes of the PDFs entering the likelihood function, as previously estimated from the Monte Carlo samples, are assumed to provide a good approximation to the expected number of events and to the underlying distributions if an infinite amount of Monte Carlo statistics would be available. The number of signal and background events in the Monte Carlo pseudo-experiments is generated according to a Poisson distribution around the expectation values, while the PDFs are sampled by using the *hit and miss* Monte Carlo method, except for the PDFs represented by step and Gauss functions, where the sampling relies directly on the explicit expression for the cumulative function.

The two generation steps are slightly modified to account for two peculiar features of the likelihood function used in this analysis:

- The expectation values for the Wt and $t\bar{t}$ backgrounds for each pseudo-experiment are smeared first by a Gaussian uncertainty of 3 and 12%, respectively, according to the expected systematic uncertainty on their ratio.
- The conditional dependence of the Higgs mass PDF on $p_T(\text{Higgs})$ can be handled explicitly during generation, to avoid the need of sampling from a two-dimensional, non analytically integrable, PDF⁷. This problem can be solved by first generating the values for the $p_T(\text{Higgs})$ variable, by randomly sampling them from their related PDF, and then, in a second step, by generating the values for all remaining variables, replacing the the value of $p_T(\text{Higgs})$ in the Higgs mass PDF with the actual value of $p_T(\text{Higgs})$.

A first validation is performed by evaluating whether the fit can correctly reproduce, on a statistical basis, the expected number of events for the signal and for the single background contributions. Due to the presence of the extended term in the likelihood function, the fit provides directly an estimation of the expected number of events and includes the error originating from the associated Poisson fluctuations. This can be analysed by looking at the deviation of the fitted yield from the Poisson expectation value, divided by the estimated fit error (defined as *pull* distribution), separately for the various signal and background components. The result of this validation is presented in Figs. 8.32-8.37, separately for the three considered mass points. Together with the pull distribution for all signal and background components, also the distribution of fitted number of events is shown.

The likelihood fit behaves correctly for all three mass points: most of the pulls are well described by a Gauss function with mean 0 and width 1. Some exceptions can be seen for the WZ and signal components, in particular for example for the signal at $m_H = 130$ GeV. This is expected, because for these components the values for the yield is very low and the Poisson nature of the statistical fluctuations would require the computation of asymmetric errors in the fit by dropping the hessian approximation for the error estimation; this, however, does not affect the distribution of the test statistics, which is only based on the value of the likelihood function in the maximum. Also the $t\bar{t}$ and Wt backgrounds show a slightly non-Gaussian behaviour in their pulls: this effect is mainly due to the Gaussian constraint applied on the

⁷The analytical integration of the Higgs mass PDF has been explicitly carried out only over the mass variable.

ratio of their expectation values⁸. In addition, the W +jet background seems to have a small negative bias in the yield determined in the fit, but since this has no significant influence on the other signal or background components, this effect is considered as negligible.

8.5.2 Discovery potential

In order to complete the definition of the statistical hypothesis test, the distribution expected for the test statistics (the profile likelihood ratio, PLR , defined in Eq. 8.20) needs to be determined on a large set of pseudo-experiments with no signal contribution. By integrating the background only distribution of the PLR , it is in principle possible to calculate the p-value corresponding to a certain profile likelihood ratio value, i.e. the probability that, in the absence of a signal, the fit provides a value of the profile likelihood ratio equal or above the chosen value. However this would require to run more than 10^7 experiments to cover the region up to 5σ significance, which is unpractical. Instead, some sort of extrapolation to high values of the PLR is needed. Since in the Gaussian case, the distribution for $2 \cdot \log(PLR)$ is expected to follow a χ^2 distribution with one degree of freedom, according to the prescriptions adopted in Ref. [44], the following function is used:

$$f(x = 2 \cdot \log(PLR)) = c_1 \cdot \delta(x) + (1 - c_1) \cdot f_{\chi^2}(x), \quad (8.22)$$

determining the value for c_1 directly on the distribution obtained from background only Monte Carlo pseudo-experiments. By integrating over this function, it is then possible to estimate the p-values as a function of the PLR . The p-values can be translated into values of Gaussian significances by using the following simple transformation:

$$\text{significance} = \sqrt{2}\Phi^{-1}(2 \cdot \text{p value}), \quad (8.23)$$

where

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2} dt. \quad (8.24)$$

Once the statistical hypothesis test is defined, the likelihood fit can be performed on real data and will provide a single value of the PLR . The statistical significance related to the presence of a Higgs boson signal with a mass of 115, 120 or 130 GeV is then provided by the p-value corresponding to that value of PLR , as determined from the background only experiments. In the absence of real data, in order to obtain an estimate of the median statistical significance and therefore of the discovery potential of a Standard Model Higgs boson in this channel, a set of pseudo-experiments containing both signal and background is run and the distribution of the PLR in the presence of a signal is obtained. This distribution can be directly translated into the corresponding distribution of p-values and, therefore, into the distribution of discovery significances, from which the median discovery sensitivity and the 68% (1σ) and 95% confidence level (2σ) intervals can be estimated. It is worth noticing that, while the distribution of p-values is not sensitive to any assumption about the signal cross section and acceptance, the median expected statistical significance obviously is.

Both the distributions for the profile likelihood ratio for background only and signal plus background experiments (expressed in the plots as $2 \cdot \log(PLR)$ for convenience) are shown in Fig. 8.38 for a Higgs boson mass of 120 GeV, together with the expected distribution of signal significances for 30 fb^{-1} of integrated luminosity. The same is shown also for the other

⁸ A Gaussian constraint applied on a ratio of the form $A/(A+B)$ is clearly non-Gaussian with respect to both A and B .

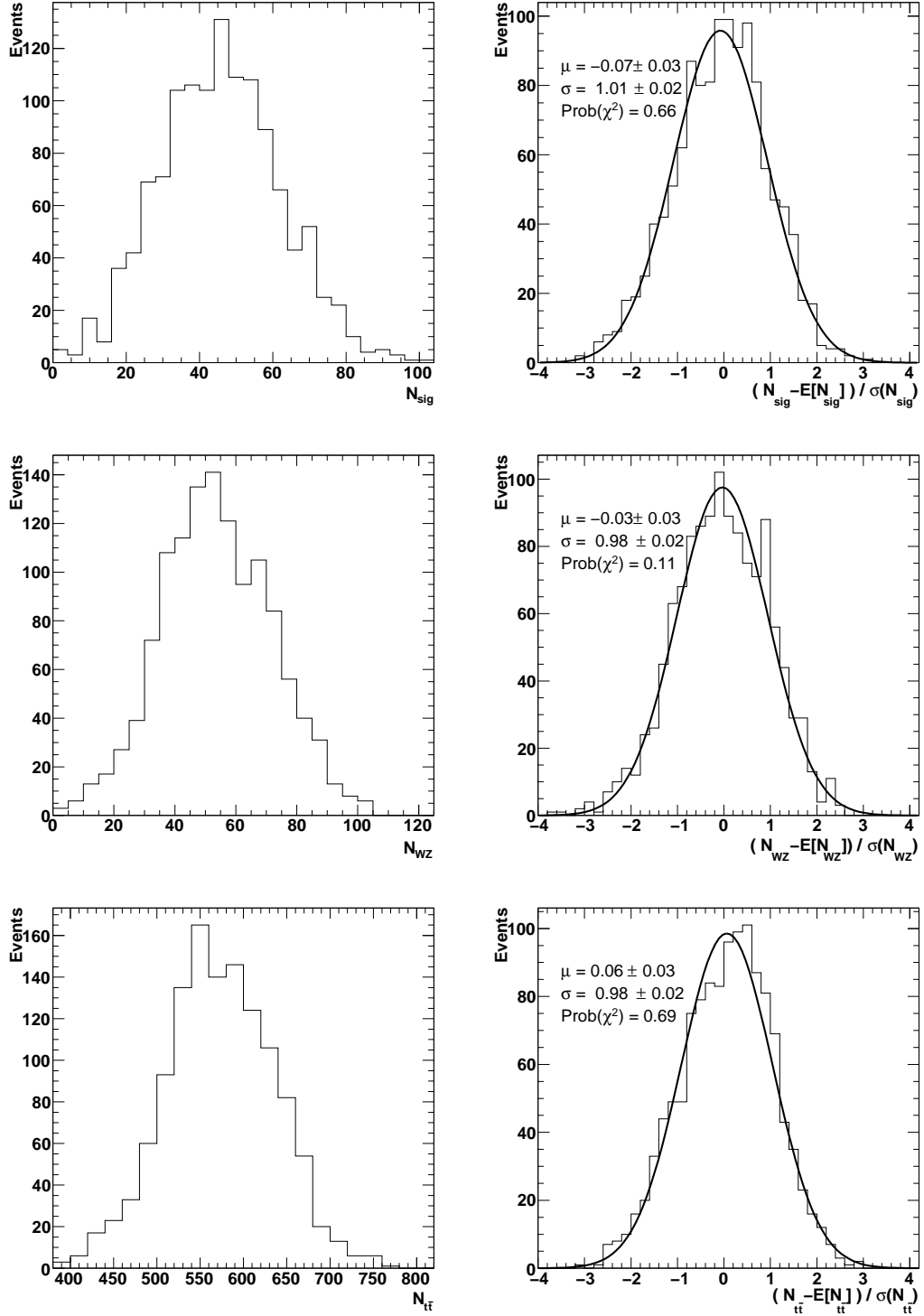


Figure 8.32: Fit validation for $m_H = 120$ GeV (I/II): the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

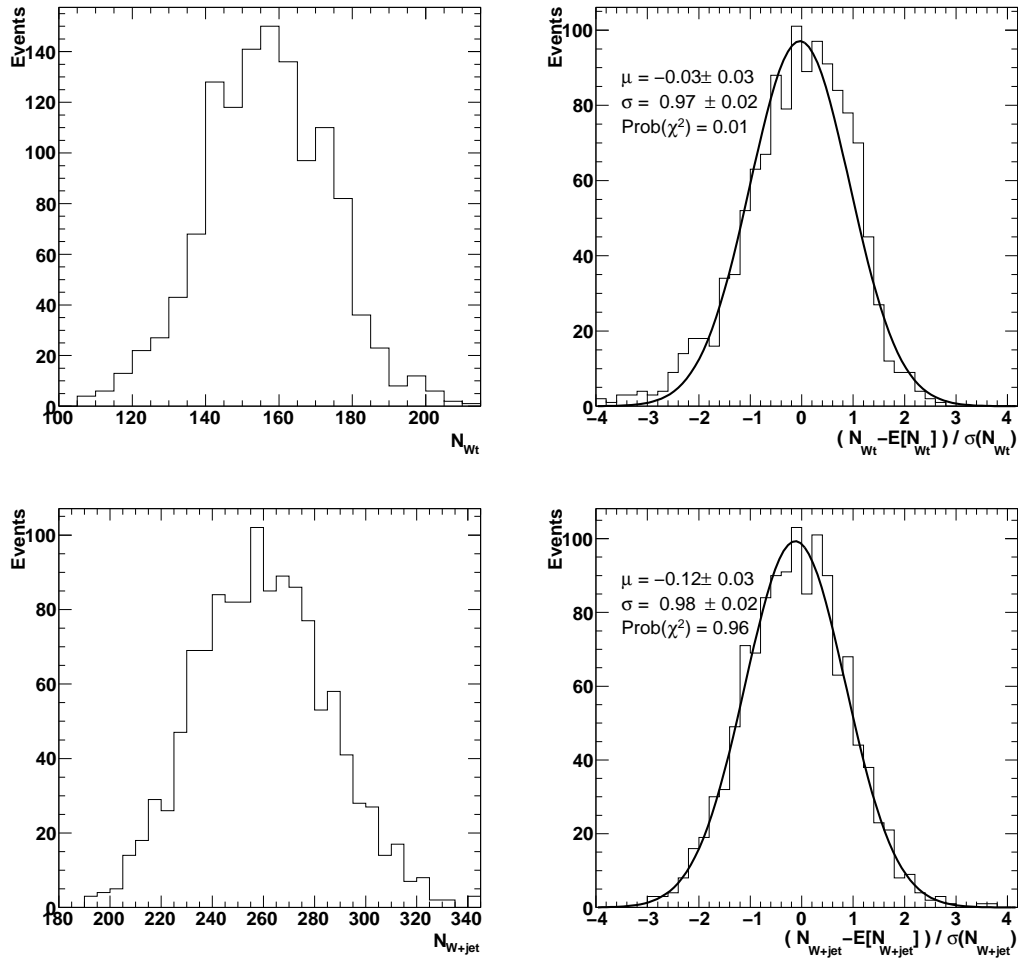


Figure 8.33: Fit validation for $m_H = 120$ GeV (II/II): the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

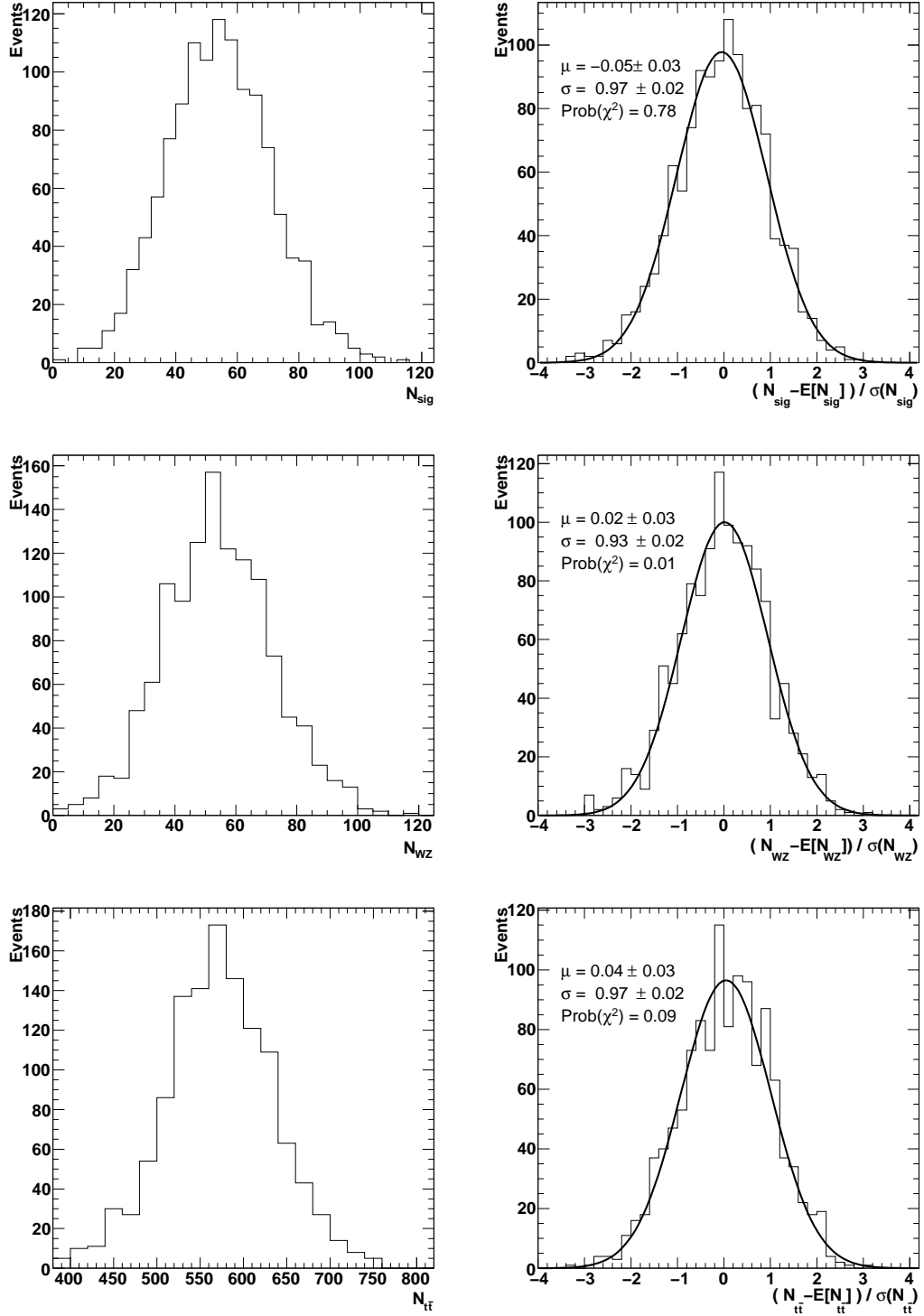


Figure 8.34: Fit validation for $m_H = 115$ GeV (I/II): the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

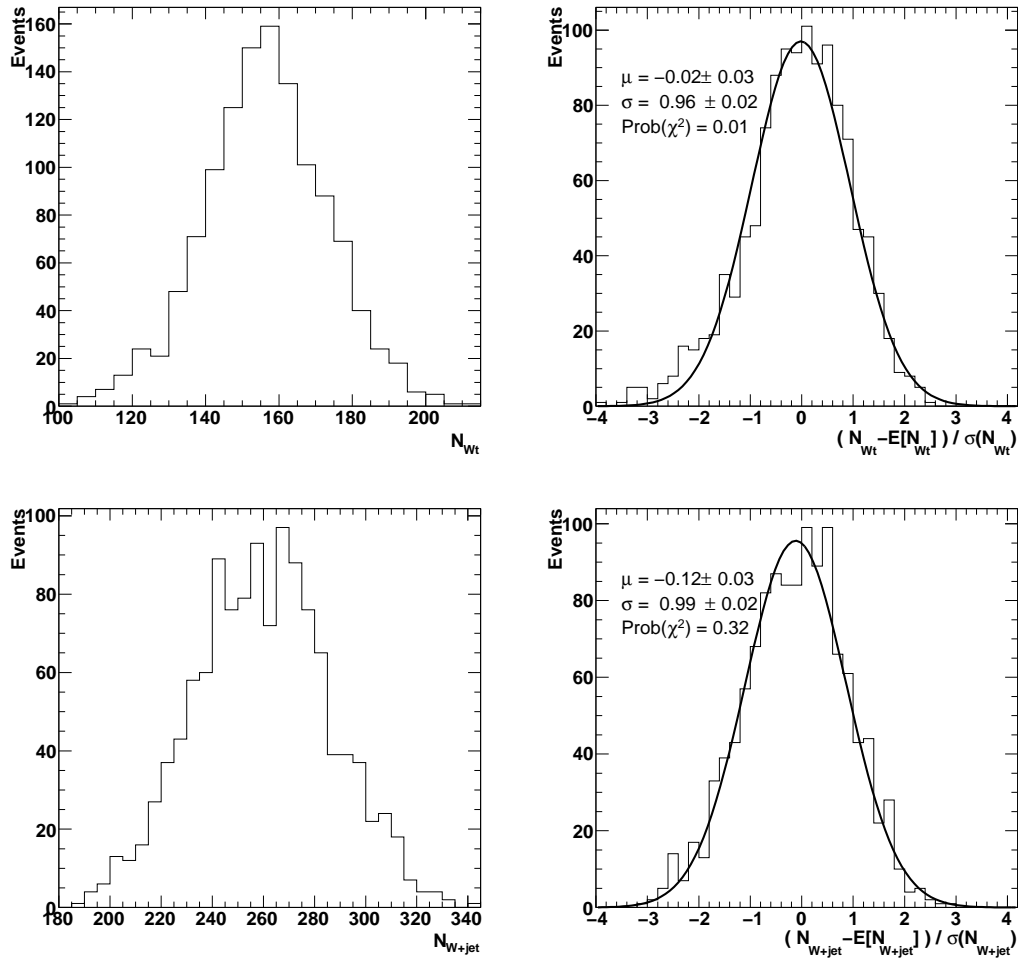


Figure 8.35: Fit validation for $m_H = 115$ GeV (II/II): the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

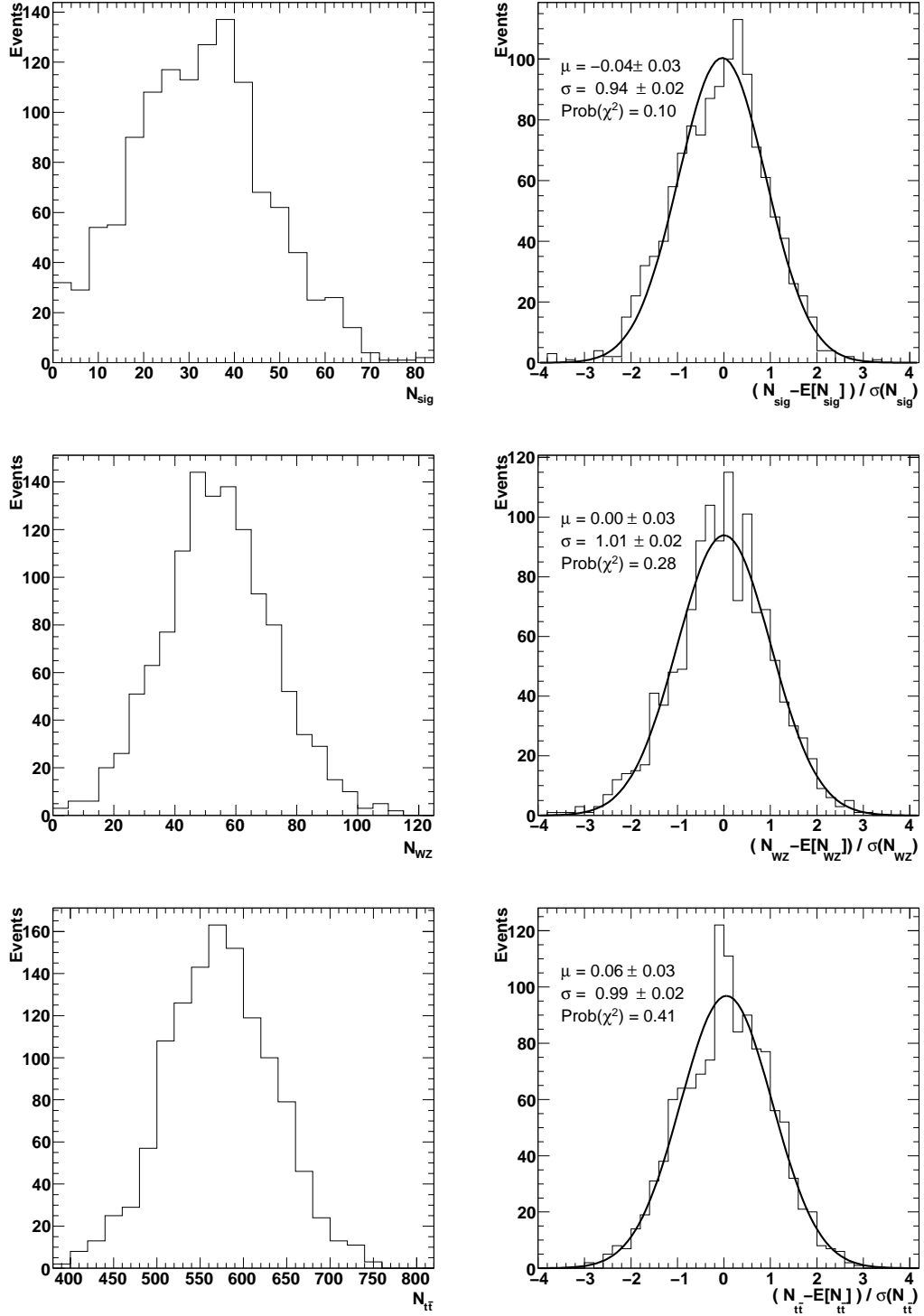


Figure 8.36: Fit validation for $m_H = 130$ GeV: the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

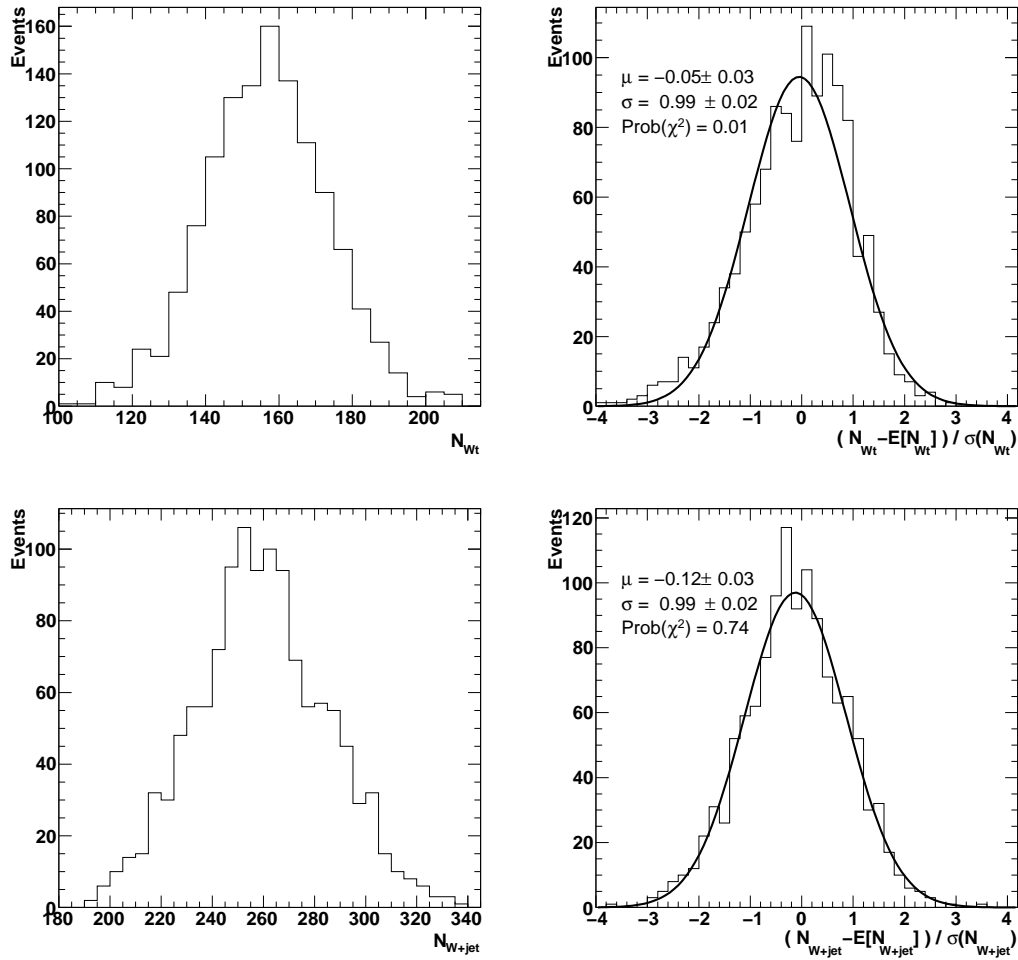


Figure 8.37: Fit validation for $m_H = 130$ GeV: the value for the single yields as determined on the Monte Carlo pseudo-experiments is shown, together with the pull with respect to their expectation value.

mass points in Figs. 8.39 and 8.40. The compatibility of the distribution of the PLR with the functional form assumed in Eq. 8.22 is cross-checked with a simple χ^2 test (reported in the plots) and found to be fairly good for all three mass points.

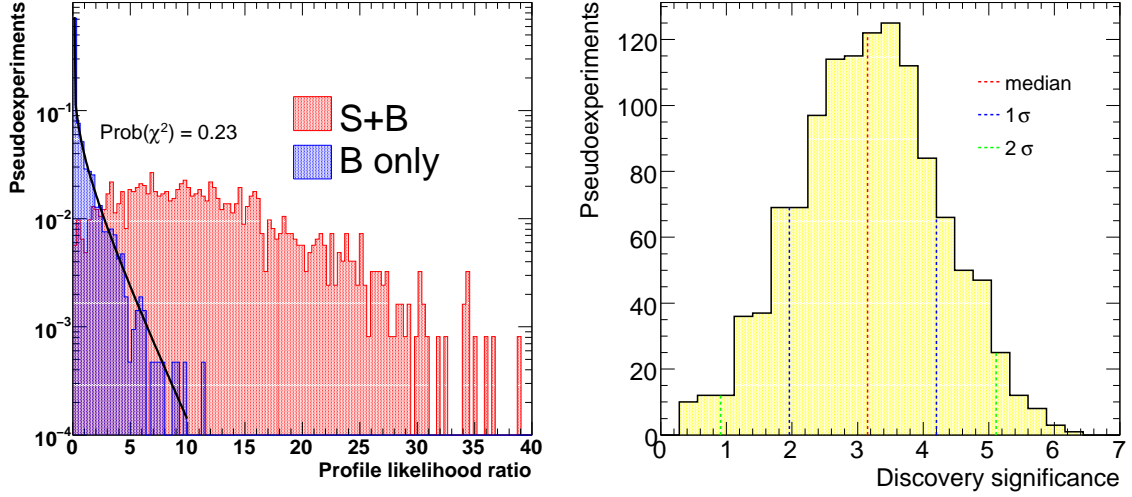


Figure 8.38: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to $m_H = 120$ GeV. The fit for the background only distribution is also shown. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 120$ GeV, with vertical lines corresponding to a confidence interval of 69% and 95% (1 and 2σ bands).

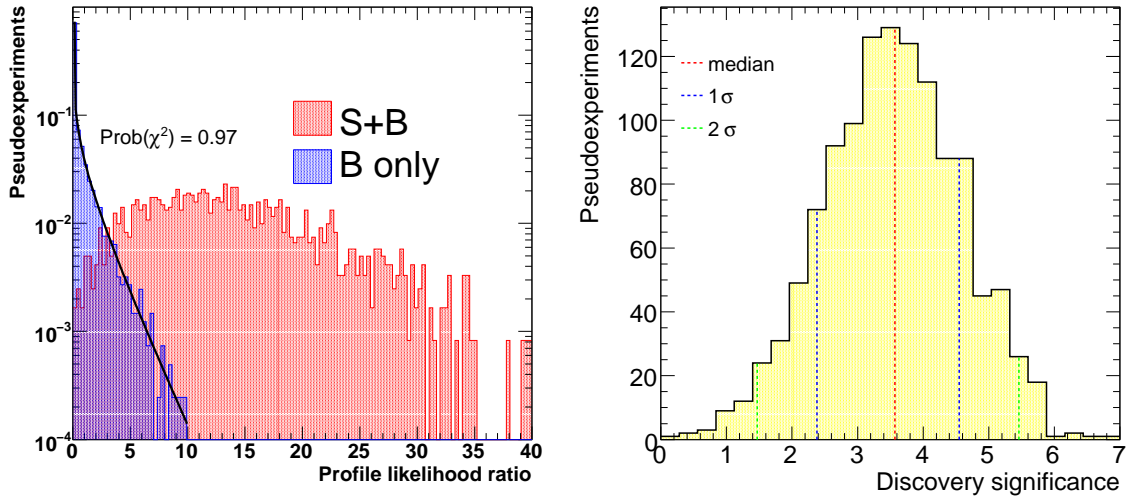


Figure 8.39: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to $m_H = 115$ GeV. The fit for the background only distribution is also shown. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 115$ GeV, with vertical lines corresponding to a confidence interval of 69% and 95% (1 and 2σ bands).

As a result, without considering the remaining systematic uncertainties which will be evaluated in the next section, it can be seen, that, at least for Higgs boson masses of 115 and 120 GeV, a median discovery sensitivity above 3σ can be obtained with 30 fb^{-1} of integrated

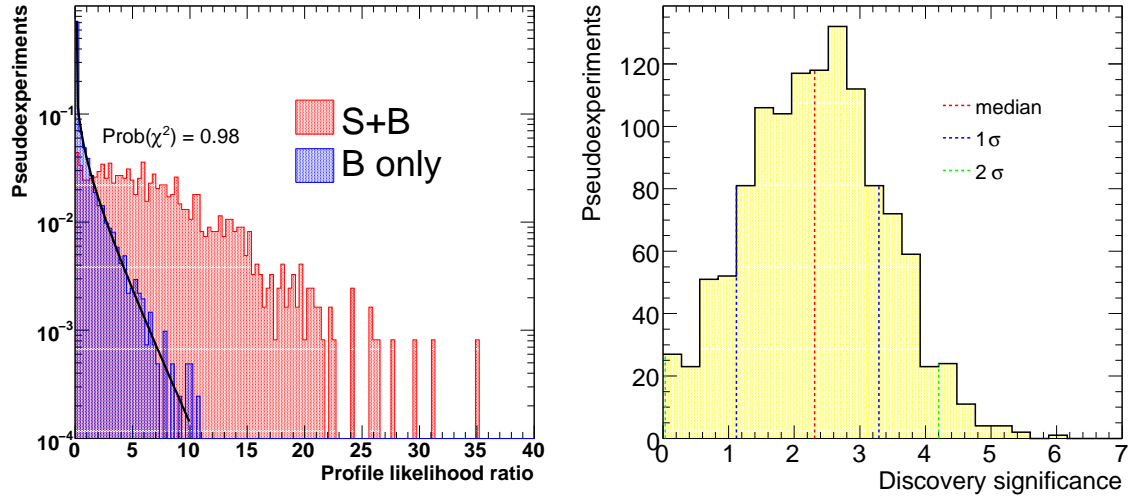


Figure 8.40: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to $m_H = 130$ GeV. The fit for the background only distribution is also shown. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 130$ GeV, with vertical lines corresponding to a confidence interval of 69% and 95% (1 and 2σ bands).

luminosity. At the same time, since the number of signal events, in the region of phase space where the signal purity is high, is small, the values of significances are significantly spread around their median values. This is more clearly shown in Fig. 8.41, which represents the discovery sensitivity, with corresponding 1 and 2σ bands, as a function of the Higgs boson mass.

8.6 Introduction of experimental systematic uncertainties

Before introducing the methods used to evaluate the impact of systematic uncertainties on the analysis and before taking them into account in the likelihood fit itself, the sources of systematic uncertainties considered for this study need to be defined.

As a first step, the performance of the detector in the Monte Carlo simulation is artificially reduced, in order to take the effect of the expected residual misalignment on b -tagging into account and to correct for differences between the ATLFast-II and full simulation. The motivation for the single steps is given below:

- The b -tagging light-jet rejection is reduced by relative 30%, according to the decrease in b -tagging performance expected due to the effect of residual misalignment after a realistic alignment procedure has been carried out [44].
- The trigger efficiency is accounted for, as studied on the WH sample using a sample of fully simulated Monte Carlo events in Section 7.6 (providing an efficiency drop of 0.7% for events with $E_T^{\text{miss}} < 250$ GeV).
- The electron reconstruction efficiency is reduced by 3%, to account for the different electron reconstruction efficiency in full simulation and in AtI Fast-II, discarding randomly electron candidates (according to Section C.4.1).

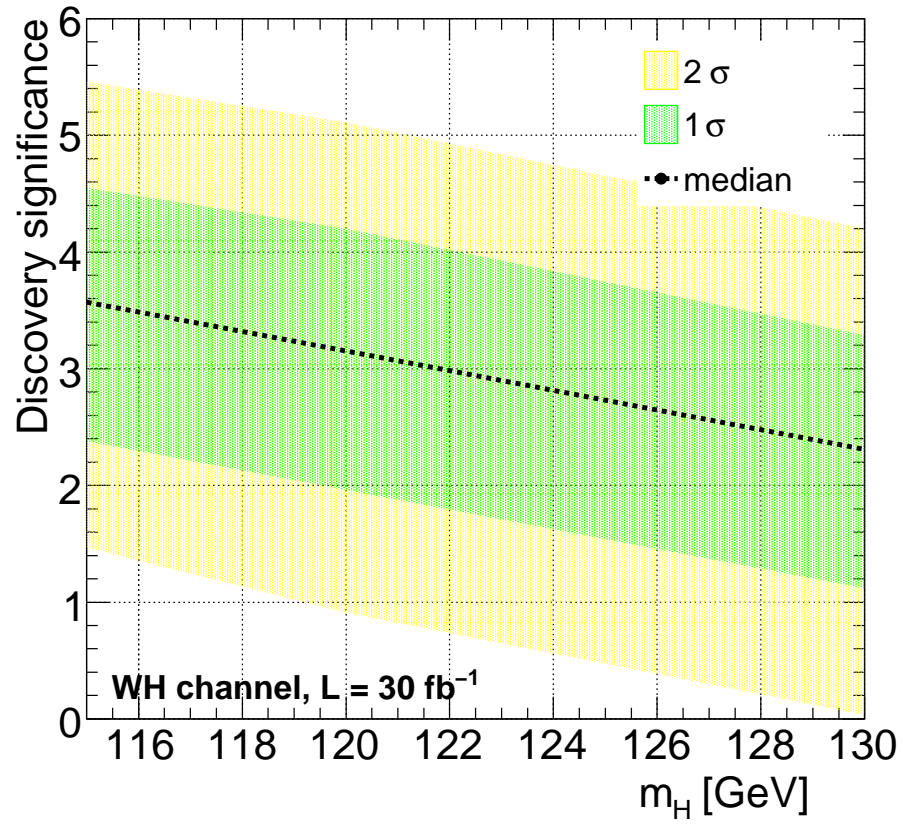


Figure 8.41: Discovery sensitivity as a function of the Higgs boson mass, for an integrated luminosity of 30 fb^{-1} . The effect of the systematic degradation of the detector performance due to the residual misalignment of the tracking detector, of differences between the fast and full detector simulation and of systematic deformations of the PDFs are not included here.

- The electron transverse momenta are corrected by scaling them up by 1.5%, again correcting AtlFast-II to full simulation (according to Section C.4.1).
- The signal efficiency is reduced by 5%, accounting for the difference in the $H \rightarrow b\bar{b}$ hadronic reconstruction efficiency between full simulation and AtlFast-II, as analysed in Section C.4.2. Most probably the same decrease in efficiency will affect the backgrounds as well, but, before a more detailed full simulation study of the backgrounds will be available, a conservative assumption is made here and only the signal efficiency is scaled down.

These corrections are applied directly on the reconstructed Monte Carlo events and physics objects (jets, leptons, E_T^{miss} , ...) before the analysis selection procedure is started. For artificially decreasing the b -tagging light-jet rejection, instead of relying on a random increase of the amount of light-jets or c -jets, which would be uncorrelated with the b -weight values, a deformation of the b -weight variables is applied on the truth labelled light and c -jets. This is performed in such a way to induce a fixed relative change in the b -tagging rejection values independently on the actual value of the b -weight cut chosen, in the range of all b -weight cut values which are of interest for the analysis (corresponding to a b -tagging efficiency from around 20 to 90%). The same procedure is applied later for artificially reducing the b -tagging efficiency, but in this case only jets truth labelled as b -jets are affected and a fixed relative change in the b -tagging efficiency is induced, instead of a change in the b -tagging rejection rate. This procedure is sufficiently realistic, but it clearly cannot substitute a more detailed study on the influence of systematic uncertainties on the kinematic variables used as input to the b -tagging algorithms and therefore on the b -weight distribution.

While the data samples of the original Monte Carlo samples (used for the analysis selection introduced in Section 8.2) are referred as *perfect* in the following, the data samples where the above mentioned systematic degradations of the detector performance have been applied are defined as *nominal*.

An eventual influence of the trigger acceptance on the shapes of the PDFs of the discriminating variables in the background samples was not considered. In general, given the very high p_T cut adopted in the analysis selection on the Higgs boson candidate ($p_T > 200$ GeV), this correction should be very small and this is confirmed by the trigger simulation on the signal sample, where the trigger efficiency is found to be very high (99.6%) and the influence on it on the shape of the PDFs for the discriminating variables is found to be negligible. However this needs to be checked also on all background samples, as soon as Monte Carlo samples with information about the trigger simulation will be available.

Applying the *loose* analysis selection on top of the *nominal* samples one obtains the analysis selection given in Table 8.11, for the signal and the main backgrounds, and in Table 8.12, for the remaining (minor) backgrounds. It can be clearly seen that the signal efficiency is reduced, so that the expected signal yield is decreased by $\approx 8.5\%$ for $m_H = 120$ GeV, mainly due to the applied AtlFast-II to full simulation correction factor, and the background efficiency is significantly increased, mainly due to the decrease of the b -tagging performance due to residual misalignment (e.g. the $t\bar{t}$ background is increased by $\approx 15\%$, while the W +jet background is increased by $\approx 25\%$).

The shapes for the PDFs to be fed into the likelihood function are determined again on the *nominal* samples, according to the procedure already introduced in Section 8.2.2. The differences with respect to the *perfect* case was already shown in Figs. 8.3-8.7. The differences in shape of the new PDFs with respect to the old ones is very small (except for the b -weight distribution): therefore the main impact on the discovery significance is expected to come from the new expectation values for the signal and background yields.

	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	W+jets
After filter cuts	964.4 \pm 5.9	860.5 \pm 6.4	656.5 \pm 4.1	3950 \pm 26	1233769	93868	584197
1 Higgs candidate	646.2 \pm 2.8	570.3 \pm 3.0	429.6 \pm 1.9	3509.0 \pm 8.1	806163	69377	562041
filtered $p_T > 200$ GeV	581.9 \pm 2.9	513.1 \pm 3.2	387.8 \pm 2.0	3108 \pm 11	709372	60256	413412
Missing $E_T > 30$ GeV	414.0 \pm 2.9	362.6 \pm 3.2	273.7 \pm 2.0	2184 \pm 13	552408	46779	318415
$p_T(W) > 200$ GeV	193.5 \pm 2.4	169.7 \pm 2.6	127.3 \pm 1.6	1214 \pm 12	137623	18434	205248
$p_T(e/\mu) > 30$ GeV	164.9 \pm 2.2	144.0 \pm 2.4	107.1 \pm 1.5	991 \pm 11	114485	15610	176480
$p_T(\text{additional } \mu) < 10$ GeV	163.9 \pm 2.2	143.0 \pm 2.4	106.4 \pm 1.5	937 \pm 11	106292	14880	176019
$p_T(\text{additional } e) < 10$ GeV	161.5 \pm 2.2	141.2 \pm 2.4	104.8 \pm 1.5	879 \pm 11	97021	13799	173452
$\Delta\phi(W,H) < \frac{2}{3}\pi$	160.5 \pm 2.2	140.6 \pm 2.4	103.9 \pm 1.5	835 \pm 11	84633	12922	166273
no additional b -jets $p_T > 15$ GeV	149.4 \pm 2.2	128.9 \pm 2.3	94.1 \pm 1.4	784 \pm 10	30341	7713 \pm 46	159002
jets on W side $p_T < 60$ GeV	132.1 \pm 2.0	114.2 \pm 2.2	82.6 \pm 1.3	633.1 \pm 9.5	19237	5797 \pm 41	120145
jets on H side $p_T < 60$ GeV	118.8 \pm 2.0	101.3 \pm 2.1	74.3 \pm 1.3	522.0 \pm 8.7	13750	4328 \pm 35	93105
one subjet b -tagged	107.5 \pm 1.9	90.0 \pm 2.0	66.2 \pm 1.2	133.2 \pm 4.7	8733	2442 \pm 27	8538
both subjects b -tagged	53.4 \pm 1.4	43.8 \pm 1.4	32.01 \pm 0.88	44.6 \pm 2.7	660 \pm 20	180.5 \pm 7.4	332 \pm 17
loose fit cuts	53.2 \pm 1.4	43.7 \pm 1.4	31.91 \pm 0.88	43.9 \pm 2.7	645 \pm 20	174.8 \pm 7.3	322 \pm 17

Table 8.11: Number of expected events passing through the analysis selection procedure as applied to the *nominal* samples for the signal and the main backgrounds, projected to 30 fb⁻¹ of collected data. LO Monte Carlo cross sections are used. For the signal, the degradation of 5% in efficiency due to the AtI Fast-II to full simulation correction factor is not considered in this table.

	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z+jets
After filter cuts	440.4 \pm 2.9	8649 \pm 33	4094 \pm 21	1807654	277706
1 Higgs candidate	295.9 \pm 1.3	8427.5 \pm 5.3	3372.3 \pm 8.2	160098	225579
filtered $p_T > 200$ GeV	267.9 \pm 1.4	7356 \pm 12	2995.1 \pm 9.5	103095	175078
Missing $E_T > 30$ GeV	141.8 \pm 1.3	6217 \pm 15	1415 \pm 10	75074	62882
$p_T(W) > 200$ GeV	33.86 \pm 0.76	4145 \pm 17	460.1 \pm 6.8	3327	13511
$p_T(e/\mu) > 30$ GeV	8.95 \pm 0.40	3433 \pm 16	264.7 \pm 5.3	2905	10133
$p_T(\text{additional } \mu) < 10$ GeV	5.59 \pm 0.32	3411 \pm 16	177.6 \pm 4.4	2524	7167
$p_T(\text{additional } e) < 10$ GeV	3.57 \pm 0.26	3332 \pm 16	122.7 \pm 3.6	2063	4788
$\Delta\phi(W,H) < \frac{2}{3}\pi$	3.45 \pm 0.25	3231 \pm 16	112.7 \pm 3.5	1667	4100
no additional b -jets $p_T > 15$ GeV	3.06 \pm 0.24	3099 \pm 16	102.8 \pm 3.3	500 \pm 28	3763
jets on W side $p_T < 60$ GeV	2.32 \pm 0.21	2500 \pm 15	82.2 \pm 3.0	326 \pm 22	2754 \pm 46
jets on H side $p_T < 60$ GeV	2.13 \pm 0.20	2069 \pm 14	68.1 \pm 2.7	236 \pm 19	2154 \pm 41
one subjet b -tagged	1.94 \pm 0.19	331.3 \pm 6.4	17.6 \pm 1.4	142 \pm 15	248 \pm 14
both subjects b -tagged	0.94 \pm 0.13	8.5 \pm 1.0	5.37 \pm 0.77	9.1 \pm 3.7	13.3 \pm 3.2
loose fit cuts	0.94 \pm 0.13	8.5 \pm 1.0	5.37 \pm 0.77	9.1 \pm 3.7	13.3 \pm 3.2

Table 8.12: Number of expected events passing through the analysis selection procedure as applied to the *nominal* samples for the remaining backgrounds, projected to 30 fb⁻¹ of collected data. LO Monte Carlo cross sections are used.

The discovery potential is re-evaluated for the *nominal case*, to obtain a reference before introducing the effects of systematic deformations of the PDFs. The result of running a complete set of new pseudo-experiments is shown in Figs. 8.42-8.44. The updated discovery sensitivity plot as a function of the Higgs boson mass is shown in Fig. 8.45.

Using as a reference the event selection as applied on the *nominal* samples, a variety of systematic uncertainties are considered on top, following the prescriptions adopted for the most recent ATLAS physics analyses [44], except for the case of the electron efficiency, the error for which is inflated by the uncertainty of the related AtI Fast-II / full simulation correction factor used in the definition of the *nominal* sample. The single uncertainties are presented in more detail in Table 8.13. In case both upwards and downwards fluctuations for a certain systematic uncertainty are considered, a weight of 0.5 is assigned, otherwise the weight is 1. For every systematic scenario considered, the basic reconstructed physics objects (jets, leptons, E_T^{miss} , ...) of the *nominal* samples are corrected accordingly and the full analysis selection is re-applied. The jet scale and resolution is assumed to be affected in the same way for both light- and b -jets, since a dedicated treatment of the b -jets energy calibration is still being developed in ATLAS.

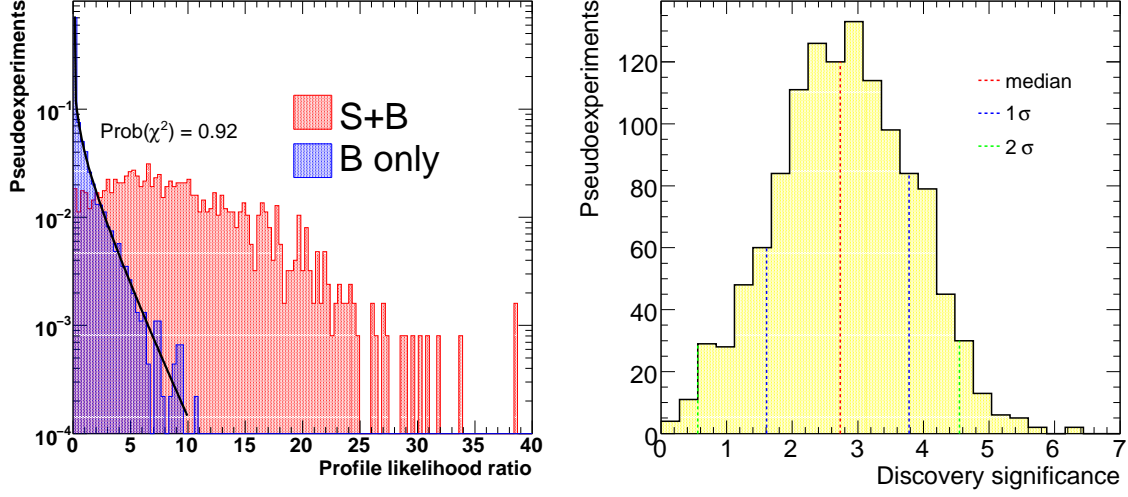


Figure 8.42: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 120$ GeV. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 120$ GeV.

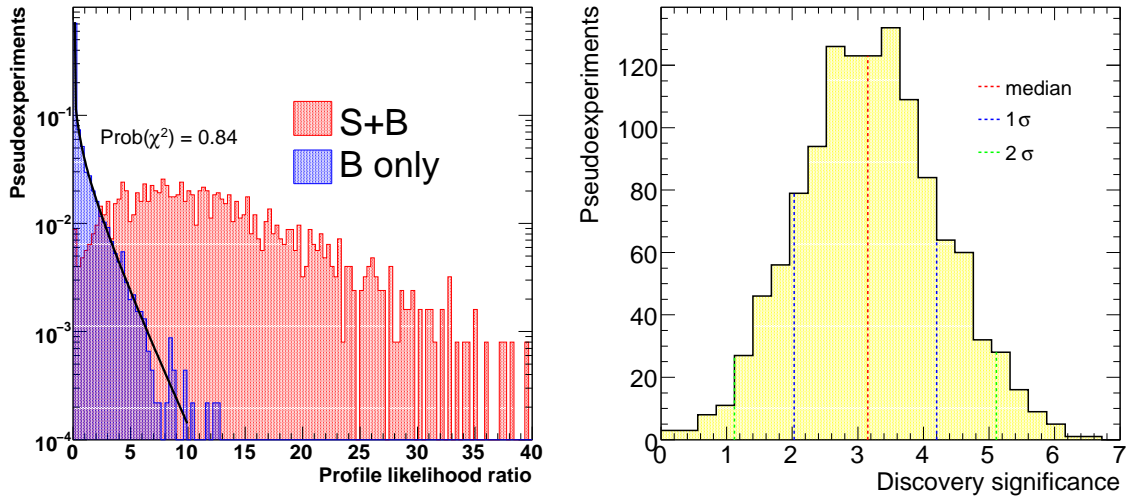


Figure 8.43: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 115$ GeV. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 115$ GeV.

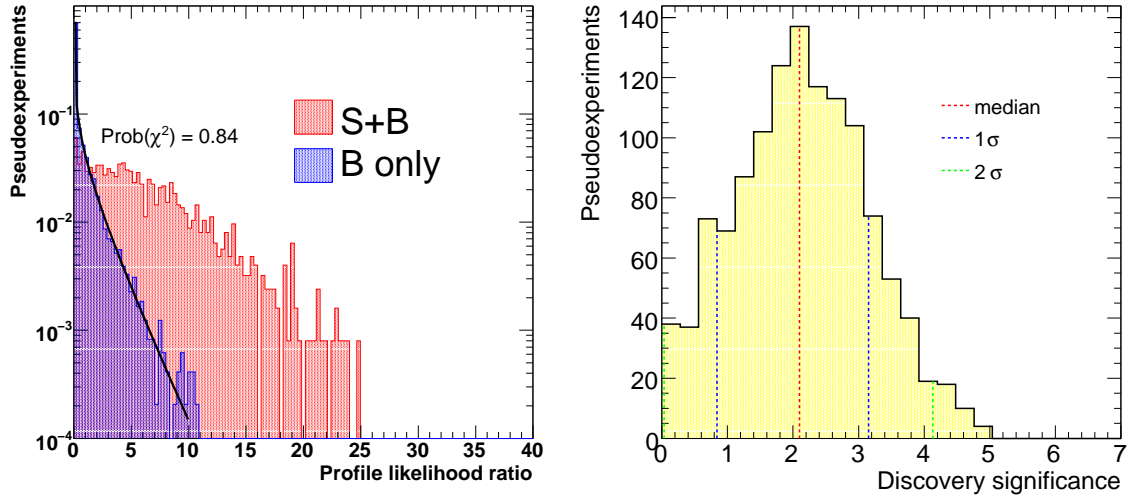


Figure 8.44: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 130$ GeV. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 130$ GeV.

Number	Systematic shift	Weight	Description
1	Jet energy scale up	0.5	+3% for $ \eta(\text{jet}) < 3.2$ and +10% for $ \eta(\text{jet}) > 3.2$
2	Jet energy scale down	0.5	-3% for $ \eta(\text{jet}) < 3.2$ and -10% for $ \eta(\text{jet}) > 3.2$
3	Jet energy smearing	1	$ \eta < 3.2$: smear by $\sigma(E) = 0.45\sqrt{E}$ $ \eta > 3.2$: smear by $\sigma(E) = 0.63\sqrt{E}$ (E in GeV)
4	Muon efficiency	1	-1% muon efficiency (randomly discard muons)
5	Muon energy smearing	1	smear by $\sigma\left(\frac{1}{p_T}\right) = 0.011 \frac{1}{p_T} \oplus 0.00017$ (p_T in GeV)
6	Muon energy scale up	0.5	+1% muon energy
7	Muon energy scale down	0.5	-1% muon energy
8	Electron energy smearing	1	smear E_T by $\sigma(E_T) = 0.0073 E_T$
9	Electron energy scale up	0.5	+0.5% electron energy
10	Electron energy scale down	0.5	-0.5% electron energy
14	Electron efficiency up	0.5	+1.73% electron efficiency
11	Electron efficiency down	0.5	-1.73% electron efficiency
12	b -tagging eff. down	1	decrease ϵ_B by 5 % (deforming b -weight)
15	b -tagging rej. up	0.5	+10% light jet rejection (deforming b -weight)
16	b -tagging rej. down	0.5	-10% light jet rejection (deforming b -weight)

Table 8.13: List of experimental systematic uncertainties considered.

The effect of the single systematics on the yields is shown in Tables 8.14 and 8.15. The total effect, summing up in quadrature the contribution of all systematic deviations, is also shown. This is mostly for illustration, since the correlations between the different signal and background components for every single systematic uncertainty need to be considered, to avoid overestimating the impact of such uncertainties on the sensitivity of the analysis. Although in the following the effects of these variations on the likelihood fit will be taken into account, it should be stressed that, since the likelihood fit extracts the number of signal and all background events directly from data, these systematic deviations are not expected to change the median significance, but will just broaden the distribution of the expected significances

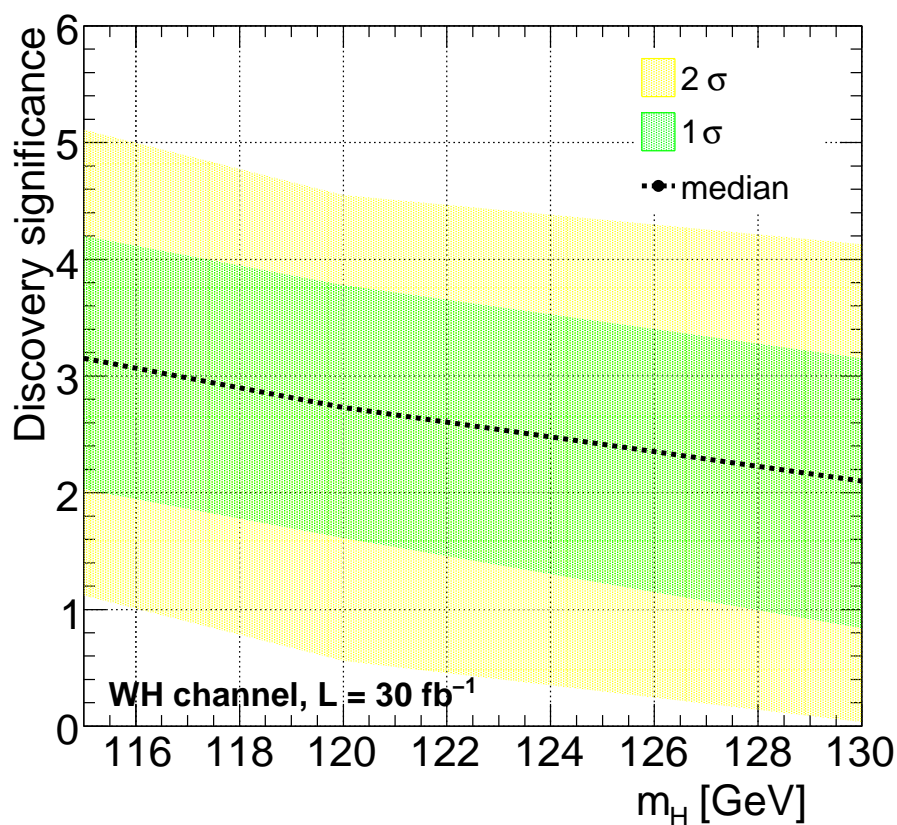


Figure 8.45: Discovery sensitivity as a function of the Higgs boson mass, for an integrated luminosity of 30 fb^{-1} , for the *nominal* scenario described in the text. The effect of the systematic degradation of the detector performance due to the residual misalignment of the tracking detector and of differences between the fast and full detector simulation are included. Further experimental systematic uncertainties leading to systematic deformations of the PDFs are not included here.

around the same median value.

Systematic shift	WH(115)	WH(120)	WH(130)	WZ	$t\bar{t}(p_T^{min})$	Wt	W +jets
nominal	50.6 ± 1.3	41.5 ± 1.3	30.3 ± 0.8	43.9 ± 2.7	645.4 ± 20.1	174.8 ± 7.3	321.8 ± 17.2
Jet energy scale up	+3.3 %	+4.6 %	+3.2 %	+5.4 %	+4.0 %	+8.5 %	+10.0 %
Jet energy scale down	-3.6 %	-2.8 %	-2.8 %	-6.2 %	-2.2 %	-4.0 %	-7.4 %
Jet resolution	-1.4 %	-0.2 %	+0.4 %	-1.5 %	+0.9 %	+1.6 %	-2.3 %
μ efficiency	-1.2 %	+0.4 %	-0.9 %	-0.4 %	-1.7 %	-1.0 %	-1.4 %
μ resolution	-0.1 %	+0.4 %	-1.4 %	-0.4 %	-0.1 %	+0.7 %	-1.1 %
μ energy scale up	+0.0 %	+0.6 %	-1.3 %	+0.0 %	-0.6 %	+1.2 %	-0.3 %
μ energy scale down	-0.6 %	+0.2 %	+0.6 %	-0.4 %	-1.4 %	-0.5 %	-0.6 %
e resolution	-0.8 %	+0.7 %	-0.1 %	-0.8 %	-1.0 %	+1.7 %	-0.9 %
e energy scale up	-0.1 %	+0.2 %	-0.3 %	+0.0 %	-0.8 %	+1.0 %	+0.3 %
e energy scale down	+0.0 %	+0.4 %	-0.5 %	-1.9 %	-0.2 %	-0.7 %	-0.3 %
e efficiency down	-0.9 %	-2.1 %	-0.6 %	+0.4 %	-1.4 %	-1.6 %	-0.3 %
e efficiency up	+0.3 %	+1.1 %	+0.3 %	+1.2 %	+0.6 %	+1.7 %	+0.9 %
b -tagging eff	-12.3 %	-10.4 %	-11.1 %	-12.3 %	-0.8 %	-4.9 %	-9.7 %
l -rejection down	-2.0 %	-1.4 %	-0.2 %	+1.5 %	+3.5 %	+7.8 %	+4.6 %
l -rejection up	+0.6 %	+1.1 %	+0.6 %	+0.0 %	-4.1 %	-1.9 %	-6.6 %
Total uncertainty:	± 13.0 %	± 11.3 %	± 11.7 %	± 13.9 %	± 5.7 %	± 10.6 %	± 14.6 %

Table 8.14: Effect of the single systematic uncertainties on the expected number of events for the signal and the main background samples, in terms of percentage of variation with respect to the yields listed in the first row.

Systematic shift	ZH(120)	WW	ZZ	$t\bar{t}(p_T^{max})$	Z +jets
nominal	0.9 ± 0.1	8.5 ± 1.0	5.4 ± 0.8	9.1 ± 3.7	6.2 ± 2.2
Jet energy scale up	+5.9 %	+0.0 %	+16.7 %	+0.0 %	-12.5 %
Jet energy scale down	-13.7 %	-9.1 %	-2.1 %	-33.3 %	-12.5 %
Jet resolution	+2.0 %	+0.0 %	+2.1 %	-33.3 %	+0.0 %
μ efficiency	+13.7 %	-1.5 %	+2.1 %	+0.0 %	+0.0 %
μ resolution	+5.9 %	-1.5 %	+6.3 %	-16.7 %	+12.5 %
μ energy scale up	+2.0 %	+0.0 %	+2.1 %	+0.0 %	+0.0 %
μ energy scale down	+2.0 %	-3.0 %	+2.1 %	+0.0 %	+12.5 %
e resolution	-3.9 %	-1.5 %	+2.1 %	+0.0 %	+12.5 %
e energy scale up	+3.9 %	-3.0 %	+0.0 %	+0.0 %	+0.0 %
e energy scale down	-3.9 %	-7.6 %	+2.1 %	+0.0 %	+0.0 %
e efficiency down	+7.8 %	-4.5 %	+2.1 %	+0.0 %	+0.0 %
e efficiency up	+3.9 %	-1.5 %	+0.0 %	+0.0 %	+0.0 %
b -tagging eff	-3.9 %	-1.5 %	-6.3 %	-16.7 %	-12.5 %
l -rejection down	+3.9 %	+7.6 %	+0.0 %	+0.0 %	+0.0 %
l -rejection up	+3.9 %	-15.2 %	+2.1 %	+0.0 %	+0.0 %
Total uncertainty:	± 21.0 %	± 15.6 %	± 15.6 %	± 47.1 %	± 26.5 %

Table 8.15: Effect of the single systematic uncertainties on the expected number of events for the reference signal and the remaining background samples, in terms of percentage of variation with respect to the yields listed in the first row.

The systematic shifts considered in the various scenarios will also induce a deformation of the distributions of the discriminating variables which enter the likelihood fit. If these deformations are taken into account properly, the statistical hypothesis test will remain unbiased, but a decrease in the discovery potential is expected.

8.7 Modelling of systematic uncertainties

In order to take both the systematic variations in the expected number of events and in the shapes of the PDFs into account, it is desirable to define a way to model all these effects analytically. This should permit:

- to preserve the correlations between deformations of PDFs of different signal or background components due to a common systematic uncertainty (e.g. a systematic shift due to the energy scale variation affects different signal or background distributions in a very similar and correlated way)
- to preserve the correlations among shape parameters describing the same PDF (e.g. an increase of events in the first bin of a step function causes usually a decrease in the second bin and so on)
- to preserve all linear correlations between the shape parameters and the yields (e.g. an increase in the W +jet background yield would not affect at all the discovery significance, if at the same time its mass distribution shows less events in the signal region, and viceversa)
- to combine all systematic variations in the expected number of events and in the shapes of the PDFs, so that the combined effect of all such variations can be simultaneously accounted for.

The method adopted to achieve this scope will be introduced here. The following notation is adopted: every single scenario of Table 8.13 will be referred to with the symbol s (set). Each signal or background component is represented by the symbol c (component). So a certain sample representing a certain component c and a certain scenario s , will be denoted $[c; s]$. A maximum likelihood fit is performed on each of these samples, which determines the parameters $\vec{\alpha}_c^s$ uniquely defining the shapes of the PDFs. In the case also the uncertainties on the yields need to be considered, the yields can just be added to the list of these parameters.

The parameters of the likelihood function describing the shapes of the PDFs (and, in case, the yields) suffer from two sources of systematic uncertainty: one is of statistical nature, which is due to the limited amount of available Monte Carlo statistics used to estimate their expectation values, the second is purely systematic, and its experimental component is taken into account by the scenarios introduced above.

The statistical uncertainty connected to the parameters $\vec{\alpha}_c^s$ is condensed in the covariance matrix returned by the likelihood fit on the single Monte Carlo samples, which is denoted $\text{COV}(\vec{\alpha}_c^s)$. Since there is no statistical correlation between the different samples corresponding to different signal or background components c , the $\vec{\alpha}_c^s$ and matrices $\text{COV}(\vec{\alpha}_c^s)$ for different samples c are uncorrelated. Since the systematic uncertainties will introduce correlations between these different components c , for every scenario s a single vector $\vec{\alpha}^s$ is constructed from $(\vec{\alpha}_0^s, \vec{\alpha}_1^s, \dots, \vec{\alpha}_N^s)$ and a single block diagonal covariance matrix merging the single covariance matrices

$$(\text{COV}(\vec{\alpha}_0^s), \text{COV}(\vec{\alpha}_1^s), \dots, \text{COV}(\vec{\alpha}_N^s)). \quad (8.25)$$

The parameters $\vec{\alpha}^s$ for the scenario with $s = 0$ (nominal scenario) provide an estimate for the expectation values, with statistical uncertainty given by the covariance matrix $\text{COV}(\vec{\alpha}^0)$. Then the systematic uncertainties, provided by the scenarios starting from $s = 1$, need to be considered on top. Since the systematic shifts of the various different scenarios are applied on top of exactly the same events, the statistical uncertainty on the variation of the parameters

$\vec{\alpha}^s$ with respect to $\vec{\alpha}^0$ can be assumed to be negligible. Such variations are therefore considered as corresponding to 1σ Gaussian fluctuations due to the systematic uncertainty introduced by the scenario considered. Using these, an additional covariance matrix can be built up:

$$\text{COV}_{\text{sys}}(\vec{\alpha})_{i,j} = \text{COV}(\alpha_i, \alpha_j) = \sum_{s=1}^{N_s} (\alpha_i^s - \alpha_i^0)(\alpha_j^s - \alpha_j^0) \quad (8.26)$$

Thanks to the non diagonal terms, all correlations are correctly handled; these terms also involve correlations between different signal and background components. Finally the two previously defined covariance matrices are simply summed up together:

$$\text{COV}_{\text{tot}}(\vec{\alpha})_{i,j} = \sum_{s=1}^{N_s} (\alpha_i^s - \alpha_i^0)(\alpha_j^s - \alpha_j^0) + \text{COV}_{\text{stat}}(\vec{\alpha}^0)_{i,j}. \quad (8.27)$$

The PDF modelling the global systematic variations expected for the parameters $\vec{\alpha}$ is then given by the multi-Gaussian function:

$$\text{PDF}(\vec{\alpha}) = \exp \left[-\frac{1}{2} (\vec{\alpha} - \vec{\alpha}^0)^T \text{COV}_{\text{tot}}^{-1} (\vec{\alpha} - \vec{\alpha}^0) \right] \quad (8.28)$$

In order to make this PDF easier to handle, after a full inversion of the covariance matrix, the property that symmetric matrices can be diagonalised through an orthogonal transformation is exploited: this removes the non diagonal terms of the quadratic form in the exponential. In this way a factorised PDF is obtained:

$$\text{PDF}(\vec{\alpha}) = \prod_{t=0}^{N_{\text{par}}} \exp \left[-\frac{(\beta_t(\vec{\alpha}) - \beta_t(\vec{\alpha}^0))^2}{2\sigma_t^2} \right], \quad (8.29)$$

where

$$\begin{aligned} \vec{\beta}_i &= O_{ij} \vec{\alpha}_j \\ \vec{\beta}_i^0 &= O_{ij} \vec{\alpha}_j^0 \\ \text{COV}^{-1}(\vec{\beta})_{is} &= (1/\sigma_i)^2 \delta_{is} = O_{ij} \text{COV}_{\text{tot},jk}^{-1} O_{ks}^T, \end{aligned}$$

and O_{ij} is the orthogonal transformation obtained with the diagonalisation. In this way the PDF is now expressed as a simple product of Gaussian PDFs. The parameters σ_t and β_t^0 are constants, while the variables β_t need to be dealt with as linear combinations of the fit parameters $\vec{\alpha}$.

Once the multi-Gaussian PDF is defined, it is easy to generate random configurations of the fit parameters α according to its distribution. In fact, after the factorisation of the multi-Gaussian PDF in a product of Gaussian functions, this is simply realised by generating random Gaussian shifts of the parameters $\vec{\beta}$ around $\vec{\beta}^0$ with Gaussian widths $\vec{\sigma}_\beta$ and then translating these shifts back into corresponding shifts of the parameters $\vec{\alpha}$.

The effect in terms of systematic deformations of the signal PDFs, for $m_H = 120$ GeV, are shown in Fig. 8.46. For each of the generated sets of PDF parameters $\vec{\alpha}$, the PDF is plotted after normalisation to an area of 1: the PDFs corresponding to many different configurations $\vec{\alpha}$ fill up a two-dimensional space of probabilities, which is used, as a function of the discriminating variable, to compute the median PDF value (on the y axis) and to define the 1 and 2σ intervals around the median, corresponding respectively to a confidence level of 68 and 95%. The impact of the statistical covariance term and of the purely systematic

ones are first considered separately and then combined into an overall uncertainty, according to the formula of Eq. 8.27. The same is shown in Fig. 8.47 for W +jet, in Fig. 8.48 for WZ , in Fig. 8.49 for $t\bar{t}$ and in Fig. 8.50 for the Wt background. In order for this method to correctly describe the effect of the systematic uncertainty on the shapes of the PDFs, the parametrisation chosen for the PDFs must be able to account for the deformations induced by the systematic uncertainties introduced by the various scenarios: this is always trivially the case for all distributions described by a step function, but it is not necessarily the case when a more complex functional form is assumed, with a low number of parameters. This should not be the case for the PDFs used in this study, but when the final analysis will be performed on data, it may be a good idea to use alternative functional forms to describe the same PDFs and compare the results, in order to exclude any effect of this kind.

Some of the involved uncertainties are easier to understand by analysing directly the variations of the parameters $\vec{\alpha}$, some of which represent well defined physics observables and detector resolution parameters. This is for example the case for shifts in the jet energy scale and of changes in the jet energy resolution, which for example have a direct impact on the mean value and width of the (modified) bifurcated Gaussian used to model the signal and WZ mass PDFs. The uncertainty on the mean value and on the width of the Gauss function used to parametrise the signal PDF is shown in Fig. 8.51. The mean squared shift of the peak is around 3%, exactly as expected from the assumed energy scale uncertainty, and the width of the core Gaussian of the Higgs mass PDF is allowed to change by around 10%. The same is shown in Fig. 8.52 for the WZ background. Since in this PDF there is an important cross talk between the right σ of the bifurcated modified Gaussian used in the PDF and its mean value, the impact of the systematic uncertainties on the Gaussian peak and on the resolution seems to be higher, however the negative correlations between the two is taken into account, as shown in Fig. 8.52 (bottom plot).

Since the multi-Gaussian PDF takes all linear correlations into account, but most of the plots shown so far cannot give any indication of these, some correlations plots are shown in Fig. 8.53. It is important for example to account for the fact that if the b -tagging efficiency is increased or the jet energy scale is shifted to higher values, the amount of signal and W +jet background events will scale very similarly: the plots show that the correlation between the two due to common systematic uncertainties is around 71%. It is also interesting to notice that there is not such a clear correlation between the expected number of events of the $t\bar{t}$ background between the signal region and the control sample: this can be explained by the fact that in the signal region the presence of an additional b -jet (beyond the two b -quarks from the Higgs boson candidate) is vetoed, while in the control sample it is required, so that the b -tagging uncertainty will affect the two samples in a partially opposite way.

It is also important that the bin-to-bin correlations of the distributions described by a step function (for example the PDF for the b -weight) are correctly modelled. The effect due to the systematic uncertainties can be different according the signal or background component considered, as shown in Fig. 8.54.

When modelling the systematic uncertainties with the use of the multi-Gaussian PDF some important assumptions are also made:

- A Gaussian distribution is assumed for the systematic shifts considered in the various different scenarios, where the 68% confidence interval is mapped onto their $\pm 1\sigma$ variations.
- A linear relation is assumed between variations in the space of parameters $\vec{\alpha}$ and the corresponding systematic shifts of the various different scenarios. The corresponding variations in the parameters $\vec{\alpha}$ are therefore again Gaussian distributed.

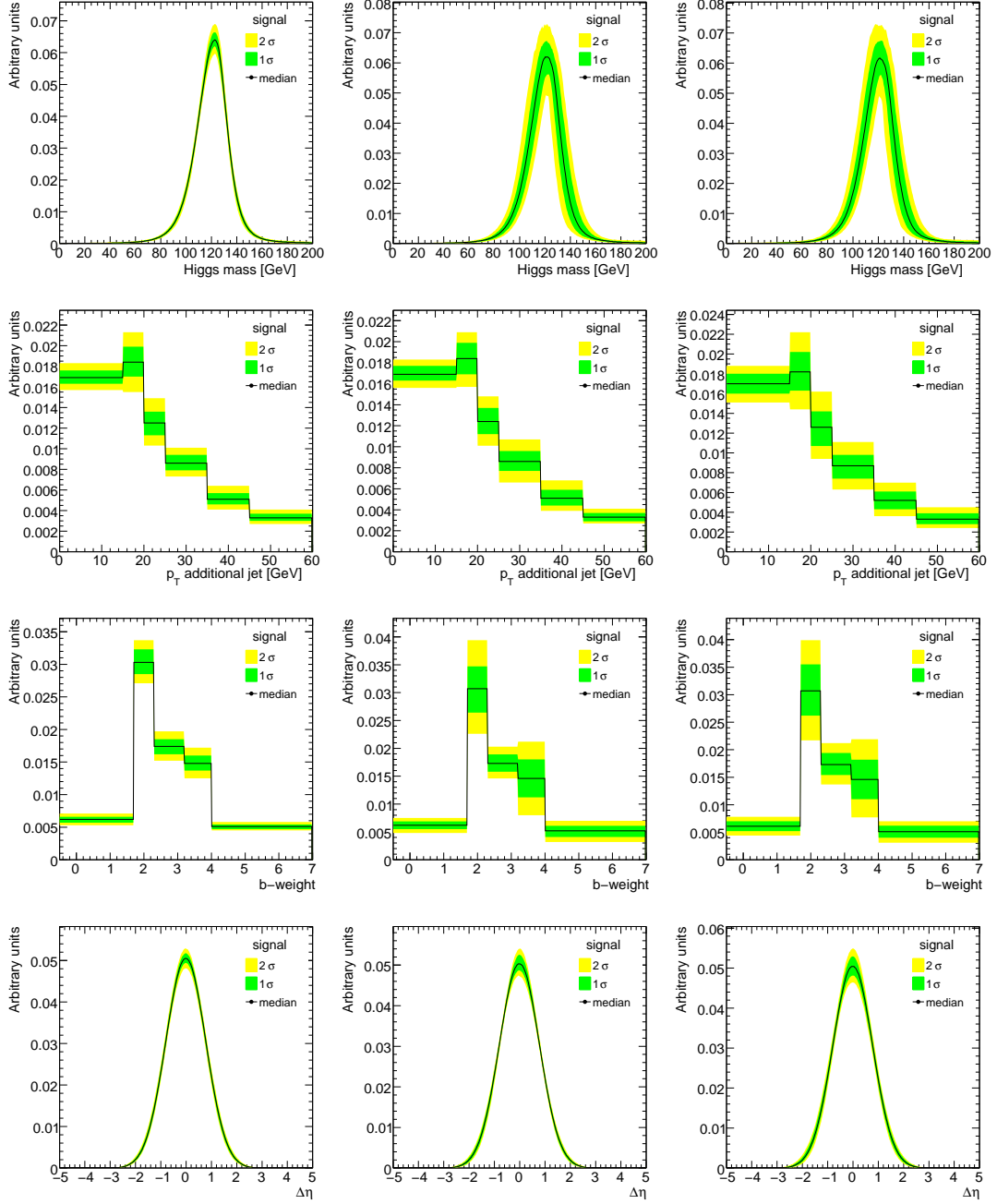


Figure 8.46: Systematic uncertainties on the signal PDFs in terms of uncertainty bands at a confidence level of 68% (1σ) and 95% (2σ). In the left column only the statistical contribution is shown, in the middle one only the systematic contribution, in the right one the combined uncertainty.

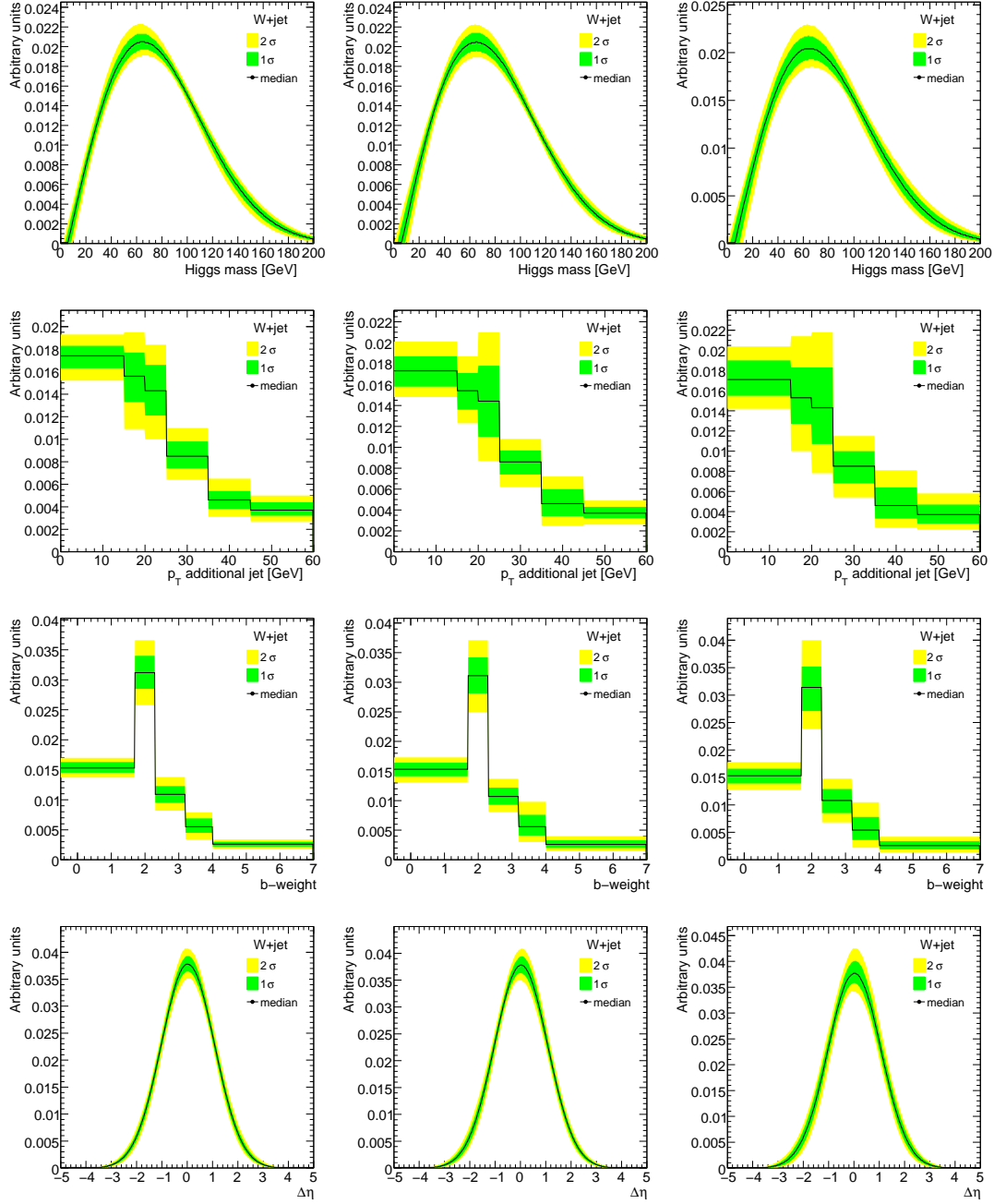


Figure 8.47: Systematic uncertainties on the PDFs for the W +jet background in terms of uncertainty bands at a confidence level of 68% (1σ) and 95% (2σ). In the left column only the statistical contribution is shown, in the middle one only the systematic contribution, in the right one the combined uncertainty.

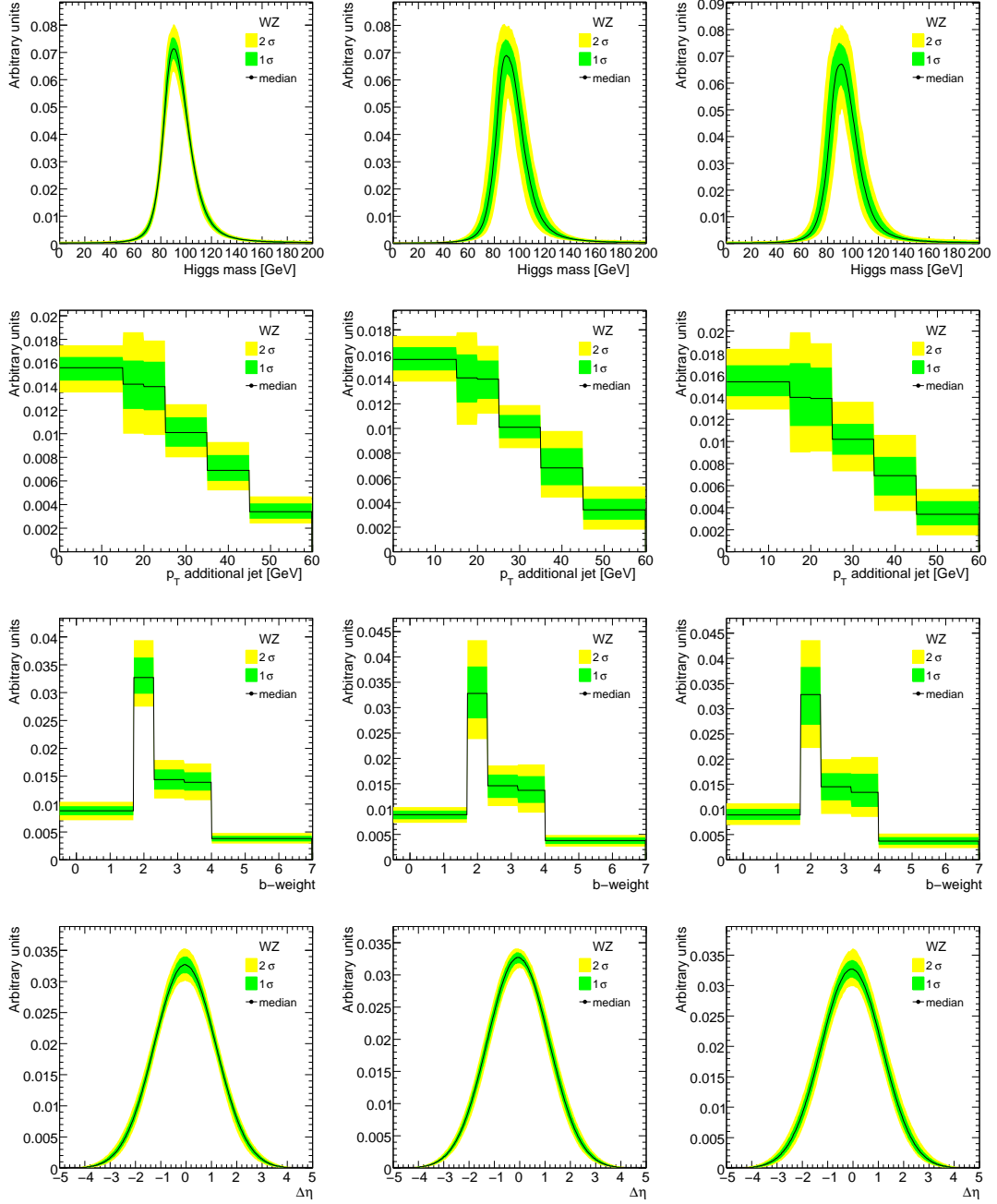


Figure 8.48: Systematic uncertainties on the PDFs for the WZ background in terms of uncertainty bands at a confidence level of 68% (1σ) and 95% (2σ). In the left column only the statistical contribution is shown, in the middle one only the systematic contribution, in the right one the combined uncertainty.

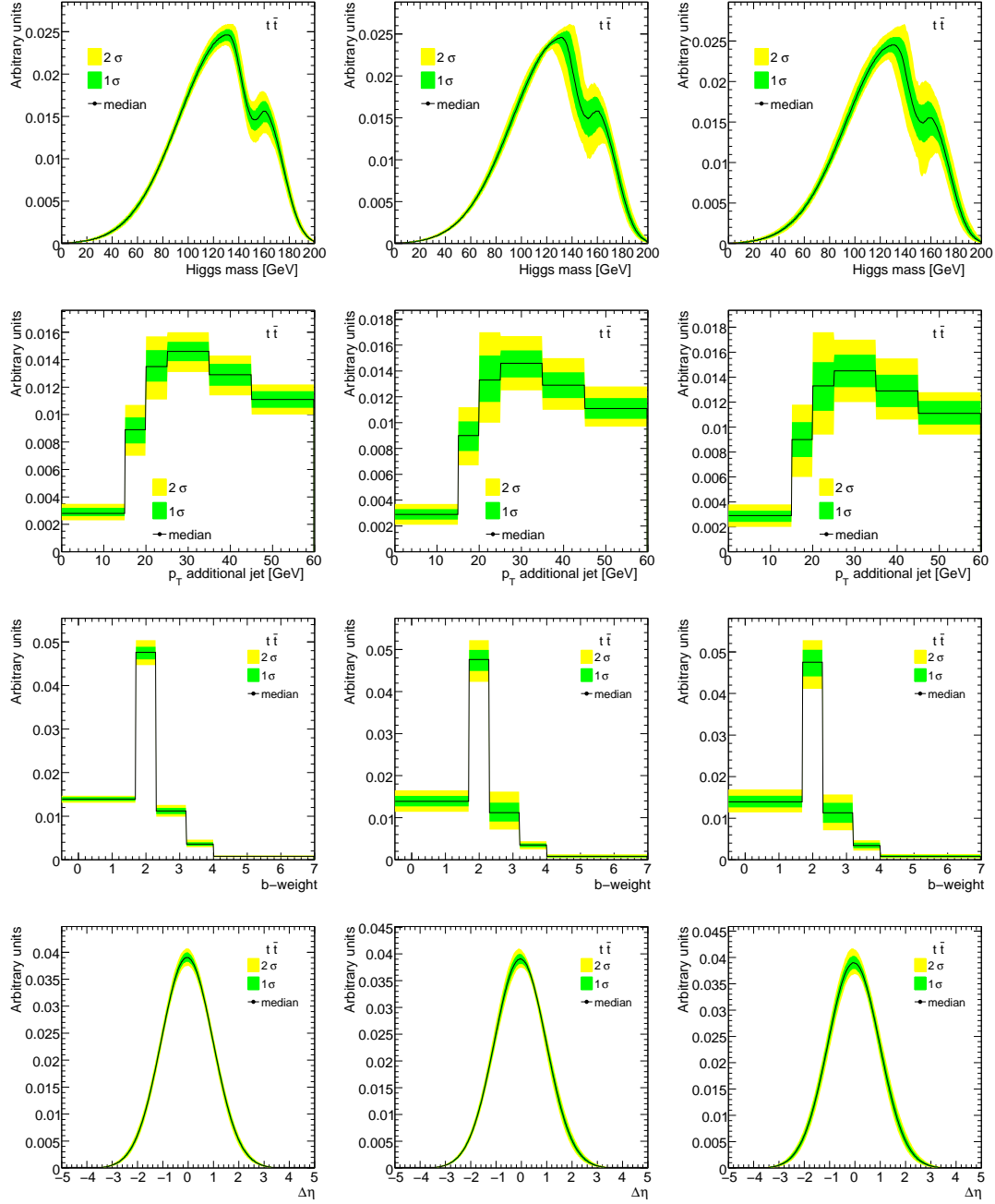


Figure 8.49: Systematic uncertainties on the PDFs for the $t\bar{t}$ background in terms of uncertainty bands at a confidence level of 68% (1σ) and 95% (2σ). On the left column only the statistical contribution is shown, in the middle one only the systematic contribution, in the right one the combined uncertainty.

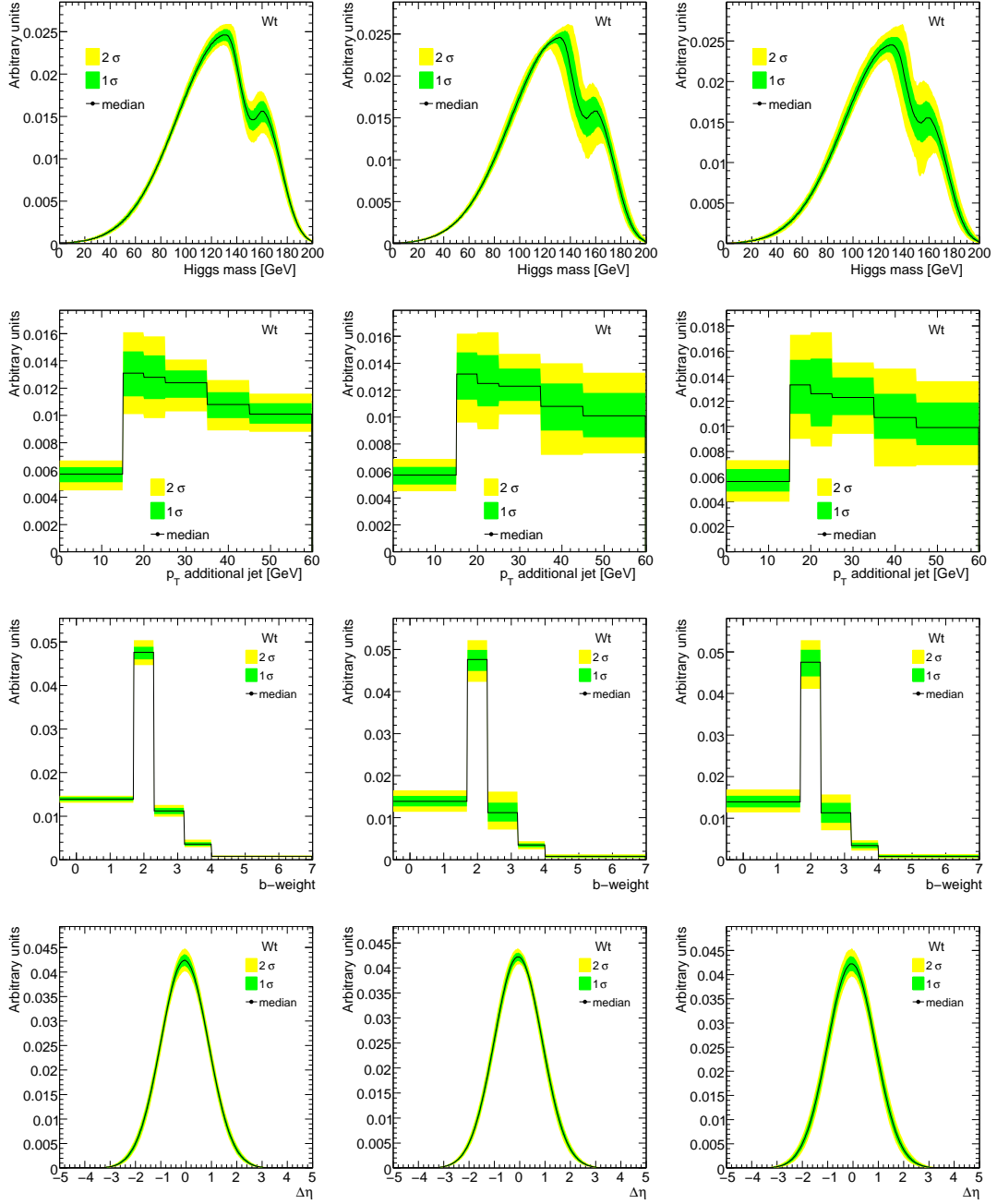


Figure 8.50: Systematic uncertainties on the PDFs for the Wt background in terms of uncertainty bands at a confidence level of 68% (1σ) and 95% (2σ). On the left column only the statistical contribution is shown, in the middle only the systematic contribution, in the right the combined uncertainty.

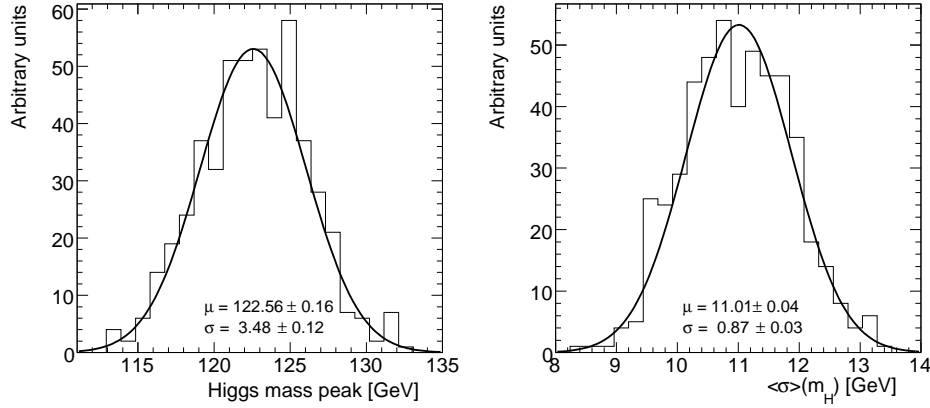


Figure 8.51: (Left) Distribution of systematic variations of the mean value of the the Gaussian PDF used in the likelihood fit to model the invariant mass of the Higgs boson in the WH signal with $m_H = 120$ GeV. (Right) Distribution of systematic variations of the width of the same Gaussian, obtained from the average of the left and right σ of the signal PDF.

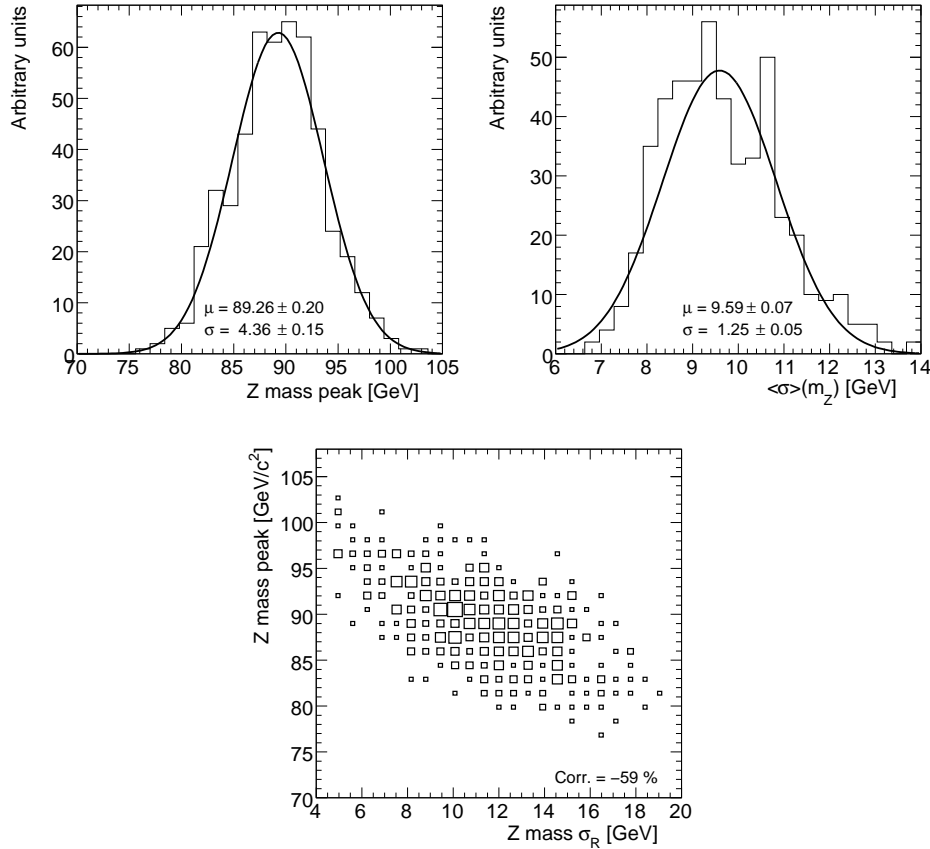


Figure 8.52: (Left) Distribution of systematic variations of the mean of the Gaussian PDF used in the likelihood fit to model the invariant mass of the Higgs boson candidate in the WZ background. (Right) Distribution of systematic variations of the width of the same Gaussian PDF, obtained from the average of the left and right σ . (Bottom) Correlation between systematic variations of the mean value and of the left σ of the invariant mass PDF for the WZ background.

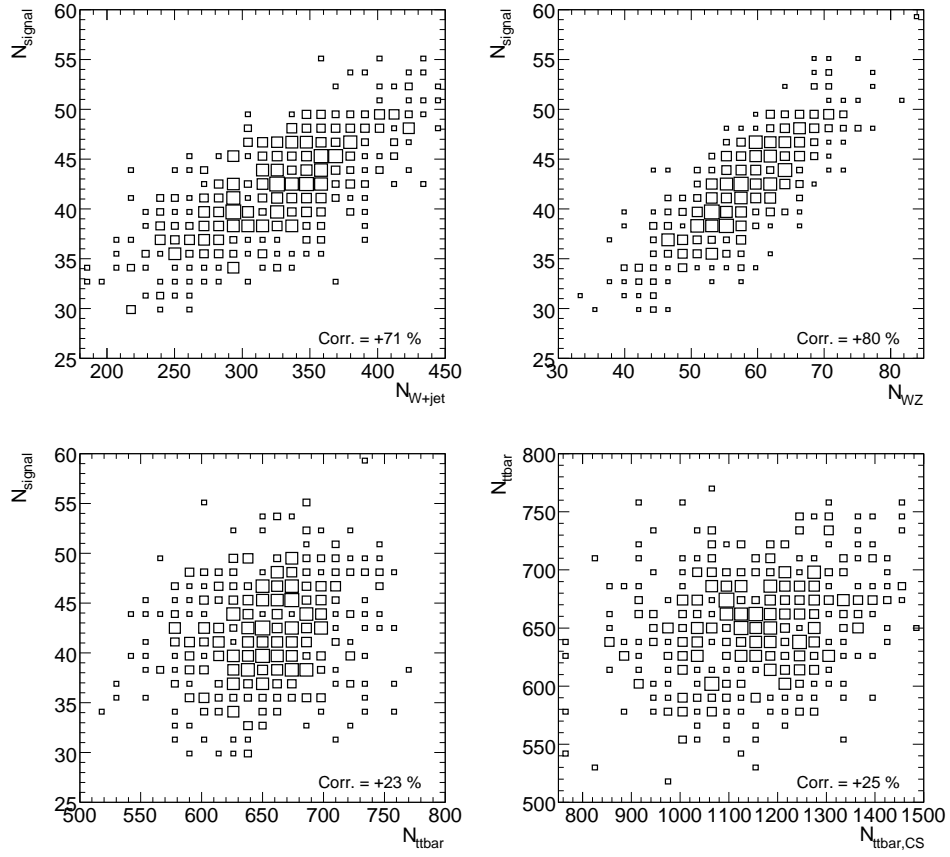


Figure 8.53: Correlations between systematic variations of the signal yield and variations of the background yields for the WZ , $W+\text{jet}$ or $t\bar{t}$ backgrounds. In addition, the same correlation plot is shown between the yield for the $t\bar{t}$ background in the signal region and in the control sample.

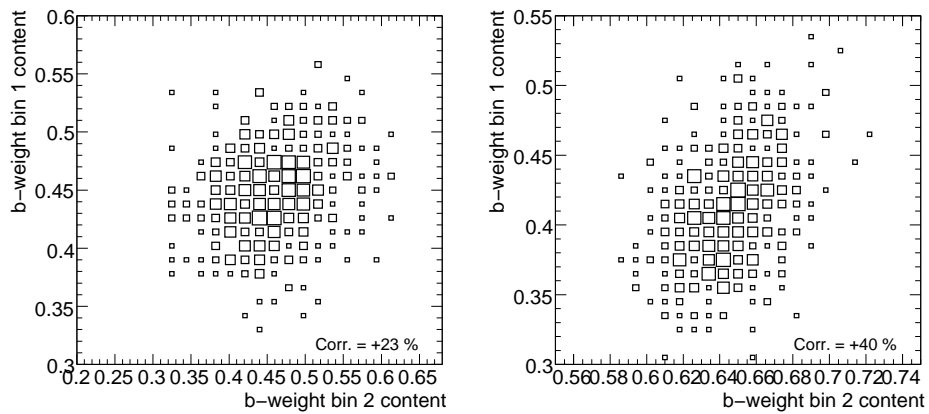


Figure 8.54: Correlations between systematic variations of the first two subsequent bins of the step function used for modelling the $W+\text{jet}$ (left) and the $t\bar{t}$ (right) background.

- When both upward and downward fluctuations for a certain systematic shift are considered, the corresponding variations of the parameters $\vec{\alpha}$ are just averaged together (summed up with weight 0.5). Due to the Gaussian assumption the final variations need in fact necessarily to be symmetric with respect to the original upward or downward fluctuation.

The first assumption is in general absolutely reasonable: many of the experimental parameters affected by systematic shifts (like shifts in the jet energy scale, in the jet resolution and in the b -tagging efficiency and rejection, which have the most important effect on this analysis) will be measured on data, so that assuming these measurements to be Gaussian distributed around the experimental errors is well motivated. The second and third assumptions are justified only if either the systematic shifts considered result in relatively small deviations of the shapes of the PDFs around their expected configuration in the nominal scenario or they affect the PDF parameters in a close to linear way. Most of the single systematic effects considered have indeed a very small impact on the shapes of the PDFs. For the effects where the impact is greatest, the parameters are still varied by in general less than relative 5 to 10%, so the linear approximation should work reasonably well.

8.8 Impact of systematics on the maximum likelihood fit

The main impact of the systematic uncertainties on the likelihood fit is related to the fact that the *true* distributions of the discriminating variables in data are not perfectly reproduced by the PDFs in the likelihood fit. The effect on the sensitivity of the analysis can be investigated by properly simulating the effect of the systematic uncertainties during the generation of the Monte Carlo pseudo-experiments. Therefore, for each pseudo-experiment, the expectation values of the parameters $\vec{\alpha}$ of the PDFs used for the generation are first randomly varied with respect to their nominal parameters, according to the multi-Gaussian PDF introduced in Eq. 8.29; the yield parameters of the signal and various background components are also included in the multi-Gaussian PDF, as considered in the previous section, to simulate possible effects of their variations.

In order to reduce the impact of such systematic uncertainties on the discovery sensitivity of the analysis, the statistical hypothesis test is also modified. The proper way to take systematic deformations of the PDFs into account in the likelihood fit, when using the profile likelihood ratio as a test statistics, is to describe them through the use of additional parameters in the likelihood function (usually called *nuisance parameters*). To do so, the parameters $\vec{\alpha}$ of the PDFs which were considered as fixed in the fit so far, are interpreted as nuisance parameters and left free to be determined on the fit data sample. In order to constraint these parameters in the limit of the expected systematic uncertainties, the multi-Gaussian term of Eq. 8.29 is used as an external constraint to the maximum likelihood fit, after having removed all the yield variables and the parameters of the $t\bar{t}$ and Wt PDFs which are determined directly on the control sample (mass and b -weight PDF). As a consequence the likelihood function of Eq. 8.16 to be maximised changes into:

$$\begin{aligned} \log(L_{TOT}) &= \log(L_{CS}) + \log(L_{SR}) + \log\left(L\left(\frac{N_{Wt}}{N_{Wt} + N_{t\bar{t}}}\right)\right) \\ &\quad - \sum_{t=0}^{N_{par}} \left(\frac{(\beta_t(\vec{\alpha}) - \beta_t(\vec{\alpha}^0))^2}{2\sigma_t^2} \right) \end{aligned} \quad (8.30)$$

The use of the profile likelihood ratio (Eq. 8.20) method in combination with the additional nuisance parameters permits to correctly account for the systematic uncertainties, due to the

separate maximisation of the likelihood function representing the signal plus background and background only hypotheses also in the space of possible values of the nuisance parameters. In addition, the nuisance parameters can also get constrained during the fit itself, while the correlations between the various parameters are correctly taken into account. This means for example that if, within the expected systematic uncertainty, the mass peak of the Z boson in the WZ background is significantly constrained by the fit sample itself, also all other fit parameters which are strongly affected by the same jet energy scale variation are constrained as well. Few validation tests have shown that this indeed works very reliably: as an example, the error on the Z mass peak provided by the fit on Monte Carlo toys generated with the signal plus background hypothesis is presented in Fig. 8.55, which shows that, starting from a systematic Z mass width of $\frac{\sigma(m_Z)}{m_Z} \approx 5\%$ (see Fig. 8.52), which reflects the expected systematic uncertainty contained in the multi-Gaussian constrain term, the fit itself can constrain the Z mass peak further down to a relative precision of $\approx 3.5\%$ in average, either directly being sensitive to the position of the Z mass peak itself or by exploiting the correlations to all other fit parameters which are also directly affected by jet energy scale variations.

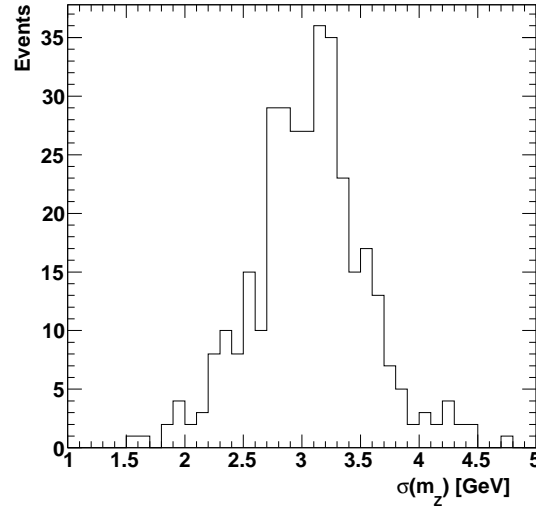


Figure 8.55: Error on the Z mass peak as provided by fits on Monte Carlo toys generated with background plus signal ($m_H = 120$ GeV). If the fit would not be able to constraint the Z mass peak further, the error would be fixed to ≈ 4.5 GeV, the upper limit which is given by the multi-Gaussian constraint term.

The method proposed here has, however, a clear drawback in terms of speed execution of a single likelihood fit: due to the increased complexity of the fit and to the increase in the number of free parameters from ≈ 20 to ≈ 100 , the execution speed of a single fit increases from around 1 minute to around 30 minutes (including the computation of the full hessian matrix of the likelihood function around the found maximum). This limits the amount of statistics which is possible to collect for the background only experiments needed to define the statistical hypothesis test.

The discovery sensitivity of the analysis was thus re-evaluated, including all experimental systematic uncertainties using the method just described. The results are shown in Figs. 8.56-8.58, for the three Higgs boson mass hypotheses considered in this study. The discovery potential of a low mass Higgs boson in the $WH \rightarrow \ell\nu b\bar{b}$ channel as a function of the Higgs mass is summarised in Fig. 8.59, together with the 1 and 2σ bands around the median values of the significances. It can be seen that, after including the effect of the experimental systematic

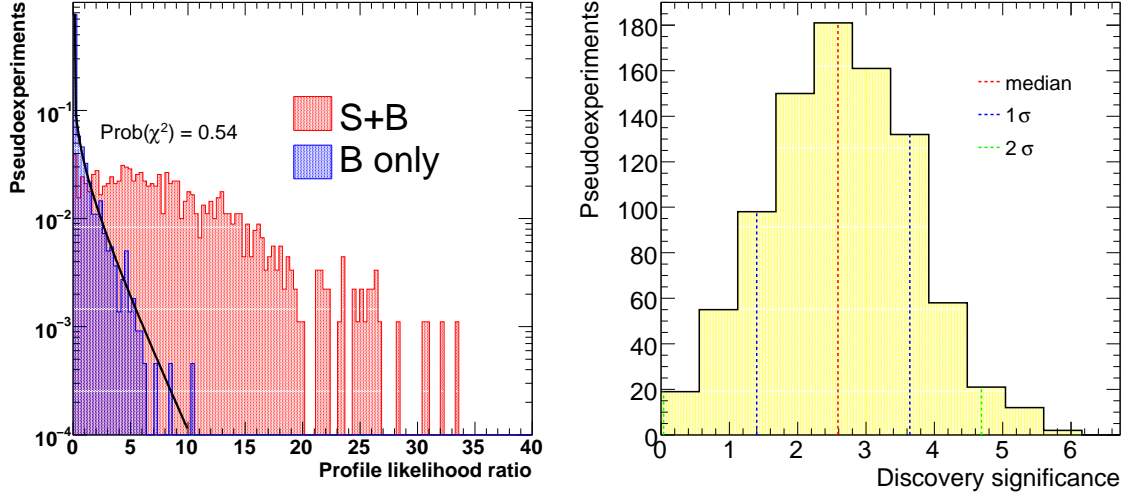


Figure 8.56: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 120$ GeV. The result of a χ^2 fit to the background only distribution by using a χ^2 function with one degree of freedom is also shown, together with its associated $\text{Prob}(\chi^2)$. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 120$ GeV.

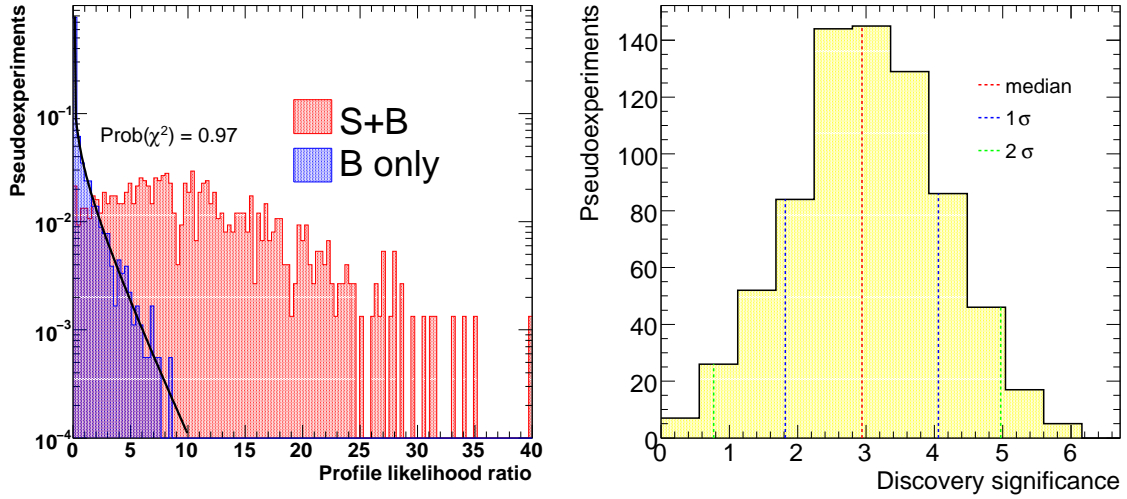


Figure 8.57: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 115$ GeV. The result of a χ^2 fit to the background only distribution by using a χ^2 function with one degree of freedom is also shown, together with its associated $\text{Prob}(\chi^2)$. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 115$ GeV.

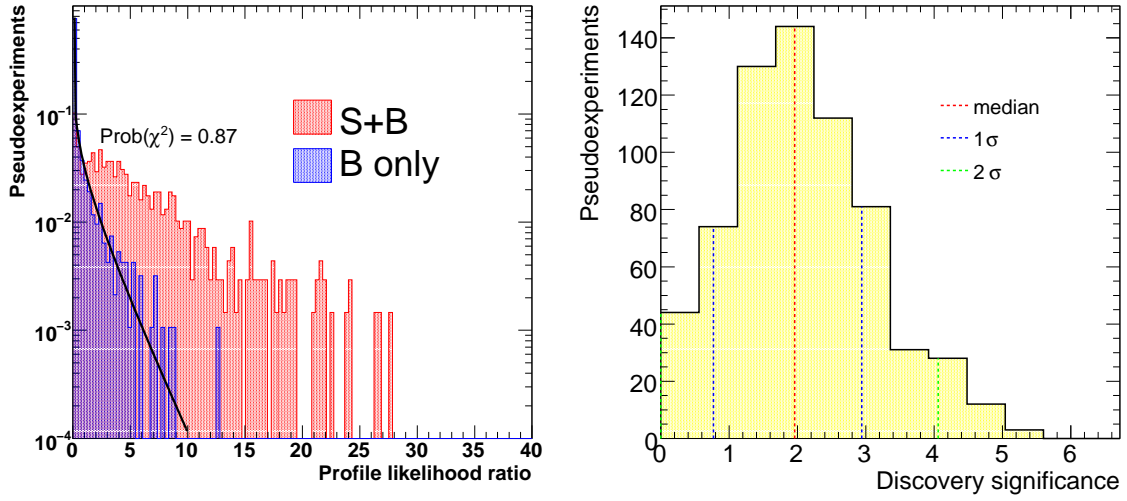


Figure 8.58: (Left) Distribution of $2 \cdot \log(PLR)$ for the background only and signal plus background pseudo-experiments corresponding to a $m_H = 130$ GeV. The result of a χ^2 fit to the background only distribution by using a χ^2 function with one degree of freedom is also shown, together with its associated $\text{Prob}(\chi^2)$. (Right) Expected distribution of significances as a result of the Monte Carlo pseudo-experiments for $m_H = 130$ GeV.

uncertainties, a median discovery significance of 3σ (2.6σ) is expected for a Higgs boson of 115 (120) GeV in mass with 30 fb^{-1} of integrated luminosity. It is also worth noticing that the biggest impact on the sensitivity is not originating from taking into account possible deformations of the PDFs used in the likelihood fit due to systematic effects, but it is mostly due to the degradations in the detector performance assumed moving from the *perfect* to the *nominal* scenario, as described in Section 8.6.

The final expected discovery sensitivity of the $WH \rightarrow \ell\nu b\bar{b}$ analysis for a Higgs boson with $m_H = 120$ GeV is higher than the discovery sensitivity assumed for the combination of all VH (W, Z) channels used as an input to the recent global fit analysis of the Higgs boson couplings presented in Ref. [49], once the uncertainty of 10% on the background level assumed in this analysis is taken into account. Since the VH search channels are found to play a crucial role in constraining the Higgs boson couplings, the result of the WH analysis represents an important indirect confirmation of the feasibility of the global fit analysis for Higgs boson masses around 120 GeV.

The effect of next-to-leading order QCD corrections and of additional theoretical uncertainties has not yet been included, but a dedicated study of the $Wb\bar{b}$ background process, for which a complete next-to-leading order Monte Carlo is missing, is ongoing. Furthermore, the experimental influence of pile-up events has not yet been fully considered: a first full simulation study, where the influence of pile-up has been taken into account, shows that the degradation of the mass resolution and of the b -tagging efficiency is very small, however the possible impact of pile-up on the jet veto will need to be carefully addressed.

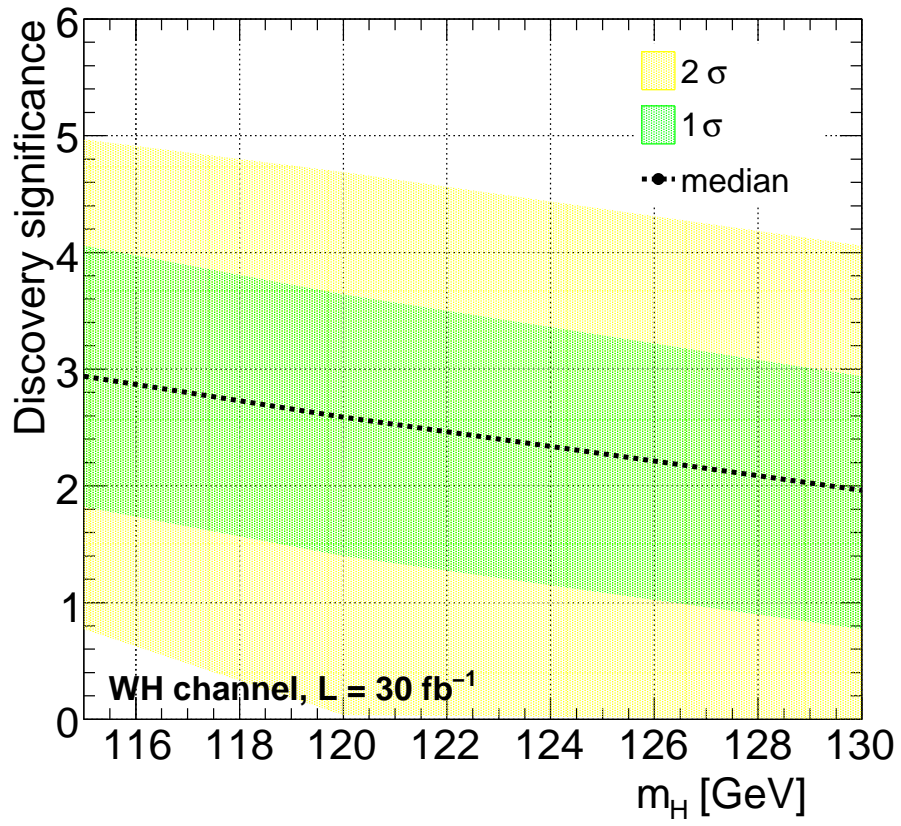


Figure 8.59: Discovery sensitivity as a function of the Higgs boson mass, for an integrated luminosity of 30 fb^{-1} , for the *nominal* scenario described in the text. All experimental systematic uncertainties have been included.

Summary and conclusions

In the course of the present thesis several new primary vertex finding and fitting methods to determine the primary interaction vertex were implemented in the ATLAS reconstruction software.

All vertex fitting algorithms rely on a Kalman filter based approach. One of the fundamental inputs to the Kalman filter is the linearisation of the equations of motion of the charged particle trajectories around a certain position in space: for the first time an exact computation of the related Jacobians was performed, which allows a faster convergence of the fit procedure. The first-order result was found to agree with previous calculations.

Due to the noisy environment of hadron collisions at the LHC, the determination of the primary vertex position needs to cope with substantial contamination from particles from pile-up events and from secondary interactions: adaptive finding methods were introduced to optimally reduce the bias on the fit result due to outlying measurements. Furthermore, a primary vertex finding algorithm was implemented, the adaptive multi-vertex finding algorithm, which deals with the problem of dynamically associating M charged particle tracks to N vertices and is therefore particularly useful in all cases where the distances between the vertices get comparable to the experimental resolutions of the charged particle tracks.

The primary vertex finding performance was evaluated for various physics processes and for different instantaneous luminosities, with an average number of in-time pile-up interactions between 0 and 6.9.

Without contamination from pile-up events, in $t\bar{t}$ events the primary vertex position can be determined with a precision of $30\text{ }\mu\text{m}$ in the longitudinal direction and of $9\text{ }\mu\text{m}$ in the transverse plane. The use of adaptive vertex finding methods permits to improve the core resolution in the z direction by $\approx 13\%$ and the root mean squared of the residual distribution up to $\approx 34\%$. Thanks to the implementation of a special $3d$ -vertex seeding method, the adaptive vertex finding algorithms can also deal correctly with a poor knowledge of the beam spot position in the transverse plane: the degradation of the vertex resolution in the z direction turns out to be marginal.

In the presence of up to 6.9 additional in-time interactions from pile-up, only a small degradation in the vertex reconstruction performance is seen: depending on the process considered, the resolution in the z direction is degraded up to 7% and the fraction of not correctly reconstructed primary interaction vertices is between 0.2% (for $t\bar{t}$ events) and 1.4% (for $H \rightarrow \gamma\gamma$ events). The fraction of events where the signal vertex is not correctly identified among all reconstructed primary vertices is significantly higher, strongly depending on the process considered: however, for physics events where the presence of a lepton with high transverse momentum ($p_T > 10\text{ GeV}/c$) can be required, this fraction can essentially be reduced to the mentioned intrinsic fraction of not correctly reconstructed primary vertices.

In addition, in the course of the present thesis a novel algorithm, denoted *JetFitter*, was implemented for the inclusive identification of b - to c -hadron cascades in b -jets, as opposed to the inclusive determination of a single b -/ c -vertex, aimed at improving the ability of ATLAS to identify b -jets. The algorithm is able to determine a variety of possible different decay chain topologies, thanks also to the possibility to reconstruct vertices with a single reconstructed charged particle track. On a set of inclusive $WH \rightarrow \ell\nu b\bar{b}$ events, the decay chain fit can

resolve the two b/c -hadron vertices with one or more charged particle tracks in $\approx 20\%$ of the cases.

The result of the decay chain fit, including the information about the decay chain topology, was used as input to two dedicated b -jet identification algorithms, the first one based on a likelihood ratio approach and the second one based on an artificial neural network. The b -jet identification performance, after combining the secondary vertex based algorithms with the best performing impact parameter based algorithm available in ATLAS, was evaluated on different physics samples of fully simulated events and in different kinematic ranges, and the results were compared to the standard combined algorithm available in ATLAS. As an example, for $t\bar{t}$ and $t\bar{t}jj$ events, if a b -jet identification efficiency of 60% is required, with the JetFitter algorithm based on the neural network approach only one every ≈ 230 u , d and s (light)-quark jets is misidentified as a b -jet, with a relative improvement of $\approx 15\%$ and $\approx 40\%$ compared to JetFitter based on the likelihood approach and to the standard combined algorithm, respectively.

The b -tagging algorithm based on JetFitter with the neural network approach was also specifically optimised to reject c -quark jets: as a result, the c -quark rejection can be increased, at the cost of a reduced light-quark rejection, by tuning a parameter which can be interpreted as the relative background composition in terms of c - and light-quark jets. As an example, if a b -jet identification efficiency of 60% is required, the c -jet rejection can be increased by up to 25%, while losing up to 50% of the light-quark rejection. Larger gains in c -jet rejection can be obtained for smaller b -jet efficiencies.

In the present thesis, a first detector level study of the $WH \rightarrow \ell\nu b\bar{b}$ Higgs boson search channel, where the W and Z bosons are required to have large transverse momenta and the Higgs boson decays into a pair of b -quarks, was presented. This study relies on Monte Carlo events generated with LO Monte Carlo generators and passed through the full simulation of the ATLAS inner detector and muon system and through a full-granularity fast simulation of the calorimeter.

The identification of the Higgs candidate relies on a recently proposed algorithm [1] where the two b -subjects are reconstructed in a single large jet and the two subjects are then searched for by using an iterative splitting procedure: after the jet energy calibration is applied, a good mass resolution of $\approx 10\%$ is obtained, which is sufficient to distinguish Higgs bosons with a mass above $115 \text{ GeV}/c^2$ from the nearby Z boson peak originating from WZ events, which are an irreducible background for this search channel.

To reject the large W +light jet background, a very good b -jet identification performance is requested on both subjects, despite their proximity. Both the JetFitter and standard combined algorithms were considered to be applied to the subjects from the Higgs candidate and the performance was evaluated on the same event samples as used for the analysis: if a b -subject identification efficiency of 70% is required, a remarkable rejection against light-quark jets of ≈ 100 can be obtained with JetFitter. After b -tagging is applied, the dominant background contributions are given by the $Wb\bar{b}$ process and by $t\bar{t}$ events. In the second case the Higgs boson is mimicked by the combination of a b - and a c -jet. By specifically optimising the JetFitter algorithm to reject c -jets, the signal-to-background ratio of the analysis was improved further.

The statistical significance of the $WH \rightarrow \ell\nu b\bar{b}$ analysis based on pure event counting, in terms of $\frac{S}{\sqrt{B}}$, was found to be 3.0 ± 0.3 , assuming a Higgs boson mass of $120 \text{ GeV}/c^2$, leading order cross sections, a centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$ and an integrated luminosity of 30 fb^{-1} .

The systematic uncertainty on the estimated number of background events can significantly decrease the discovery significance with respect to the quoted statistical significance. In

this thesis a maximum likelihood fit was developed, based on the signal and background distributions of four uncorrelated discriminating variables, in order to extract the signal and the main background contributions directly from data and, therefore, to basically remove this source of uncertainty. The statistical hypothesis test is based on the profile likelihood ratio and the estimation of the median discovery significance relies on a large set of background-only and signal-plus-background Monte Carlo pseudo-experiments.

An important source of systematic uncertainties is given by possible differences in the shapes of the distributions of the discriminating variables for the signal or the various background components between simulated and real data. Control sample are therefore defined for the $t\bar{t}$ and Wt backgrounds, to extract the probability density functions (PDFs) for most of their discriminating variables directly from data, in regions of the phase space where a negligible signal contribution is expected. The distributions in the control sample and in the signal region were found to be compatible in the limit of the available Monte Carlo statistics. The systematic uncertainty on the shapes of the remaining PDFs is considered by adding corresponding nuisance parameters to the likelihood function and by constraining these parameters in the limit of the variations expected by considering shifts in the energy scale, in the b -tagging efficiency and rejection, in the lepton reconstruction and identification efficiency and several other experimental systematic variations. In addition, a novel method was derived, to take all correlations between the systematic variations of the PDFs due to common systematic uncertainties correctly into account.

The final result is that, including the effect of experimental systematic uncertainties, the expected median discovery sensitivity to a Standard Model Higgs boson is $\approx 3\sigma$ for $m_H = 115 \text{ GeV}/c^2$, $\approx 2.6\sigma$ for $m_H = 120 \text{ GeV}/c^2$ and $\approx 2.0\sigma$ for $m_H = 130 \text{ GeV}/c^2$, based on leading order cross sections, on a centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$ and on an integrated luminosity of 30 fb^{-1} .

The Higgs $\rightarrow b\bar{b}$ decay mode has been thus re-established as a discovery mode for Higgs bosons at the LHC. In addition this channel turns out to be extremely important to extract the Higgs boson couplings.

The effect of next-to-leading order QCD corrections and of additional theoretical uncertainties has not yet been included, but a dedicated study of the $Wb\bar{b}$ background process, for which a complete next-to-leading order Monte Carlo is missing, is ongoing. Furthermore, the experimental influence of pile-up events has not yet been fully considered: the possible impact of pile-up on the jet veto will need to be carefully addressed.

A Primary vertex reconstruction

A.1 Details on computation of track parameterisation

A.1.1 Computation of Jacobian derivatives

In Section 5.3.3, in order to get the Jacobians A and B to feed into Eq. 5.16, the derivatives of $(d_0, z_0, \phi_P, \theta_P, q/p)$ with respect to $(x_V, y_V, z_V, \phi_V, \theta, q/p)$ need to be computed, as explicitly derived here:

$$\begin{aligned}
 \frac{\partial d_0}{\partial x_V} &= -\text{sgn}(\rho) \frac{X}{S} \\
 \frac{\partial d_0}{\partial y_V} &= -\text{sgn}(\rho) \frac{Y}{S} \\
 \frac{\partial d_0}{\partial z_V} &= 0 \\
 \frac{\partial d_0}{\partial \phi_V} &= -\rho \text{sgn}(\rho) \frac{(x_V - x_R) \cos(\phi_V) + (y_V - y_R) \sin(\phi_V)}{S} \\
 \frac{\partial d_0}{\partial q/p} &= -\frac{\rho}{q/p} \left[1 - \text{sgn}(\rho) \frac{((x_V - x_R) \sin(\phi_V) + (y_V - y_R) \cos(\phi_V))}{S} \right] \\
 \frac{\partial d_0}{\partial \theta} &= \frac{\rho}{\tan(\theta)} \left[1 - \text{sgn}(\rho) \frac{((x_V - x_R) \sin(\phi_V) + (y_V - y_R) \cos(\phi_V))}{S} \right]
 \end{aligned} \tag{A.1}$$

$$\begin{aligned}
 \frac{\partial \phi_P}{\partial x_V} &= -\frac{Y}{S^2} \\
 \frac{\partial \phi_P}{\partial y_V} &= \frac{X}{S^2} \\
 \frac{\partial \phi_P}{\partial z_V} &= 0 \\
 \frac{\partial \phi_P}{\partial \phi_V} &= \frac{\rho^2 + \rho((x_V - x_R) \sin(\phi_V) - (y_V - y_R) \cos(\phi_V))}{S^2} \\
 \frac{\partial \phi_P}{\partial q/p} &= \frac{\rho}{q/p} \frac{(x_V - x_R) \cos(\phi_V) + (y_V - y_R) \sin(\phi_V)}{S^2} \\
 \frac{\partial \phi_P}{\partial \theta} &= -\frac{\rho}{\tan(\theta)} \frac{(x_V - x_R) \cos(\phi_V) + (y_V - y_R) \sin(\phi_V)}{S^2}
 \end{aligned} \tag{A.2}$$

$$\begin{aligned}
 \frac{\partial z_0}{\partial x_V} &= -\frac{\rho}{\tan(\theta)} \frac{\partial \phi_P}{\partial x_V} \\
 \frac{\partial z_0}{\partial y_V} &= -\frac{\rho}{\tan(\theta)} \frac{\partial \phi_P}{\partial y_V} \\
 \frac{\partial z_0}{\partial z_V} &= 1 \\
 \frac{\partial z_0}{\partial \phi_V} &= \frac{\rho}{\tan(\theta)} \left[1 - \frac{\partial \phi_P}{\partial \phi_V} \right]
 \end{aligned} \tag{A.3}$$

$$\begin{aligned}\frac{\partial z_0}{\partial q/p} &= -\frac{\rho}{\tan(\theta)} \left[\frac{1}{q/p} (\phi_V - \phi_p(x_V, y_V, \phi_V, \theta, q/p)) + \frac{\partial \phi_P}{\partial q/p} \right] \\ \frac{\partial z_0}{\partial \theta} &= -\rho \left[\phi_V - \phi_p(x_V, y_V, \phi_V, \theta, q/p) + \frac{1}{\tan(\theta)} \frac{\partial \phi_P}{\partial \theta} \right],\end{aligned}$$

while the derivatives with respect to θ and q/p are all zero, except for $\frac{\partial q/p}{\partial q/p}$ and $\frac{\partial \theta}{\partial \theta}$ which are trivially equal to 1.

These derivates can be further simplified, by using the following identities:

$$\begin{aligned}(y_V - y_R) \sin(\phi_V) + (x_V - x_R) \cos(\phi_V) &= \\ (y_V - y_R - \rho \cos(\phi_V)) \sin(\phi_V) + (x_V - x_R + \rho \sin(\phi_V)) \cos(\phi_V) &= \\ Y \sin(\phi_V) + X \cos(\phi_V)\end{aligned}\tag{A.4}$$

$$\begin{aligned}\rho^2 + \rho((x_V - x_R) \sin(\phi_V) - (y_V - y_R) \cos(\phi_V)) &= \\ \rho[(x_V - x_R + \rho \sin(\phi_V)) \sin(\phi_V) - (y_V - y_R - \rho \cos(\phi_V)) \cos(\phi_V)] &= \\ \rho[X \sin(\phi_V) - Y \cos(\phi_V)].\end{aligned}\tag{A.5}$$

The final result for the Jacobians A and B , in a more compact form, is obtained in Section 5.3.3.

A.1.2 Approximation in the limit of large curvatures

The computation of Jacobians presented in the Section 5.3.3 contains no approximation. The computation of the same Jacobian was already available in Ref. [61], but only in the limit of large curvatures ($|\rho| \gg |d_0|$). This corresponds to performing a Taylor expansion of the result presented in Eq. 5.36 and Eq. 5.37 as a function of $\frac{1}{\rho}$ and then keeping only the terms up to first order, $O\left(\frac{1}{\rho}\right)$. While the result presented here is exact up to all orders in $\frac{1}{\rho}$, it is very useful to do such an expansion and compare the first order result as a first cross-check of the new result.

For the position Jacobian one obtains:

$$\begin{aligned}A &= \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(x_V, y_V, z_V)} \\ &= \begin{pmatrix} -\sin(\phi_V) & \cos(\phi_V) & 0 \\ -\cos(\phi_V) \cot(\theta) & -\sin(\phi_V) \cot(\theta) & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &+ \begin{pmatrix} -\cos(\phi_V)[x_V \cos(\phi_V) + y_V \sin(\phi_V)] & -\sin(\phi_V)[x_V \cos(\phi_V) + y_V \sin(\phi_V)] & 0 \\ \cot(\theta)[-y_V \cos(2\phi_V) + x_V \sin(2\phi_V)] & -\cot(\theta)[x_V \cos(2\phi_V) + y_V \sin(2\phi_V)] & 0 \\ \cos(\phi_V) & \sin(\phi_V) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{1}{\rho} + O\left(\frac{1}{\rho^2}\right)\end{aligned}\tag{A.6}$$

The same expansion applied on the momentum Jacobian provides the following result:

$$\begin{aligned}
 B &= \frac{\partial(d_0, z_0, \phi_P, \theta_P, q/p)}{\partial(\phi_V, \theta, q/p)} \\
 &= \begin{pmatrix} -\tilde{R} & 0 & 0 \\ \cot(\theta)\tilde{Q} & \frac{\tilde{R}}{\sin^2(\theta)} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 &+ \begin{pmatrix} \tilde{Q} \cdot \tilde{R} & \frac{1}{2}\cot(\theta)\tilde{R}^2 & -\frac{1}{2\frac{q}{p}}\tilde{R}^2 \\ -\cot(\theta)\tau & -\frac{1}{2}\frac{3+\cos(2\theta)}{\sin^2(\theta)}\tilde{R} \cdot \tilde{Q} & \frac{\cot(\theta)}{\frac{q}{p}}\tilde{R} \cdot \tilde{Q} \\ -\tilde{Q} & -\cot(\theta)\tilde{R} & \frac{1}{\frac{q}{p}}\tilde{R} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{1}{\rho} + O\left(\frac{1}{\rho^2}\right)
 \end{aligned} \tag{A.7}$$

where \tilde{R}, \tilde{Q} and τ are defined as:

$$\begin{aligned}
 \tilde{R} &= (y_V - y_R) \sin(\phi_V) + (x_V - x_R) \cos(\phi_V) \\
 \tilde{Q} &= (x_V - x_R) \sin(\phi_V) - (y_V - y_R) \cos(\phi_V) \\
 \tau &= [(x_V^2 - y_V^2) \cos(2\phi_V) + 2x_V y_V \sin(2\phi_V)]
 \end{aligned} \tag{A.8}$$

When comparing this result with the computation presented in Ref. [61], three differences should be considered:

- The ATLAS convention for the sign of the transverse impact parameter d_0 and for the curvature ρ is opposite to the one used in Ref. [61].
- The definition used for R and Q is slightly different.
- In the ATLAS tracking model the perigee parameters are given in terms of $(d_0, z_0, \phi_p, \theta_p, q/p)$, while in Ref. [61] the parameter q/p is replaced with the inverse curvature itself.

While it is straightforward to take the first two differences into account, the third one must be dealt with by relating the Jacobians expressed in the two different parameterizations; this can be done by multiplying the submatrices of A and B involving $(\theta_p, q/p)$ by an additional Jacobian $\frac{\partial(\theta, q/p)}{\partial(\theta, 1/\rho)}$. After considering these three small differences, the result in Eq. A.6 and Eq. A.7 agree perfectly with the result of Ref. [61]¹.

A.2 Implementation in the ATLAS software ATHENA

In this section of the appendix details will be given about how the vertex finding and fitting algorithms described in Section 5.3 and Section 5.4 are concretely realized.

¹In this article there is a sign error in one of the elements of the momentum Jacobian, which is corrected by an *erratum* [100] published few years later.

A dedicated framework, based on the C++ programming language, is implemented in the ATLAS software ATHENA for the reconstruction of the primary event vertex, which was put in place before the start of the present thesis work [101]. This framework was significantly extended to add all features requested by the adoption of *adaptive* finding and fitting methods and of the simultaneous fit of more than one primary vertex and will be thus described in some detail here. The framework includes:

- A set of *Event Data Model* (EDM) classes, which contain all vertex fit related informations (vertices, associated tracks, fit quality,...), hierarchically structured, which are filled up during reconstruction and then stored in the event output.
- A set of algorithms and related tools, which deal with the vertex finding or fitting tasks.

All algorithms and tools are implemented in a very flexible way. They usually inherit from abstract interfaces, where every abstract interface corresponds to a well defined task, with predefined input and output informations. As a consequence, more than one concrete implementation of a tool or of an algorithm can exist, so that the user running the ATLAS reconstruction software can easily switch from a software component to a different one, by just changing external configuration files.

A.2.1 Event Data Model

The classes containing the vertex reconstruction Event Data Model are designed in such a way that they can hold the output information produced by any present or future vertex finder. In addition these classes contain few additional informations, which are temporarily needed by the vertexing algorithms during their internal processing, but are not stored in the output as a default, like for example the result of the linearization of a track around an approximate vertex position.

In order to visualize the structure of the adopted object oriented Event Data Model, the basic classes are shown in Fig. A.1 using an UML (Unified Modeling Language) diagram. The output of any primary vertex finding algorithm is a vector of reconstructed vertex candidates (*VxCandidate*). The *VxCandidate* contains the informations about the reconstructed vertex position and quality (*RecVertex*), about the tracks associated to the vertex (vector of *VxTrackAtVertex*) and a *VertexType*, which informs about the nature of the vertex (signal, pile-up, secondary,...). The *RecVertex* contains the vertex position (which is actually stored in the base class *Vertex*), the vertex covariance matrix and the fit quality (*FitQuality*), in terms of value of χ^2 and number of degrees of freedom. The associated tracks to the vertex (*VxTrackAtVertex*) contain:

- the fit quality information (χ^2 and number of degrees of freedom corresponding to the compatibility of the track to the fitted vertex)
- the track weight (used only in the adaptive fit and corresponding to the weight ω of Eq. 5.45)
- the vertex compatibility (again used only in the adaptive fit and corresponding to the value of $\hat{\chi}^2$ to be used in Eq. 5.45)
- the track parameters after the vertex fit (where usually the tracks are constrained to pass through the vertex position)
- the linearized track

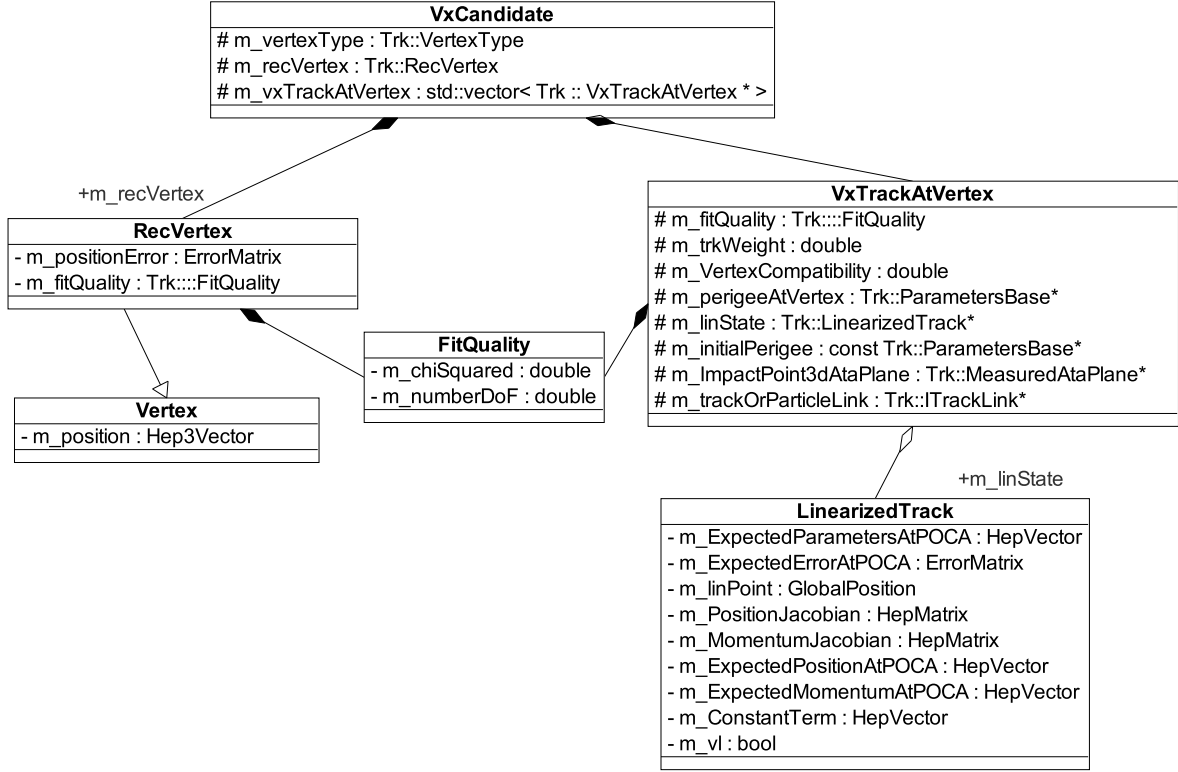


Figure A.1: UML diagram for the basic classes containing the vertex information in the ATLAS Event Data Model.

- the original track parameters (as provided by the tracking algorithm)
- a special version of the track parameters, defined at the point of closest approach to the actual vertex position in three dimensions (used only in the adaptive fit in order to get a fast estimate of the value of $\hat{\chi}^2$ which does not require the smoothing step described in Section 5.3.2)
- a link to the original track in the tracking collection

For all vertex fitting algorithms implemented in ATLAS, and in particular for the Billoir (Section 5.3.1) and Kalman filter (Section 5.3.2) methods described in this thesis, the vertex fit relies on the use of the linearised measurement equation of Eq. 5.16. All informations about the track after it has been linearised around a certain point in space, as described at the end of Section 5.3.3, are contained in the *LinearizedTrack* class, which contains the perigee parameters at the linearization point, the related covariance matrix, the linearization point, the position and momentum jacobian (matrices A and B of Eq. 5.7), the expected position and momentum at the linearization point (vectors \vec{r}_0 and \vec{p}_0 of Eq. 5.7), the constant term of the linearization equation (vector \vec{C} of Eq. 5.7) and finally a simple bool stating whether or not the content of the *LinearizedTrack* object is valid.

The object classes described so far need to be further extended in order to support the simultaneous fit of more vertices, which is required by the adaptive multi-vertex fit: such an extension is shown in Fig. A.2, again in the form of an UML diagram. In particular, what needs to be stored, at least temporarily during the fit, is the association between tracks and vertices. This is obtained by extending the information of the tracks associated to the vertex

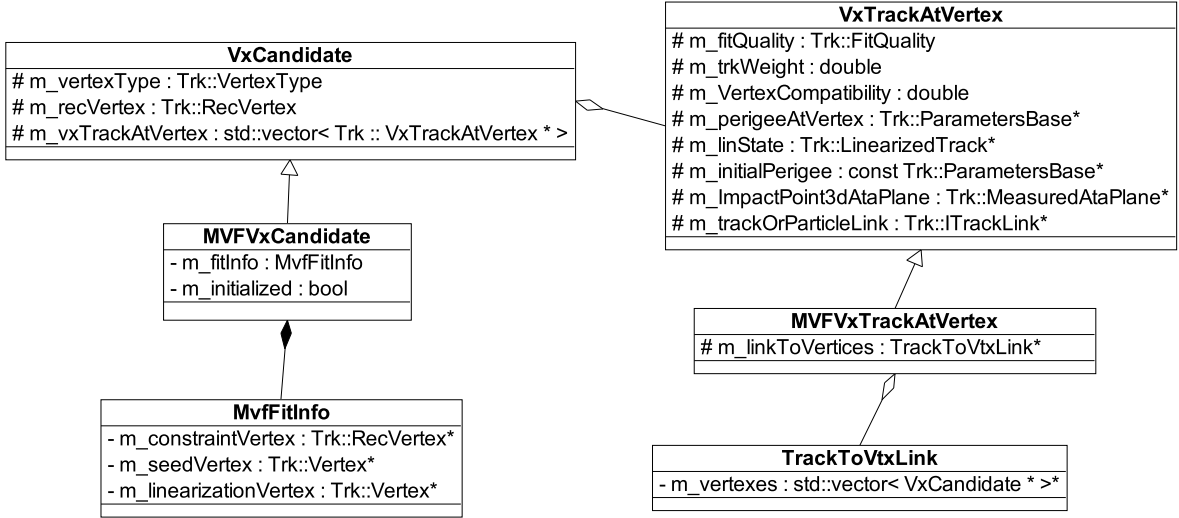


Figure A.2: UML diagram for the classes which extend the original vertex Event Data Model to support the simultaneous fit of more vertices, as implemented in the adaptive multi-vertex fitter.

(stored in *VxTrackAtVertex*) by the *MVFVxTrackAtVertex* class, which contains an additional object, the *TrackToVtxLink* class. If the same original track is contributing to more vertex candidates, they will all have a different *MVFVxTrackAtVertex*, but these tracks associated to their respective vertices will have a common link to the original track and a common *TrackToVtxLink* object, containing the link to all vertex candidates the tracks is competing to be assigned to.

In this way, during the adaptive multi-vertex fit, starting from a single vertex candidate, it is possible, after some multiple iterations over all vertex tracks, to navigate to all other vertex candidates which they share tracks with. This allows to collect all the vertex compatibility factors $\hat{\chi}^2$ needed to calculate the weights ω of Eq. 5.46 for every track associated to each vertex candidate to be fitted in the simultaneous fit. In addition also the vertex candidate (*VxCandidate* object) is extended by the *MVFVxCandidate* class, which contains also the vertex constraint (typically beam spot position + covariance matrix), the initial seed for the vertex position and the vertex position used for the last track linearization step. These informations are very useful to be kept during the vertex fit, but do not necessarily need to be stored in the event data output.

A.2.2 Vertexing algorithms and tools

The *InDetPriVxFinder* algorithm is the main algorithm governing primary vertex finding in ATLAS. As shown in Fig. A.3, it can be configured to make use of any of the primary vertex finding tools in ATLAS: *InDetPriVxFinderTool* (the default vertex finder described in Section 5.4.1), *InDetAdaptivePriVxFinderTool* (the adaptive vertex finder described in Section 5.4.3 and the *InDetAdaptiveMultiPriVxFinderTool* (the adaptive multi-vertex finder described in Section 5.4.4). The sequence diagram in Fig. A.4 shows schematically how the *InDetPriVxFinder* algorithm works: it retrieves the input tracks provided by the tracking algorithm, it passes them to the vertex finding tool of choice and, after all primary vertices in the event have been found, it stores them back in the event store (which in ATLAS is called *StoreGate*).

Every primary vertex finder interacts with the vertex fitter of choice to determine the ver-

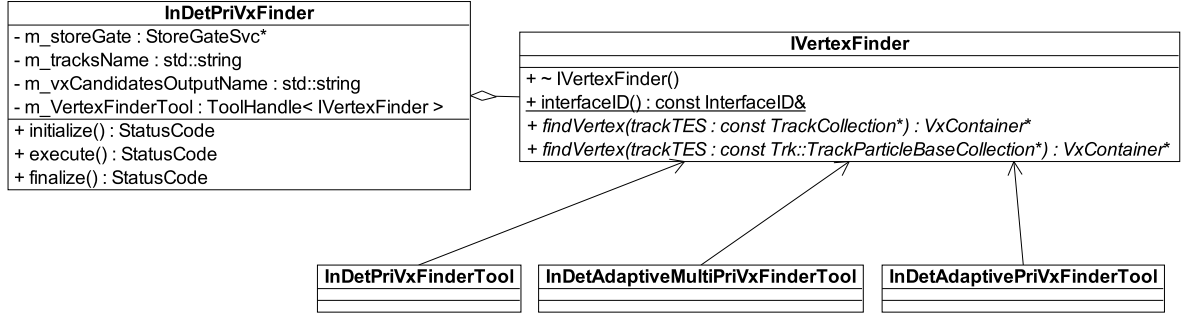


Figure A.3: The main algorithm governing primary vertex finding in ATLAS (*InDetPriVxFinder*) and the three available vertex finding tools it can be interfaced to.

tex positions. These vertex fitting tools all inherit from the base class *IVertexFitter*, except for the adaptive multi-vertex fitter (*AdaptiveMultiVertexFitter*), which requires a dedicated implementation and is more tightly bound to its related vertex finding algorithm (*InDetAdaptiveMultiPriVxFinderTool*). The vertex fitting tools are designed in a very modular way, subdivided into several different tasks, which can be easily exchanged and re-configured differently.

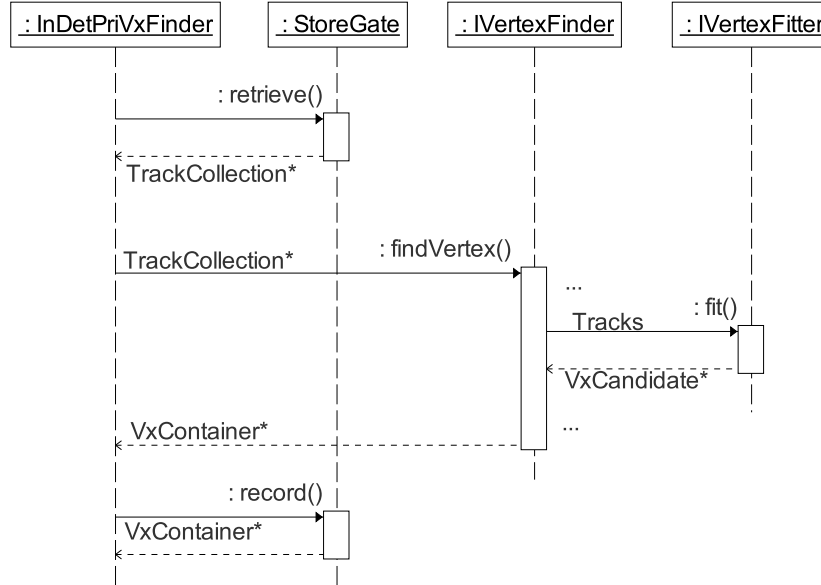


Figure A.4: Sequence diagram for the *InDetPriVxFinder* algorithm.

The Adaptive Vertex Fitter is an example of such a modular implementation, as shown in Fig. A.5, where the *AdaptiveVertexFitter* class with its properties and main methods is shown, together with all helper classes, which are used by the vertex fitter to accomplish specific tasks. Some of the helper classes are in common with other vertex fitting tools: for example, the *IVertexLinearizedTrackFactory* is in common among all vertex fitting algorithms; the *IVertexUpdater* and the *IVertexSmoother* deal with the Kalman update and smoothing step which are typical for a Kalman filter, as described in Section 5.3.2 and are in common with all Kalman Filter based fitting algorithms, while other classes are specific for the adaptive

vertex fitter, as the *ImpactPoint3dAtaPlane*, the *IVertexTrackCompatibilityEstimator* and the *IVertexAnnealingMaker*. To understand more in detail how the adaptive vertex fitter, conceptually described in Section 5.3.5, is realized in practice, it is worth giving a look at its sequence diagram (Fig. A.6), which shows how this algorithm works and interacts with the helper classes.

As a first step, the adaptive vertex fitter provides a set of tracks to the *IVertexSeedFinder*, which elaborates them according to the concrete seed finding algorithm of choice (e.g. *z*-scan or full 3d based) and returns the seed vertex position. According to this seed vertex, the tracks, already in the form of *VxTrackAtVertex*, i.e. of tracks associated to the vertex to be fitted, are prepared for the fit. On one side the *IVertexLinearizedTrackFactory* is used to linearize the tracks around the seed vertex position and the result is stored in the *LinearizedTrack* associated to its respective *VxTrackAtVertex* for later use. On the other side the *ImpactPoint3dAtaPlaneFactory* is used to define the track parameters at the point of closest approach in three dimensions to the seed vertex position, which are later used to get a fast estimate of the compatibility of the track to the vertex (i.e. an estimation of the value of

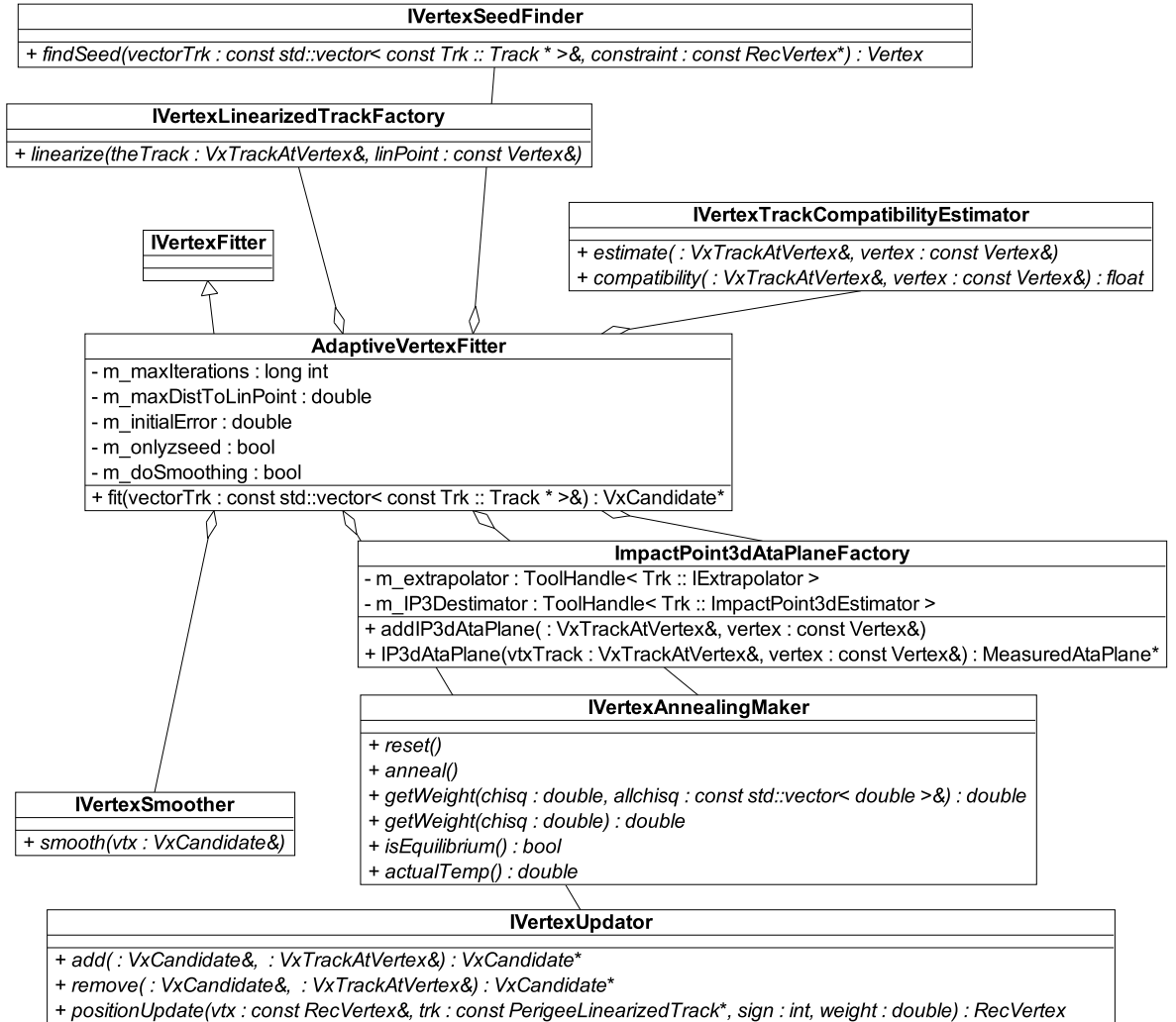


Figure A.5: The *AdaptiveVertexFitter* class and the vertex fitting tools it relies on.

$\hat{\chi}^2$): these parameters are stored in the *ImpactPoint3dAtaPlane* object, which is again associated to the *VxTrackAtVertex*. After the preparation phase, the fit iterations are started. At each further iteration, the *IVertexAnnealingMaker* is called, to decrease the temperature of the adaptive annealing procedure: the iterations are stopped when the annealing procedure reaches its last step. During a single fit iteration, the vertex is updated by iterating the Kalman update step over all tracks. For each track the *IVertexTrackCompatibilityEstimator* is first called, which uses the *ImpactPoint3dAtaPlane* of the track to update the *VxTrackAtVertex* with the new value of track to vertex compatibility $\hat{\chi}^2$, which is stored in its *VertexCompatibility* property. Then the *IVertexAnnealingMaker* is called, which, given the actual temperature of the annealing scheme and the value of the track $\hat{\chi}^2$, returns a value for the track weight ω , which is stored in the *trkWeight* associated with the same *VxTrackAtVertex*: as a default for the track weight computation the formula of Eq. 5.45 is used. Finally the *IVertexUpdater* is called, which updates the actual vertex candidate (*VxCandidate*) by performing the real Kalman update step, which uses as input the *LinearizedTrack* and the *trkWeight* associated to the *VxTrackAtVertex*. After all fit iterations have been completed, the *IVertexSmoother* is called, which iterates the Kalman smoothing step over all tracks and fills all *VxTrackAtVertex* objects with the refitted track parameters (*perigeeAtVertex*) and the correct values of track χ^2 and number of degrees of freedom (*fitQuality*).

In the case of the simultaneous adaptive fit of more vertices, a tighter integration of the ver-

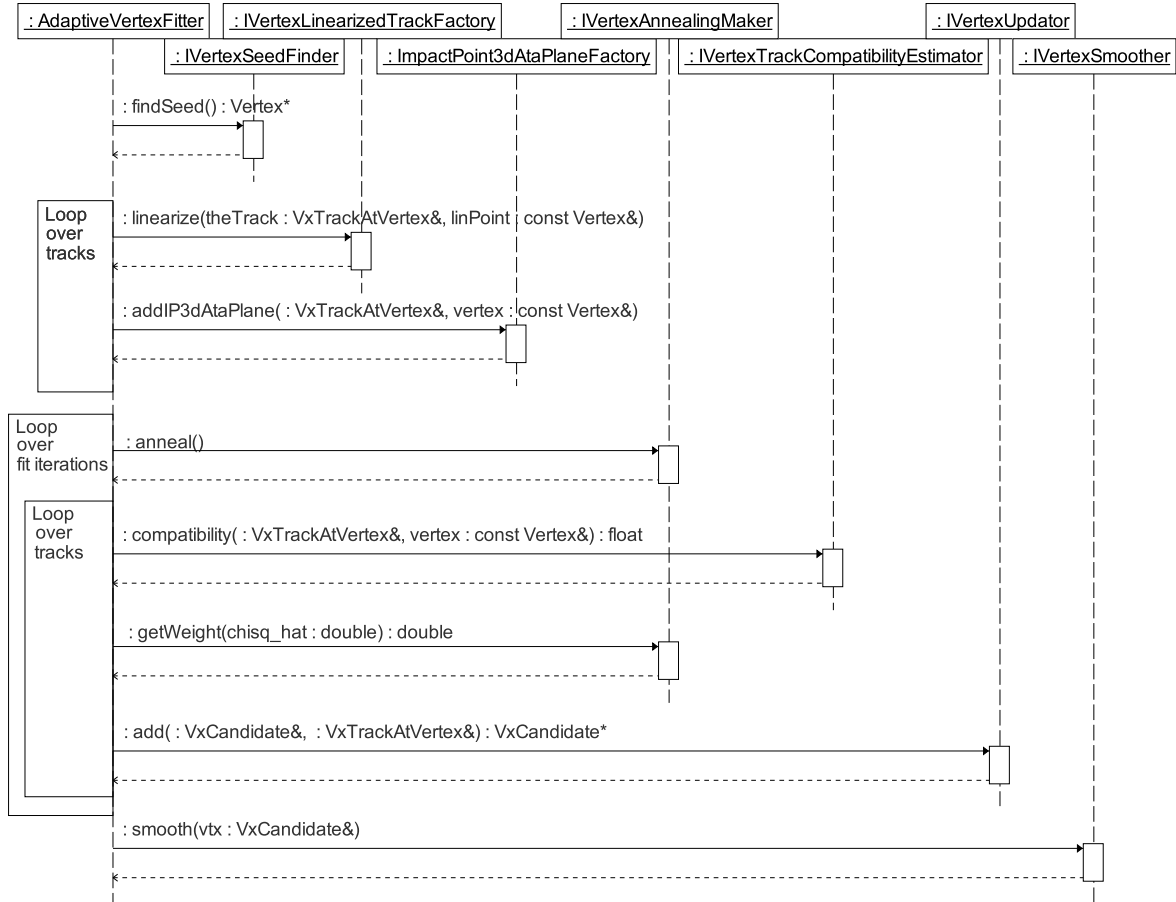


Figure A.6: Sequence diagram for the *AdaptiveVertexFitter* vertex fitting tool.

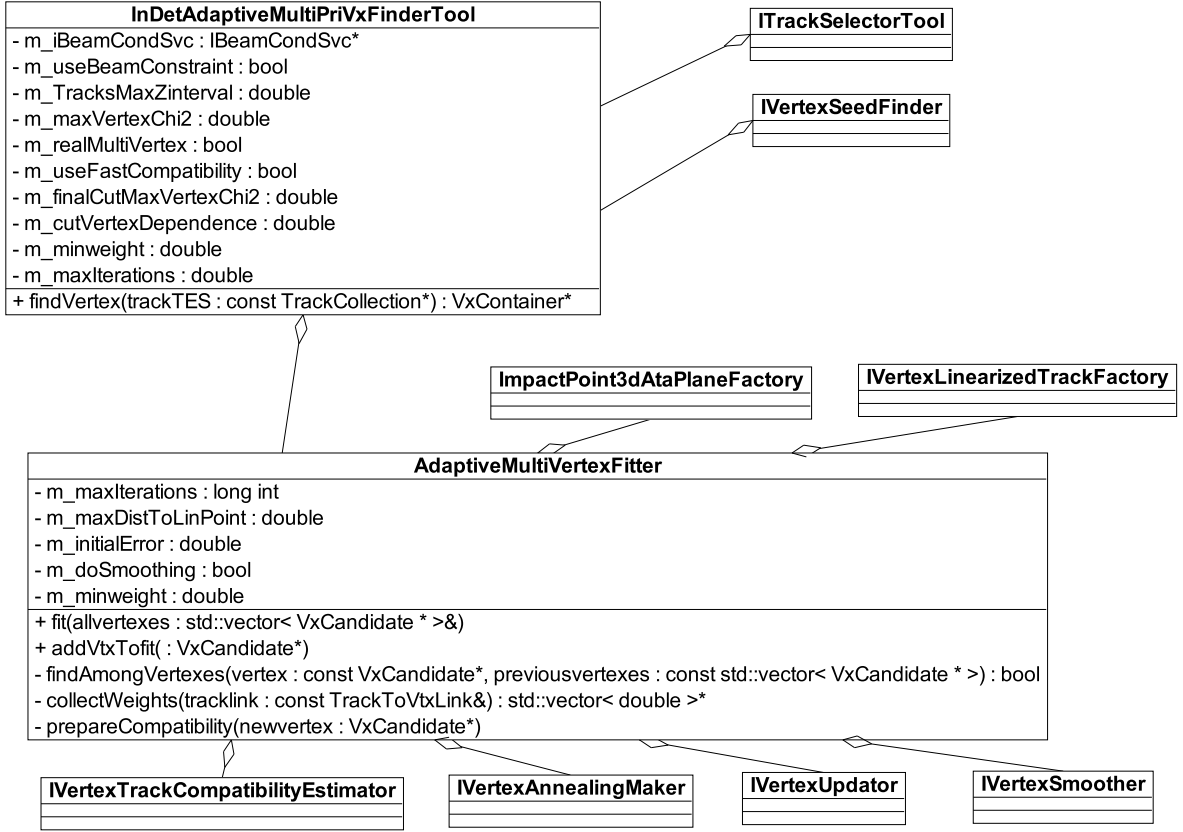


Figure A.7: The adaptive multi-vertex finding and fitting classes with all helper classes it relies on.

tex finding and fitting tools is realized; however, as shown in the UML diagram of Fig. A.7, the same fitting helper tools used in the case of the single-vertex fit can be re-used. The adaptive multi-vertex finder (*InDetAdaptiveMultiPriVxFinderTool*), as all the remaining vertex finding algorithms, deals directly with the track selection (calling the *ITrackSelectorTool*), which is needed to ensure the input tracks are of sufficiently good quality, and with the seed finder tool (*IVertexSeedFinder*), while all remaining helper classes are dealt with directly by the *AdaptiveMultiVertexFitter*, analogously to the adaptive single-vertex fitter. A simplified scheme about how the multi-vertex finding algorithm works is shown in Fig. A.8: this explains how the multi-vertex finding described conceptually in Section 5.3.5 is realized in practice.

After track selection, the multi-vertex finder iterates on the vertex seeds found iteration after iteration and passes them to the *AdaptiveMultiVertexFitter*, which, at each new iteration, tries to add the new vertex to the old ones and does a simultaneous fit of all vertices. Every new vertex needs to be prepared for the vertex fit; this is achieved by the *prepareCompatibility()* method, which cares about track linearization and estimation of track parameters at point of closest approach in *3d*. Then the simultaneous fit of all vertices is performed. This is analogous to the single-vertex fit case of Fig. A.6 (therefore some steps as calling the *IVertexAnnealingMaker::anneal()* method are not shown for simplicity), with the difference that all vertices need to be updated at every annealing iteration and that the weight ω_{vtx} of a certain track at a certain vertex is computed by a dedicated method of the *IVertexAnnealingMaker* class, which takes into account not only the compatibility of the track of interest to the vertex to be updated, but also to all other vertices the track is competing to be assigned

to, as required by Eq. 5.46. The vertex finding stops when no new seed vertices are found.

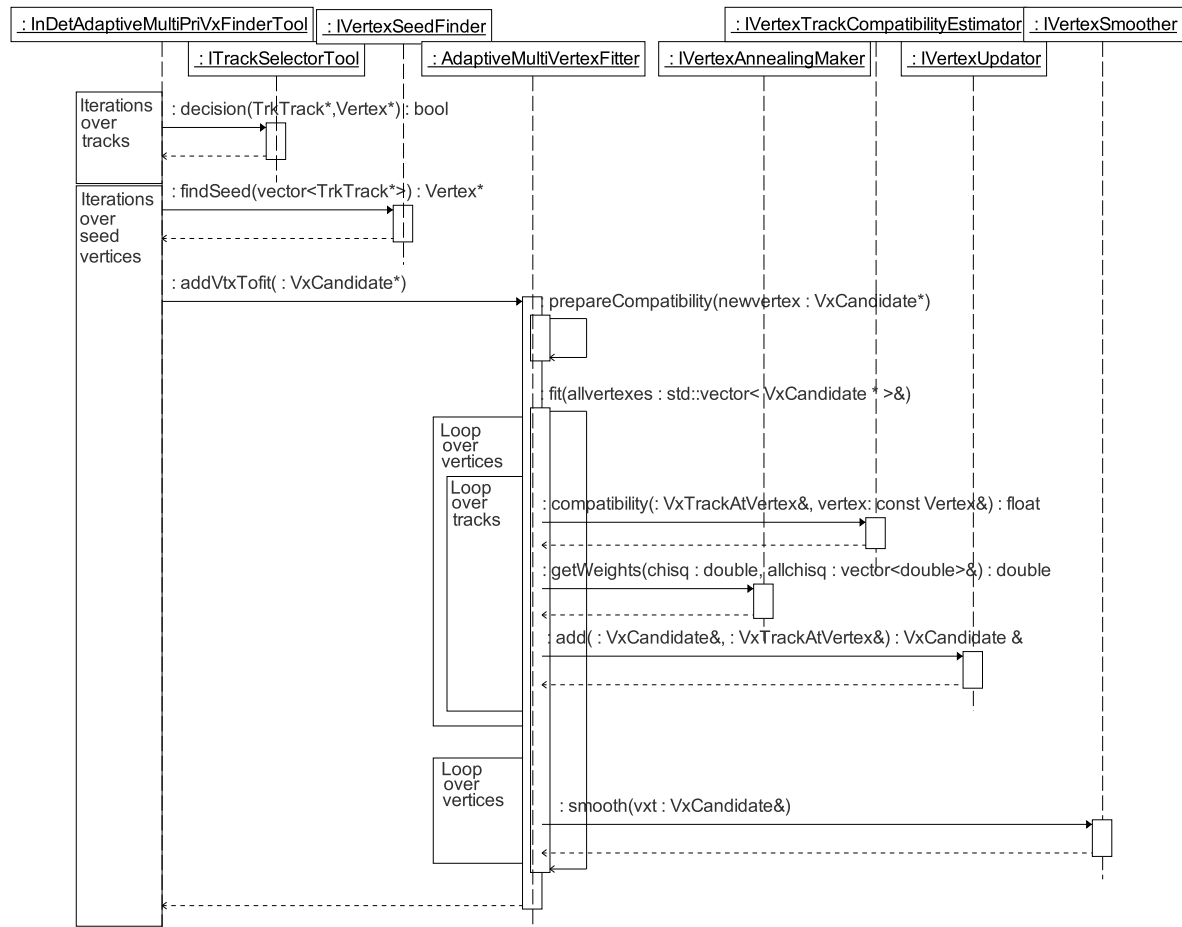


Figure A.8: Sequence diagram for the adaptive multi-vertex finder.

A.3 Tuning of parameters in the adaptive vertex fitters

The two most important parameters affecting the adaptive behaviour of the adaptive vertex fitters are the value of χ^2_{cutoff} for the track weight factor $\omega(\chi^2)$ (Eq. 5.45) and the set of temperatures used for the annealing scheme.

$\mathcal{L} = 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ with $\Delta t_{BC} = 75 \text{ ns}$, $\langle N_{pu} \rangle = 6.9$								
$t\bar{t}$								
χ^2_{cutoff}	Set of temperatures	<i>Misid.</i>	<i>Lost</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
9	[64, 16, 4, 2, 1.5, 1]	0.5%	0.2%	0.3%	0.7	32.2 μm	8.8 μm	38.7 μm
9	[16, 8, 4, 2, 1.5, 1]	0.4%	0.2%	0.3%	0.7	32.2 μm	8.9 μm	38.6 μm
9	[256, 64, 16, 4, 2, 1]	0.5%	0.2%	0.4%	0.8	32.2 μm	8.9 μm	38.9 μm
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	0.5%	0.2%	0.4%	0.8	32.2 μm	8.8 μm	38.8 μm
7	[64, 16, 4, 2, 1.5, 1]	0.5%	0.2%	0.4%	0.7	32.2 μm	8.9 μm	39.0 μm
5	[64, 16, 4, 2, 1.5, 1]	0.6%	0.3%	0.4%	0.6	32.8 μm	9.0 μm	39.8 μm
3	[64, 16, 4, 2, 1.5, 1]	23.1%	22.4%	23.7%	0.5	33.9 μm	9.0 μm	41.8 μm
$WH(\rightarrow l\nu u\bar{u})$								
χ^2_{cutoff}	Set of temperatures	<i>Misid.</i>	<i>Lost</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
9	[64, 16, 4, 2, 1.5, 1]	1.7%	0.5%	1.4%	0.7	34.0 μm	8.0 μm	44.8 μm
9	[16, 8, 4, 2, 1.5, 1]	1.6%	0.4%	1.3%	0.6	33.9 μm	8.0 μm	44.7 μm
9	[256, 64, 16, 4, 2, 1]	1.9%	0.8%	1.7%	0.7	34.1 μm	8.1 μm	44.9 μm
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	1.8%	0.6%	1.5%	0.7	34.0 μm	8.0 μm	44.8 μm
7	[64, 16, 4, 2, 1.5, 1]	1.7%	0.6%	1.5%	0.6	34.3 μm	8.1 μm	45.0 μm
5	[64, 16, 4, 2, 1.5, 1]	1.8%	0.7%	1.6%	0.5	35.2 μm	8.1 μm	45.7 μm
3	[64, 16, 4, 2, 1.5, 1]	32.5%	31.1%	33.2%	0.5	36.5 μm	8.4 μm	48.3 μm
$WH(\rightarrow l\nu b\bar{b})$								
χ^2_{cutoff}	Set of temperatures	<i>Misid.</i>	<i>Lost</i>	<i>Outliers</i>	$\langle n_{trk} \rangle_{PU}$	σ_z	σ_x	RMS _z
9	[64, 16, 4, 2, 1.5, 1]	4.0%	0.7%	3.8%	0.6	41.4 μm	10.4 μm	55.5 μm
9	[16, 8, 4, 2, 1.5, 1]	3.9%	0.6%	3.7%	0.6	41.3 μm	10.3 μm	55.5 μm
9	[256, 64, 16, 4, 2, 1]	4.3%	1.0%	4.1%	0.7	41.7 μm	10.4 μm	55.8 μm
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	4.1%	0.8%	3.9%	0.7	41.6 μm	10.4 μm	55.6 μm
7	[64, 16, 4, 2, 1.5, 1]	4.1%	0.7%	3.9%	0.6	41.9 μm	10.3 μm	55.6 μm
5	[64, 16, 4, 2, 1.5, 1]	4.3%	1.0%	4.1%	0.5	42.2 μm	10.4 μm	56.2 μm
3	[64, 16, 4, 2, 1.5, 1]	29.7%	27.2%	30.5%	0.5	43.7 μm	10.6 μm	58.8 μm

Table A.1: Comparison of the performance of the adaptive multi-vertex finding algorithm when different parameters χ^2_{cutoff} and set of temperatures are applied to the adaptive multi-vertex fit. Listed are the fraction of misidentified vertices, where the signal vertex was not correctly reconstructed and selected (*Misid.*), the fraction of not found vertices, where the signal vertex was not correctly reconstructed (*Lost*), the fraction of events where the reconstructed and selected vertex is not in a z window of 0.5 mm from the signal simulated vertex (*Outliers*), the average number of tracks from pile-up ($\langle n_{trk} \rangle_{PU}$) and the core resolutions (σ) and RMS of the vertex residual distributions along the z and x axes. These quantities are properly defined in Section 5.5.4.

The initial temperature is particularly important: a high value will recover cases where the initial vertex position estimate provided by the seed finder is only very poor, but will also allow for tracks far away from the vertex of interest to play an undesired role in the fit; on the contrary, a too low value requires the seed finder to provide a very good estimate of the initial vertex position, otherwise tracks originating from the vertex of interest may be discarded during early stages of the fit. A comparison of the effect of using various different values for χ^2_{cutoff} and set of temperatures is presented in Table A.1 and Table A.2. The performance is very similar in all cases, with a clear exception for the lowest values of χ^2_{cutoff} : a value of 3 for this cut means that all tracks with $\text{Prob}(\chi^2) < 22\%$ are removed from the fit, which means that way too many tracks are discarded from the vertex fit. Even if the difference with respect to the default setting listed as the first option is very small, the best performance is

seen by using the set of temperatures [16, 8, 4, 2, 1.5, 1]: the fact that this happens with an initial temperature of 16 is an indication of the fact that the seed finding algorithm is indeed working very well. Another thing to notice is that the width of the pulls starts to shrink when the χ^2_{cutoff} is lowered: this is simply due to the fact that the tracks with the largest residuals are removed from the fit.

As a result, the best parameters, among those considered in this study, seem to be a χ^2_{cutoff} value of 9 (corresponding to $\text{Prob}(\chi^2) \approx 1\%$) and the set of annealing temperatures [16, 8, 4, 2, 1.5, 1].

$\mathcal{L} = 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ with $\Delta t_{BC} = 75 \text{ ns}$, $\langle N_{pu} \rangle = 6.9$					
$t\bar{t}$					
χ^2_{cutoff}	Set of temperatures	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,PULL}$	$\sigma_{x,PULL}$
9	[64, 16, 4, 2, 1.5, 1]	30.0 μm	8.6 μm	1.22	1.19
9	[16, 8, 4, 2, 1.5, 1]	30.0 μm	8.6 μm	1.22	1.19
9	[256, 64, 16, 4, 2, 1]	29.9 μm	8.6 μm	1.23	1.19
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	30.0 μm	8.6 μm	1.22	1.19
7	[64, 16, 4, 2, 1.5, 1]	31.2 μm	8.9 μm	1.19	1.14
5	[64, 16, 4, 2, 1.5, 1]	33.4 μm	9.4 μm	1.13	1.07
3	[64, 16, 4, 2, 1.5, 1]	36.2 μm	10.2 μm	1.10	1.02
$WH(\rightarrow l\nu u\bar{u})$					
χ^2_{cutoff}	Set of temperatures	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,PULL}$	$\sigma_{x,PULL}$
9	[64, 16, 4, 2, 1.5, 1]	36.9 μm	9.2 μm	1.11	1.00
9	[16, 8, 4, 2, 1.5, 1]	36.9 μm	9.2 μm	1.10	1.00
9	[256, 64, 16, 4, 2, 1]	37.1 μm	9.3 μm	1.11	1.00
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	36.9 μm	9.2 μm	1.11	1.00
7	[64, 16, 4, 2, 1.5, 1]	38.3 μm	9.5 μm	1.08	0.97
5	[64, 16, 4, 2, 1.5, 1]	40.7 μm	10.0 μm	1.03	0.93
3	[64, 16, 4, 2, 1.5, 1]	43.8 μm	10.7 μm	1.02	0.91
$WH(\rightarrow l\nu b\bar{b})$					
χ^2_{cutoff}	Set of temperatures	$\langle \sigma_z \rangle$	$\langle \sigma_x \rangle$	$\sigma_{z,PULL}$	$\sigma_{x,PULL}$
9	[64, 16, 4, 2, 1.5, 1]	41.0 μm	10.3 μm	1.19	1.16
9	[16, 8, 4, 2, 1.5, 1]	41.0 μm	10.3 μm	1.19	1.16
9	[256, 64, 16, 4, 2, 1]	40.9 μm	10.2 μm	1.20	1.16
9	[64, 32, 16, 8, 4, 2, 1.5, 1]	41.0 μm	10.3 μm	1.20	1.16
7	[64, 16, 4, 2, 1.5, 1]	42.7 μm	10.6 μm	1.16	1.11
5	[64, 16, 4, 2, 1.5, 1]	45.6 μm	11.2 μm	1.10	1.05
3	[64, 16, 4, 2, 1.5, 1]	49.4 μm	12.0 μm	1.07	1.00

Table A.2: Comparison of the performance of the adaptive-multi vertex finding algorithm when different parameters χ^2_{cutoff} and set of temperatures applied to the vertex fit algorithm: the average estimated vertex error ($\langle \sigma \rangle$) and the Gaussian width of the pull distributions (σ_{PULL}) are listed, separately for the x and z directions.

B Details about the JetFitter algorithm

B.1 Optimization of the algorithmic execution speed

As discussed in Section 6.6.4, the execution speed of the JetFitter algorithm is mainly limited by the need of inverting a matrix with the size of the state vector, which in this specific case is $5+N$. In the following a strategy is set up to significantly improve the execution speed.

There is one precise reason why the JetFitter vertex fit can be made significantly faster. When adding a certain track to the vertex number i , the measurement equation contained in Eq. 6.10, once Eqs. 6.13 and 6.14 are taken into account, connects the track parameters of that track with the primary vertex position, the jet axis direction and the flight length of the corresponding vertex along the flight axis d_i , but not directly with the flight length of all remaining vertices d_j , with $j \neq i$. To make this statement stronger, if the whole decay chain fit would be expressed in terms of a likelihood function to be maximized as a function of the parameters \vec{d} , it would contain no explicit terms which involve two different d_i and d_j , for $i \neq j$. As a consequence the weight matrix $COV^{-1}(\vec{d})$ of the decay chain vertex fit at any fit stage (after adding an arbitrary number of tracks to the fit) will have all non-diagonal elements $COV^{-1}(\vec{d})_{d_i, d_j} \equiv 0$. This can be exploited when, at the end of a Kalman update step, the weight matrix of size $5 + N$ needs to be inverted to get the covariance matrix and, according to that, the value of the state vector \vec{d} (according to the general Kalman update step presented in Eq. 5.20).

In practice, the weight matrix $COV^{-1}(\vec{d})$ of the fitted decay chain will be always of the form:

$$\begin{aligned}
 COV^{-1}(\vec{d}) &= \begin{bmatrix} w(\vec{r}, \vec{r}) & w(\vec{r}, \phi) & w(\vec{r}, \theta) & w(\vec{r}, d_1) & w(\vec{r}, d_2) & w(\vec{r}, \dots) \\ w(\vec{r}, \phi) & w(\phi, \phi) & w(\phi, \theta) & w(\phi, d_1) & w(\phi, d_2) & w(\phi, \dots) \\ w(\vec{r}, \theta) & w(\phi, \theta) & w(\theta, \theta) & w(\theta, d_1) & w(\theta, d_2) & w(\theta, \dots) \\ w(\vec{r}, d_1) & w(\phi, d_1) & w(\theta, d_1) & w(d_1, d_1) & 0 & 0 \\ w(\vec{r}, d_2) & w(\phi, d_2) & w(\theta, d_2) & 0 & w(d_2, d_2) & 0 \\ w(\vec{r}, \dots) & w(\phi, \dots) & w(\theta, \dots) & 0 & 0 & w(\dots, \dots) \end{bmatrix} \\
 &= \begin{bmatrix} A_{3 \times 3} & B_{3 \times N} \\ C_{N \times 3} & D_{N \times N} \end{bmatrix}, \tag{B.1}
 \end{aligned}$$

where \vec{r} is the primary vertex position, ϕ and θ the directions of the b -flight axis and d_1, d_2, \dots the fitted distances of the various vertices on the flight axis. Using the well known block inversion formula for this matrix and expressing it as a function of the inverse of the D matrix, one gets:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}. \tag{B.2}$$

The inversion of the diagonal matrix D is extremely fast, so that, by using this formula, the slowest part of the computation is condensed in the inversion of the $A - BD^{-1}C$ matrix, which is however of fixed size 5, independently on the value of N . This is significantly faster than the inversion of the original matrix of size $N + 5$.

In addition, as discussed when demonstrating the equivalence between the Billor and Kalman filter method in the *weighted means* formalism (Section 5.3.2), the Kalman filter update step, presented in its general formulation in Eq. 5.20, can be applied for updating the fit with every track in such a way to only update the quantity $C_k^{-1} \cdot \vec{x}_k$, so that the inversion of the weight matrix of the state vector C_k^{-1} is required only at the last Kalman update step with $k = N_{trk}$, in order to get access to the final state vector \vec{x}_k . This reduces the amount of needed matrix inversions by a factor N_{trk} .

Several different optimization strategies are also possible, like for example switching from the *weighted mean* to the *Kalman gain* formalism, which in principle requires only the inversion of a matrix of the size of the measurement which is iteratively added to the fit (which would be the size of the track parameters, which is 5). However this approach has several other drawbacks, so that the solution of inverting the matrix block-wise and performing the inversion only during the last fit iteration seems to be really the optimal one.

The result of this optimization is discussed in Section 6.6.4.

B.2 Validation of neutral reconstruction

The explicit reconstruction of neutral particles corresponding to two-track vertices compatible with conversions, K_S^0 mesons and Λ baryons has been introduced in Section 6.7.3. Out of these two-track vertices and of the momenta of the tracks at the vertex a neutral track is created, by adopting the perigee parameterization for neutral tracks.

In order to cross-check if the neutral reconstruction algorithm works correctly, the track parameters of the neutral candidates have been analysed. First, by using Monte Carlo truth information in $WH \rightarrow l\nu u\bar{u}$ events, only neutral candidates corresponding to real neutral particles have been selected and then the corresponding *true* track parameter values have been computed. Then, the residuals of the track parameters with respect to the true expected value, divided by their errors, have been derived, both for K_S^0 mesons (Fig. B.1) and for conversions (Fig. B.2). K_S^0 mesons are reconstructed extremely well, with all pulls being nearly gaussian and with widths relatively close to unity: as a consequence, the impact parameter significance can be considered as a reliable estimator of whether the neutral candidate originates or not from the primary vertex. The same is expected for Λ baryons. In the case of conversions (Fig. B.2) the results are a bit worse, because of the energy losses the electrons undergo when interacting with the detector material, due to the sizable effect of bremsstrahlung. In coincidence of where the photon emission(s) took place, kinks are generated in the electron tracks, which mostly affect the measurement of the absolute value of the track momentum. As a consequence, the absolute value of the momentum of the conversion candidates has an extremely large bremsstrahlung tail to lower values (higher values of $1/p$ in the plot), but also most of the other pulls of the conversion candidates have significant tails, especially the transverse impact parameter significance, even though their behaviour can still be considered as nearly Gaussian.

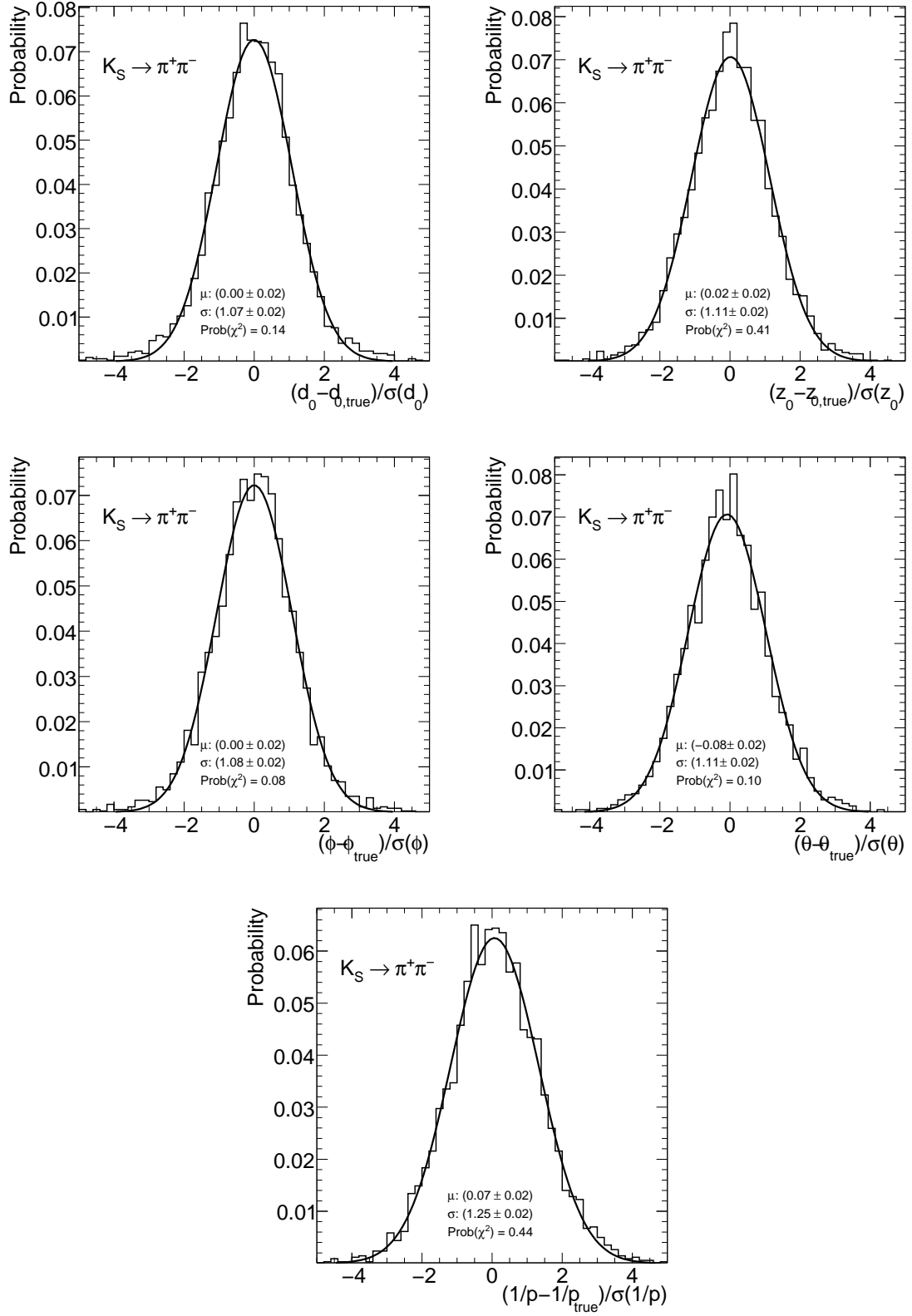


Figure B.1: Distribution of pulls (residuals divided by respective errors) of the neutral track parameters corresponding to K_S^0 meson candidates in light jets: the transverse (top left) and longitudinal (top right) impact parameters with respect to the primary vertex, the ϕ (middle left), θ (middle right) and $1/p$ (bottom) components of the momentum are shown separately.

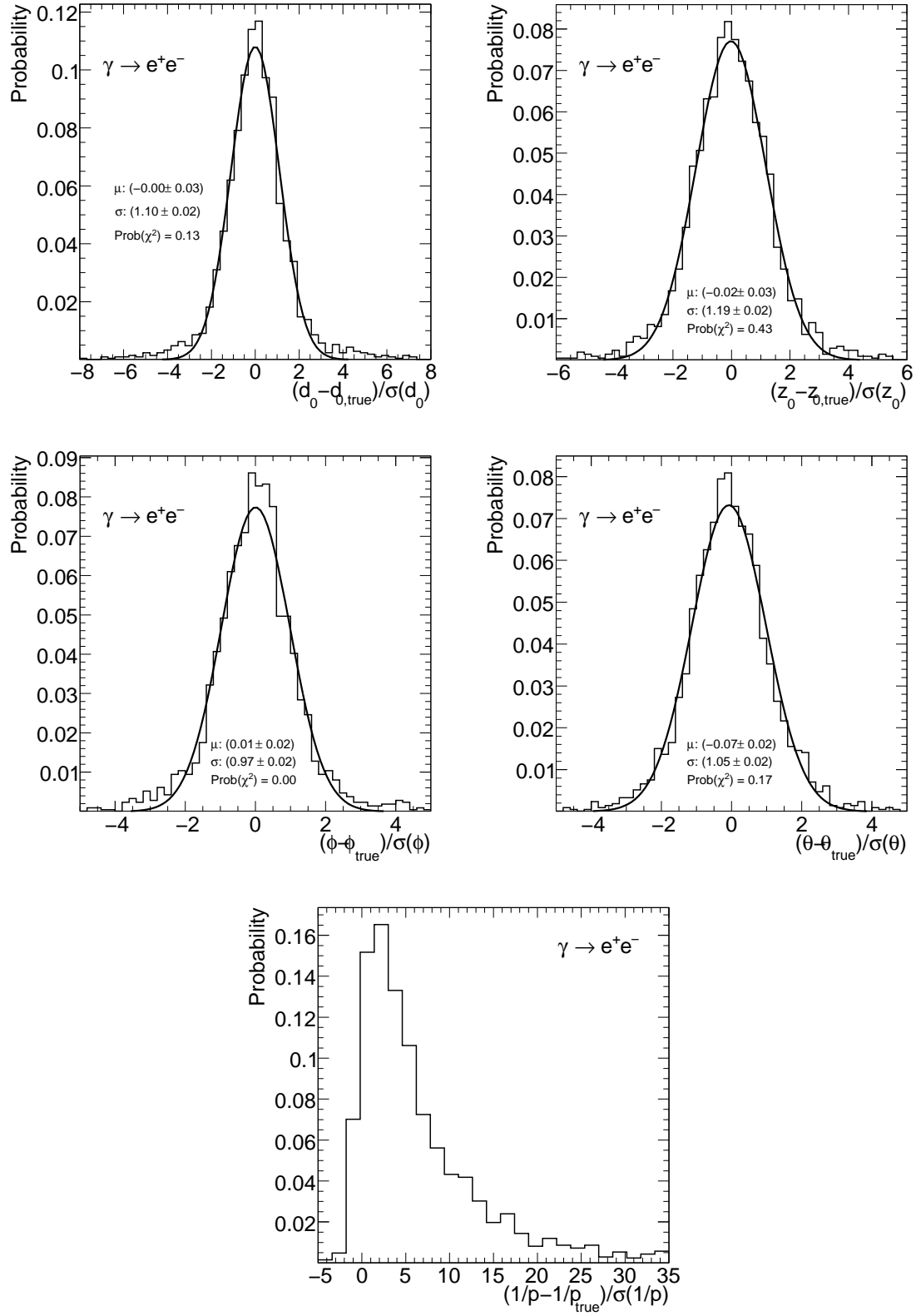


Figure B.2: Distribution of pulls (residuals divided by respective errors) of the neutral track parameters corresponding to conversion candidates in light jets: the transverse (top left) and longitudinal (top right) impact parameters with respect to the primary vertex, the ϕ (middle left), θ (middle right) and $1/p$ (bottom) components of the momentum are shown separately.

B.3 Implementation in the ATLAS reconstruction software

In this section some details will be given about how the JetFitter algorithm has been implemented in the ATLAS reconstruction software ATHENA. As in the case of primary vertex finding, the vertex fitting and finding algorithms have been completely decoupled, in order to gain flexibility and modularity.

The information about the decay chain fit during and after the vertex fit is stored in specific event data classes, which represent an extension of the event data model classes presented in Section A.2. The complete decay chain fit is represented by a *VxJetCandidate* (Fig. B.3), which inherits from the *VxCandidate* class, used for all more conventional single-vertex fitting algorithms (Fig. A.1). The *VxJetCandidate* adds to the class it inherits from all the properties which are needed to fully describe the decay chain. This includes, first of all, the parameters determined in the fit (primary vertex position, b -flight axis direction in ϕ and θ and the positions of the single vertices along the b -jet axis) and their related covariance matrix, which are both stored in the *VertexPositions* and *RecVertexPositions* classes, together with the overall vertex fit χ^2 . While the tracks used in the vertex fit are already stored as tracks at vertex (*VxTrackAtVertex*) in the vertex candidate (*VxCandidate*) class, an extra set of classes stores the information about the vertices along the jet axis (*VxVertexOnJetAxis*), which include their index number and their associated tracks. During clustering, the compatibility of all pair of vertices needs to be computed: this information is stored in a dedicated class, the clustering table (*VxClusteringTable*).

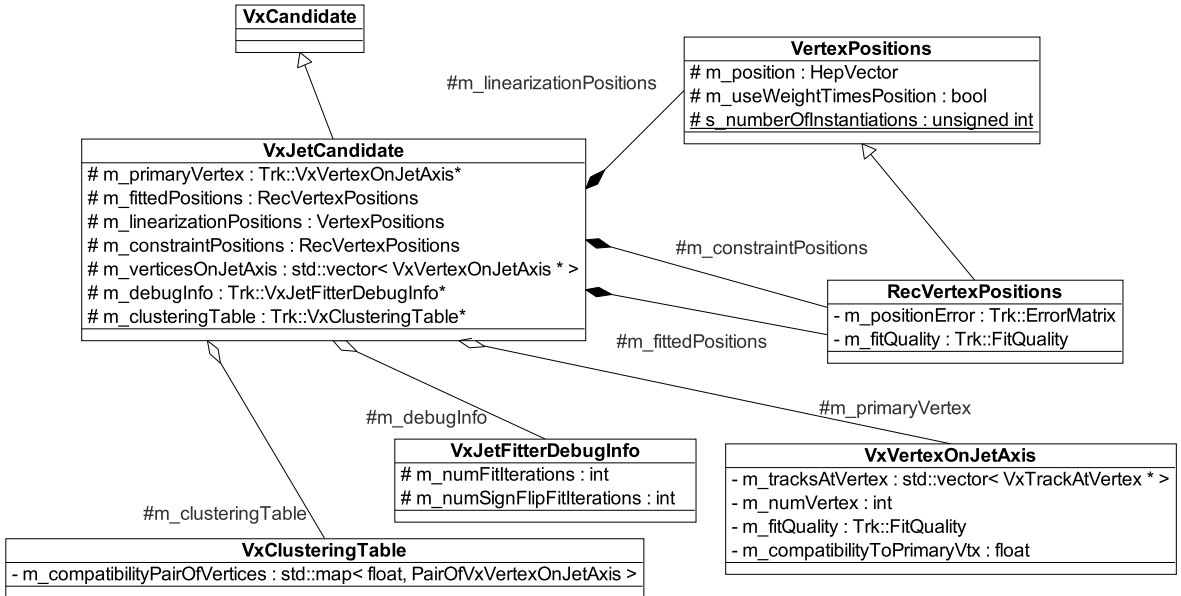


Figure B.3: Extension of the single-vertex event data model required to store all the informations of the decay chain fit for the JetFitter algorithm.

While in the case of primary vertex finding, the algorithms were designed in such a way that every vertex fitting or finding algorithm was getting a well defined set of input classes and then giving back a different set of output classes, in the case of JetFitter, due to the higher algorithmic complexity, it turned out to be more convenient to have a well defined set of objects which completely define the decay chain fit in its actual status, the *VxJetCandidate* class described above, with all its sub-classes, and then the vertex fitting algorithm implements methods which act on such objects, processing them and directly modifying them.

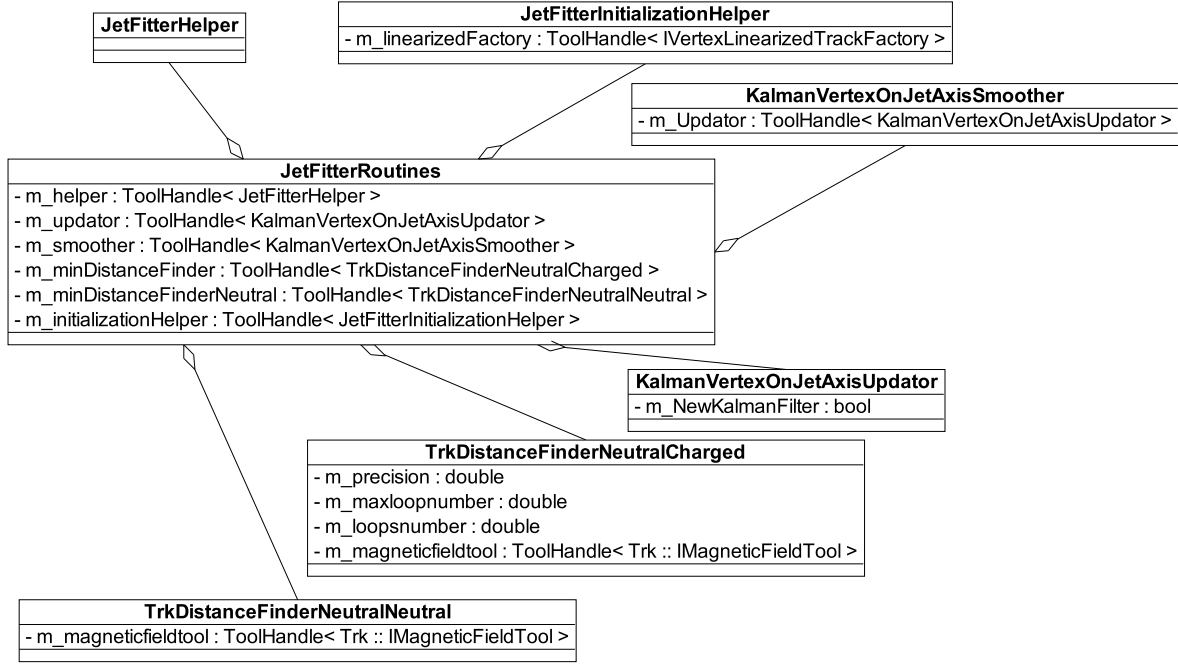


Figure B.4: Hierarchical structure of the vertex fitting tools of the JetFitter algorithms.

The tools implementing the decay chain vertex fit are shown in Fig. B.4. The most important class contains the main set of JetFitter routines (*JetFitterRoutines*) and handles the methods for performing the decay chain fit and for filling the table of probabilities for clustering. This class makes use of a series of satellite classes. The *JetFitterInitializationHelper* class contains the methods to initialize the decay chain (*VxJetCandidate* object) prior to any fit, while the *JetFitterHelper* class contains methods to modify the *VxJetCandidate*, for example to merge vertices together or to delete vertices from the decay chain. The real Kalman update and smoothing steps are performed by the *KalmanVertexOnJetAxisUpdater* and by the *KalmanVertexOnJetAxisSmoother*. Finally, in order to initialize the distances of the single-track vertices along the jet axis before the first fit iteration, the neutral-to-neutral and charged-to-neutral track distance finders have been implemented (*TrkDistanceFinderNeutralNeutral* and *TrkDistanceFinderNeutralCharged*).

The structure of JetFitter's vertex finding tool is not equally modular (Fig. B.5). It contains one main class (*InDetImprovedJetFitterVxFinder*), which steers all the vertex finding operations, by making use of large variety of tools. These include track selection tools, seed finding and single-vertex fitting tools (to manage the vertex fit of two-track vertices) and all mentioned vertex fitting tools of the JetFitter algorithm. Some more detailed mathematical expressions for processing impact parameters, creating neutral tracks and computing lifetime signs are implemented in a satellite class (*InDetJetFitterUtils*). The main vertex finding tool inherits from an abstract class (*ISecVertexInJetFinder*): this means, in practice, that all inclusive secondary vertex finders for *b*-tagging in ATLAS, including JetFitter, use the same input (primary vertex position and error matrix, set of input tracks and jet direction) and provide the same kind of output object, where all the informations about the found secondary vertex or vertices are stored.

A simplified sequence diagram showing how the JetFitter finding algorithm works is presented in Fig. B.6. First the tracks are selected, by using the track selector tool, then all

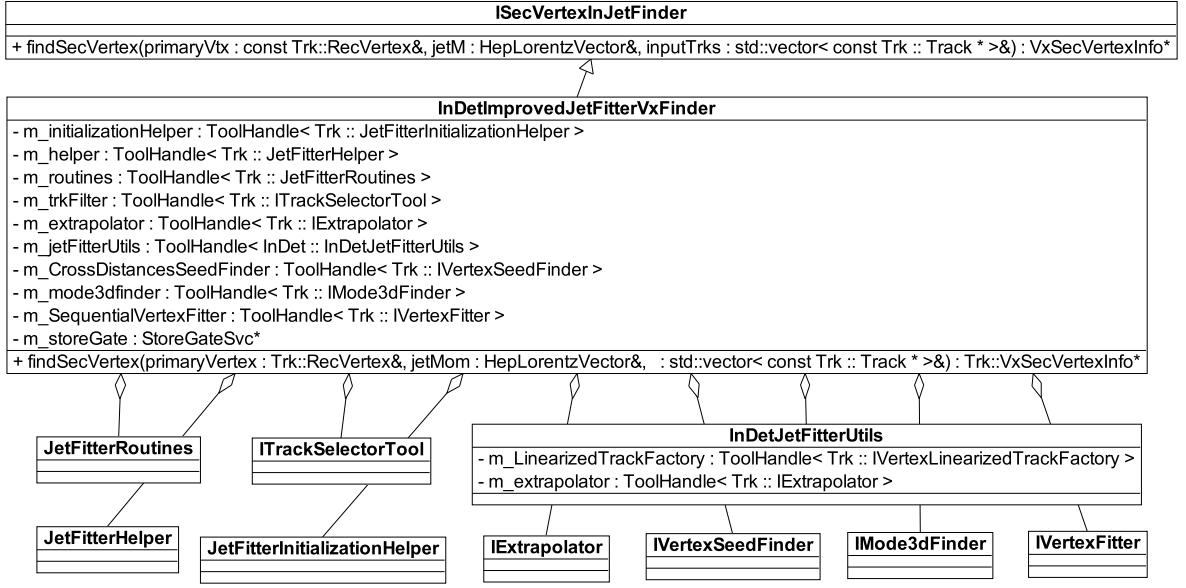


Figure B.5: Hierarchical structure of the vertex finding tools of the JetFitter algorithms.

two-track vertices are found by using a Kalman fit based vertex reconstruction algorithm, and neutral track candidates are created to be used later in the decay chain fit. The decay chain is then initialized and the distances of the initial single-track vertices are initialized to their point of closest approach to the initial b -flight axis direction, which is assumed to be equal to the jet direction. Then the iterations for the clustering along the b -flight axis are started. A single fit iteration consists of the linearization of all tracks, in doing a Kalman update step with all tracks, and, after convergence, in smoothing all vertices along the b -flight axis. After that, the table of probabilities for all pairs of found vertices to be compatible with a single vertex is filled, and this is used by the finding algorithm to decide what vertices to merge and whether to start a new iteration of the clustering process or not.

The result of the finding procedure is stored in a dedicated secondary vertex information object (*VxJetFitterVertexInfo*), which inherits from the common class used to store the output of a generic secondary vertex finding algorithm for b -tagging in ATLAS (*VxSecVertexInfo*), as shown in Fig. B.7. The fitted decay chain is stored, together with the selected tracks in the jet (primary and secondary tracks) and with all two-track vertices, including neutral particle candidates.

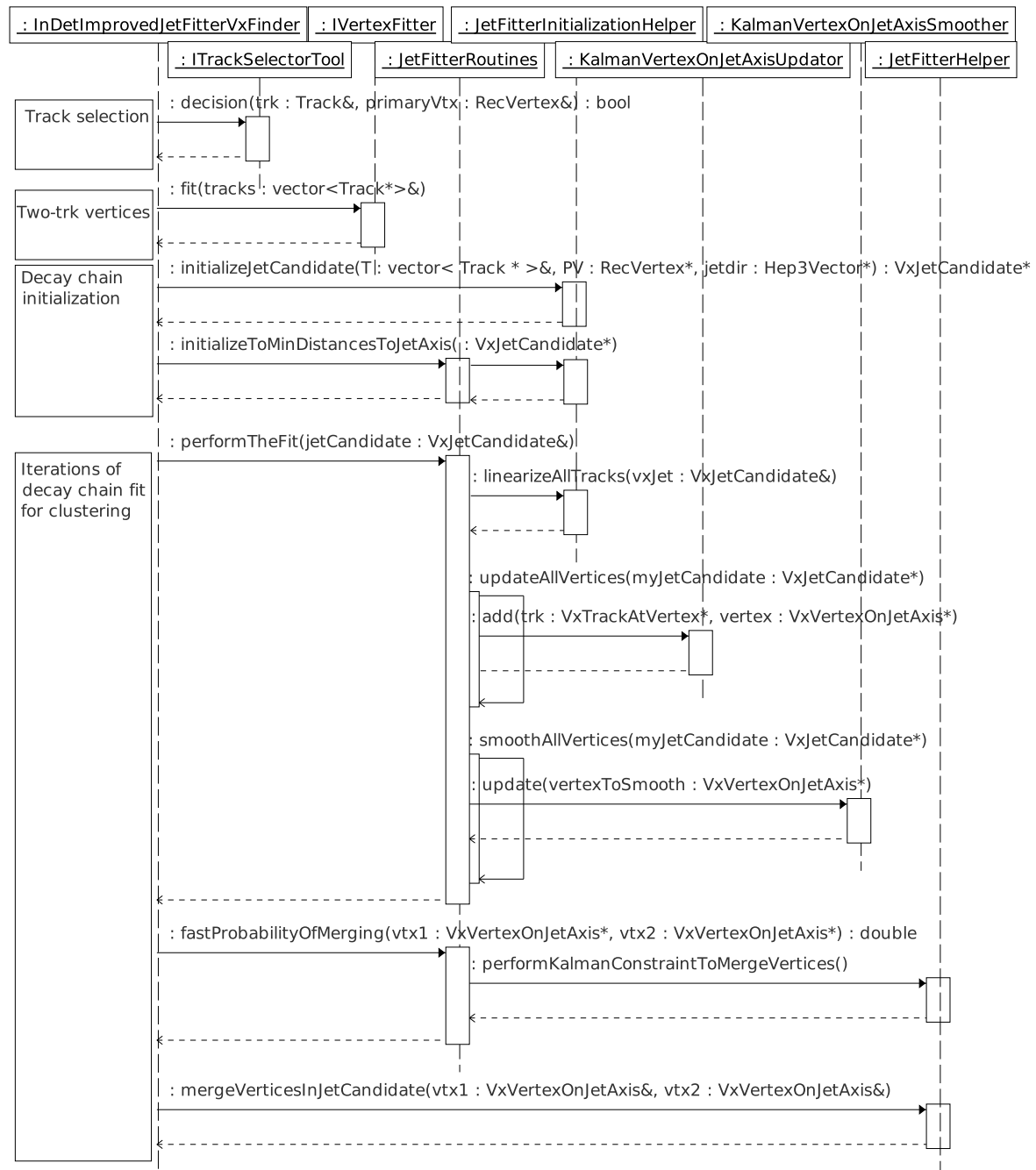


Figure B.6: Sequence diagram showing how the JetFitter finding algorithm works and how it makes use of the various vertex fitting tools.

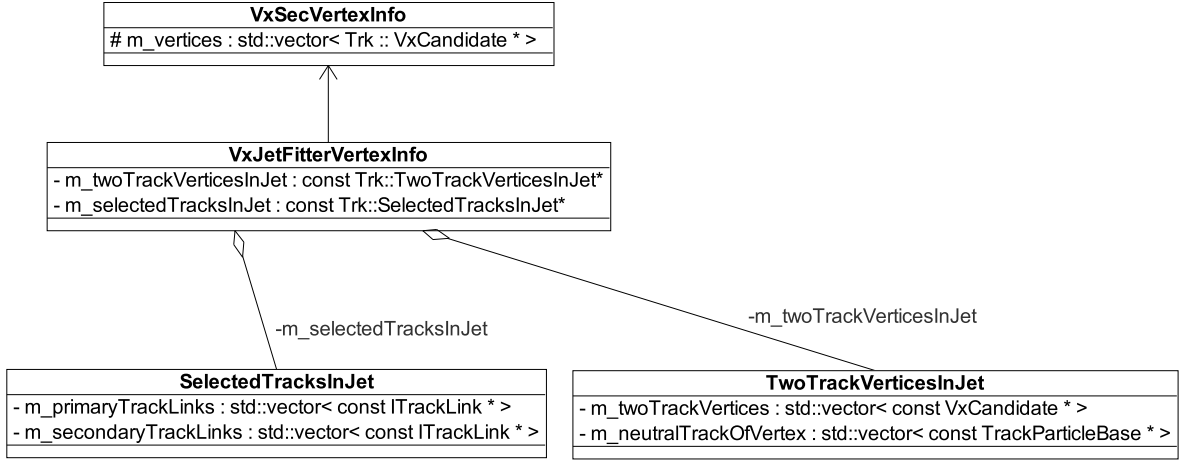


Figure B.7: Data classes used to store the secondary vertex finding result.

C WH analysis: further details

C.1 Parton level study of the $t\bar{t}Z$ background

The $t\bar{t}Z$ background is potentially dangerous for the WH analysis, characterized by the $\ell\nu b\bar{b}$ signature, in particular if the Z decays into a pair of neutrinos: here the additional transverse missing energy increases the relative background selection efficiency and the fact that the $t\bar{t}$ system is not selected back-to-back makes the emission of two b -quarks or of a c and b -quark from the two top quarks inside a cone of radius $\Delta R = 1.2$ easier with respect to the $t\bar{t}$ background. However, even if the relative selection efficiency of the $t\bar{t}Z$ background is higher with respect to $t\bar{t}$, since the $t\bar{t}Z$ production cross section is only 0.86 pb at LO, one can still hope to end up with a negligible background contribution.

This has been cross-checked on a sample of parton level events produced with the MadEvent event generator [102]. After application of the following selection cuts:

- $p_T(\ell) > 30$ GeV, $|\eta|(\ell) < 2.7$, $E_T^{\text{miss}} > 30$ GeV
- $p_T(W \text{ boson candidate}) > 200$ GeV
- no additional leptons with $p_T > 20$ GeV and $|\eta| < 2.5$
- cb - or bb -quark pair with $|\eta(b/c)| < 2.5$, $p_T(bb/cb) > 200$ GeV, $0.3 < \Delta R(bb/cb) < 1.2$
- no additional parton (light- or c -quark) with $p_T > 30$ GeV
- $(100 < m_{bb/cc} < 140)$ GeV

and adding a conservative estimate of the b - and c -(mis)tagging efficiencies (respectively 70 % and 10 %), one can conclude that, considering both the potential bb - and cb -quark contributions, the contribution of the $t\bar{t}Z$ background in the $\ell\nu b\bar{b}$ analysis is < 0.6 events with 30 fb^{-1} at 95 % confidence level. The $t\bar{t}Z$ background with the Z boson decaying into a b -quark pair is not considered in more detail, since in this case there are two more jets in the event on which the additional jet veto can apply, so that this contribution can be expected to be much smaller than $t\bar{t}Z$ with $Z \rightarrow \nu\nu$ and therefore negligible. This confirms that no significant contribution is expected from the $t\bar{t}Z$ background.

C.2 $Wb\bar{b}$ background: comparison between HERWIG and AcerMC

Since the Monte Carlo statistics available for the W +jet sample is limited, all the analysis optimization and variables distribution plots were obtained using the AcerMC $Wb\bar{b}$ sample, including the HERWIG W +jet sample after removing the $Wg \rightarrow Wb\bar{b}$ component. However, since, as discussed in Section 7.3, the HERWIG generator is expected to provide a better description of this process in the particular region of phase space selected by this analysis, the inclusive HERWIG W +jet sample was used as a reference for the analysis selection and for the final mass distribution presented in Section 7.5. It is, however, worth analysing how

the analysis selection has a different impact on the AcerMC $Wb\bar{b}$ sample and on the subset of events of the HERWIG sample which contain the $Wg \rightarrow Wb\bar{b}$ contribution: this is shown in Table C.1, starting from the lepton cut, which removes the bias due to the fact that the W bosons were forced to decay leptonically in the AcerMC sample. As the Table shows, while the number of expected events agrees within statistics after the final mass window cut due to the limited statistics of the HERWIG sample, the jet veto and b -tagging cuts show some significant differences.

	$qq \rightarrow Wb\bar{b}$	$qq \rightarrow Wg$
$p_T(e/\mu) > 30$ GeV	1176 ± 12	1359
$p_T(\text{additional } \mu) < 10$ GeV	1153 ± 12	1346
$p_T(\text{additional } e) < 10$ GeV	1131 ± 12	1326
$\Delta\phi(W,H) > \frac{2}{3}\pi$	1087 ± 12	1316
no additional b -jets $p_T > 15$ GeV	926 ± 11	1179
add. jets on W side $p_T < 20$ GeV	663.0 ± 9.9	732
add. jets on H side $p_T < 20$ GeV	446.8 ± 8.3	461
one subjet b -tagged	340.5 ± 7.3	323
both subjets b -tagged	99.6 ± 4.0	60.7 ± 7.5
loose fit cuts	97.4 ± 4.0	59.8 ± 7.4
$112 \text{ GeV} < \text{mass}(H) < 136 \text{ GeV}$	10.7 ± 1.3	6.4 ± 2.4

Table C.1: Expected number of events for the tight selection cuts for the AcerMC $Wb\bar{b}$ sample compared to the HERWIG W +jet sample, where the only $qq \rightarrow Wg \rightarrow Wb\bar{b}$ component has been selected. All numbers correspond to an integrated luminosity of 30 fb^{-1} .

In order to explain such differences, the parton level study presented in Section 7.3 needs to be recalled here, where it was explained that in a significant fraction of the events in HERWIG the widest angle splitting is $g \rightarrow gg(\rightarrow gbb)$, and not $g \rightarrow b\bar{b}$: so in a high fraction of the events either the two leading subjets describe an underlying gb or $g\bar{b}$ pair, thus failing the b -tagging requirement, or an additional hard jet is present in the event, close to the $b\bar{b}$ pair, on which the veto on jets surrounding the $H \rightarrow b\bar{b}$ candidate can easily apply.

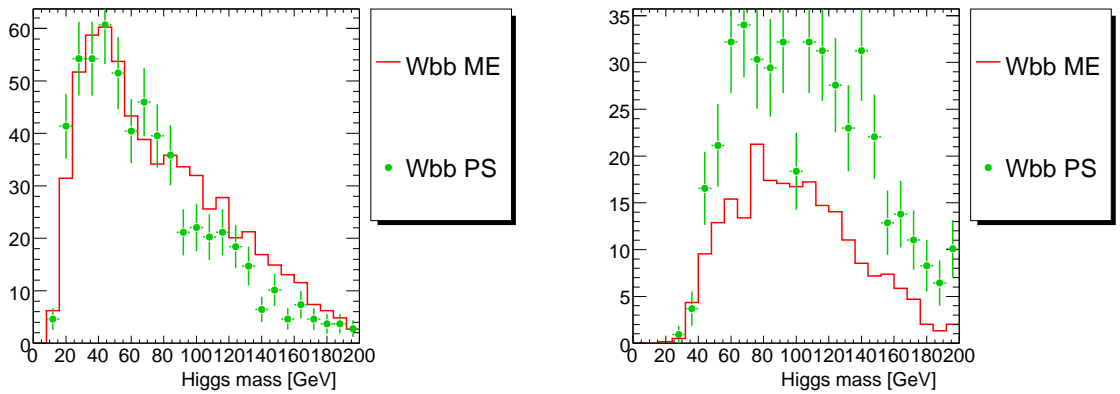


Figure C.1: Comparison between ME and PS based approaches for the simulation of the $W+b\bar{b}$ background of the invariant mass of the Higgs boson candidate after the $\Delta\phi$ cut, for the category where the two leading subjets are matched to a real $b\bar{b}$ pair (left) and for the category where they are matched to a real gb or $g\bar{b}$ pair (right). The distributions are normalized to the number of events expected in 30 fb^{-1} of collected data.

More specifically, after the $\Delta\phi$ cut, the number of events predicted by the PS is higher than those predicted by the ME approach, as expected from the parton level study. Focusing at

this stage of selection, the invariant mass distribution of the Higgs boson candidate is shown in Fig. C.1 separately for the category where the two leading subjets are matched to a real $b\bar{b}$ pair and for the category where they are matched to a real $g\bar{b}$ or $g\bar{b}$ pair¹. In the first case, the distributions are very similar for the PS and ME approaches, although the region of higher masses is slightly lower in the PS approach. In the second case, the PS predicts a much higher rate than the ME, which is most probably due to the presence of secondary $g \rightarrow b\bar{b}$ splittings. However, after b -tagging and jet vetoes are applied, the $g\bar{b}$ or $g\bar{b}$ contribution is essentially completely removed; in addition to that, the real $b\bar{b}$ contribution gets reduced more effectively in the PS than in the ME Monte Carlo, mainly due to the additional gluon radiation predicted by the PS based Monte Carlo, but also because in that case the b -jet efficiency turns out to be slightly lower. As a consequence, while at the beginning the PS approach predicts a higher rate than the ME, after all analysis cuts the situation turns into the opposite. However, as explained, the difference can be attributed to the different modeling of the jet shapes in HERWIG and AcerMC, for which the PS based approach should provide a more reliable description.

C.3 Subjet technique compared to traditional jet clustering

In Section 7.8.1 the WH analysis was repeated by using a traditional k_\perp algorithm instead of the subjet clustering method. Here some details of this comparison are discussed.

C.3.1 Improved rejection of $t\bar{t}$ background

The subjet clustering technique allows a significantly better rejection of the $t\bar{t}$ background.

However, this improvement, as shown in Fig. C.2, where the invariant mass distribution for the $t\bar{t}$ background is compared between the two jet clustering methods at two different stages of the analysis, gets visible only after the application of b -tagging. Even looking only at the combinations which are matched at truth level with a combination of b - and c -jets, the subjets clustering method results in a higher amount of background before b -tagging is applied. One possible explanation for this is that the three jets structure gets better decomposed by the subjet clustering method, so that, while one subjet turns out to be the b -jet, the other would find either a light or a c -jet. The k_\perp finding algorithm would instead typically find a single second jet, mixing together the contributions from the light and c -jets from the same W boson decay. While the true-level algorithm which matches the jets with the flavour of the original partons would still label also the light jet as a c -jet, since the c -jet would be in both cases nearer than $\Delta R = 0.3$ to the c -quark, the b -tagging algorithm would more easily reject around a half of the Higgs candidates selected by the subjet clustering algorithm, where the second selected jet is a light-jet. The better decomposition of the subjet structure by the subjets based technique is confirmed by the fact that, in particular after b -tagging is applied, the mass tends to peak at higher values, towards the top mass, which corresponds to a higher amount of cases where three subjets have been correctly included in the clustering of the Higgs candidate.

C.3.2 Impact of b -tagging on collinear $b\bar{b}$ pairs

The subjet clustering method is in principle able to reconstruct subjets which are significantly closer in pseudo-rapidity than $\Delta R = 0.4$; at the moment the analysis is, however, not

¹It should be stressed here that a subjet is labeled as b -subjet only if a b -quark is present with momentum within $\Delta R = 0.3$ from the subjet direction

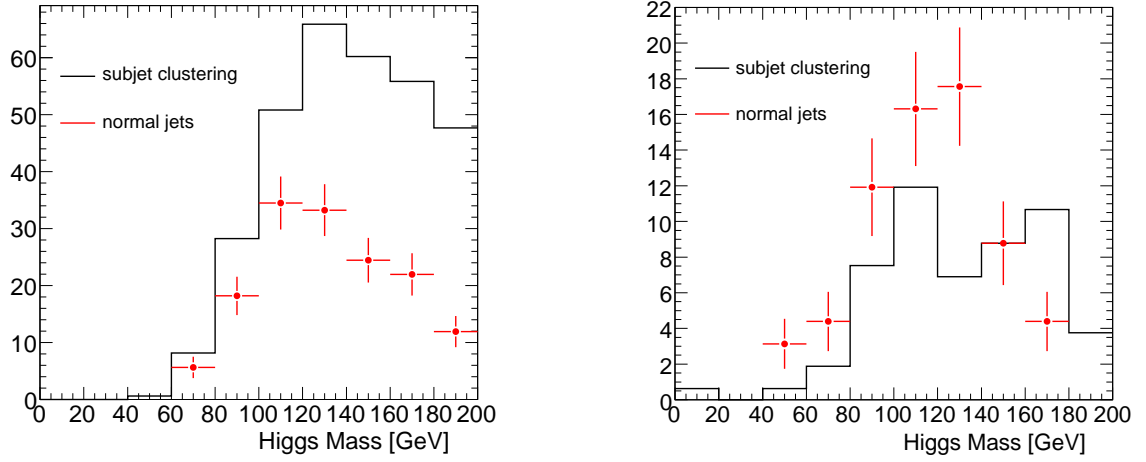


Figure C.2: (Left) Invariant mass distribution of the Higgs boson candidate for the conventional di-jet analysis (red) compared to same quantity as obtained from the subjets based analysis (black), before applying b -tagging but selecting only the c - b jet combinations at true level. (Right) Same distribution, but after b -tagging is applied. Both distributions are normalized to 30 fb^{-1} of integrated luminosity.

significantly profiting from this, since this feature is limited by the application of b -tagging, which provides an effective *turn on* curve as a function of ΔR . To clarify this effect, Fig. C.3 shows the Higgs boson mass distribution for the subjet clustering method, normalized to the number of entries in the histogram, before and after applying b -tagging on both subjets. It can be seen that the b -tagging efficiency drops for subjets nearer than $\Delta R = 0.2 - 0.3$. This effect is not so small as it may look like, because most of the higher p_T Higgs bosons which are easier to separate from the $t\bar{t}$ background, accumulate at lower ΔR . This effect could be cured by developing a dedicated b -tagging algorithm, which simultaneously fits the two $PV \rightarrow b \rightarrow c$ decay chains expected in the two overlapping b -jets. This could be developed in the future as an extension of the *JetFitter* vertexing algorithm, using eventually adaptive methods (as illustrated in this thesis for primary vertex finding applications) to dynamically associate tracks either to the first or to the second $b \rightarrow c$ -hadron decay chain in the vertex fit.

C.4 Comparison between ATLFAST-II and full simulation

C.4.1 Lepton reconstruction efficiency

The WH analysis relies on the identification of a high p_T lepton in the event. The electron and muon efficiency as a function of p_T and η for the Higgs signal samples with $m_H = 120 \text{ GeV}$ based on the ATLFAST-II simulation is shown in Fig. C.4, for all high p_T leptons from the decay of an associated W boson. In order to increase the available statistics no selection cuts are applied. The analysis relies on requiring an electron passing the medium quality electron identification requirement (denoted *medium isEM*), while the veto is applied to a looser quality identification requirement (denoted *loose isEM*) electron.

The same plots for the reconstructed *muons*, again from the W boson, are shown in Fig. C.5. The analysis requires the presence of a *combined* muon, while for the veto only a *standalone* muon is required.

The inefficiency around $\eta \approx 0$ is important for this analysis, since an eventual high p_T muon in this region will escape the veto, and translate into a very large contribution to E_T^{miss} . In

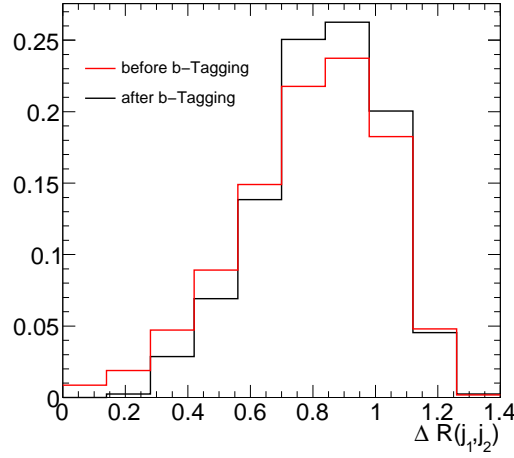


Figure C.3: $\Delta R(j_1, j_2)$ between the two subjects representing the Higgs boson candidate before and after applying b -tagging on them. Both histograms are normalized to an area of one.

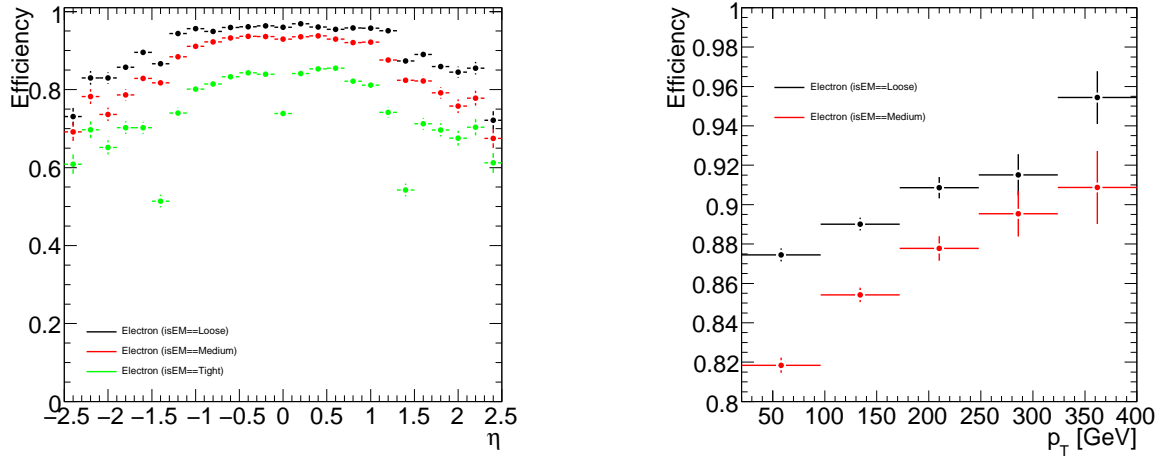


Figure C.4: Electron efficiency as a function of electron pseudo-rapidity η (left) and of electron p_T (right). Efficiencies are given for different object identification definitions: isEM=Loose, isEM=Medium and isEM=Tight (corresponding in ATLAS to a low, medium and high quality electron identification requirement).

this way a Z boson can be reconstructed as a W (in addition cases where the second lepton is outside acceptance, $|\eta| > 2.5$ for the electron or $|\eta| > 2.7$ for the muon, also contribute to this)².

Both signal and backgrounds are dominated by cases where the selected high p_T lepton comes from a W boson (or in few cases from a Z boson): it turns out that these are isolated leptons, so if ATLFAST-II can reproduce their efficiency correctly on the signal sample, it should be able to correctly reproduce their efficiency on the backgrounds too. Fig. C.6 shows the ratio between the number of electrons in ATLFAST-II and in full simulation, as a function

²In the future, the analysis could be improved by trying to veto events with a high p_T track in correspondence to $\eta \approx 0$ whose invariant mass with the highest p_T lepton (muon) in the event is compatible with the Z mass, to act against the muon reconstruction inefficiency at $\eta \approx 0$.

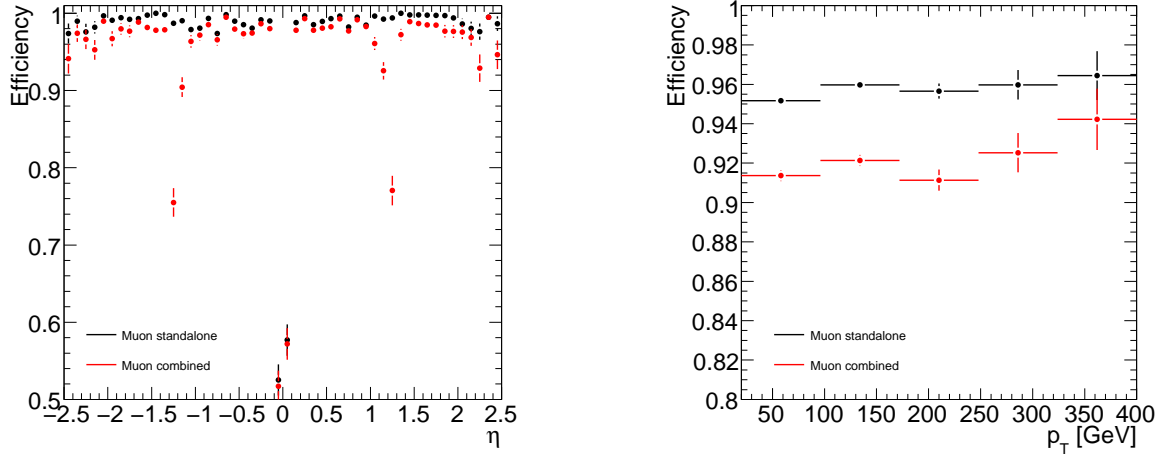


Figure C.5: Muon efficiency as a function of muon pseudo-rapidity η (left) and of muon p_T (right). Efficiencies are given for muons as reconstructed in the muon system and for muons combined with tracks reconstructed in the inner detector.

of p_T and η .

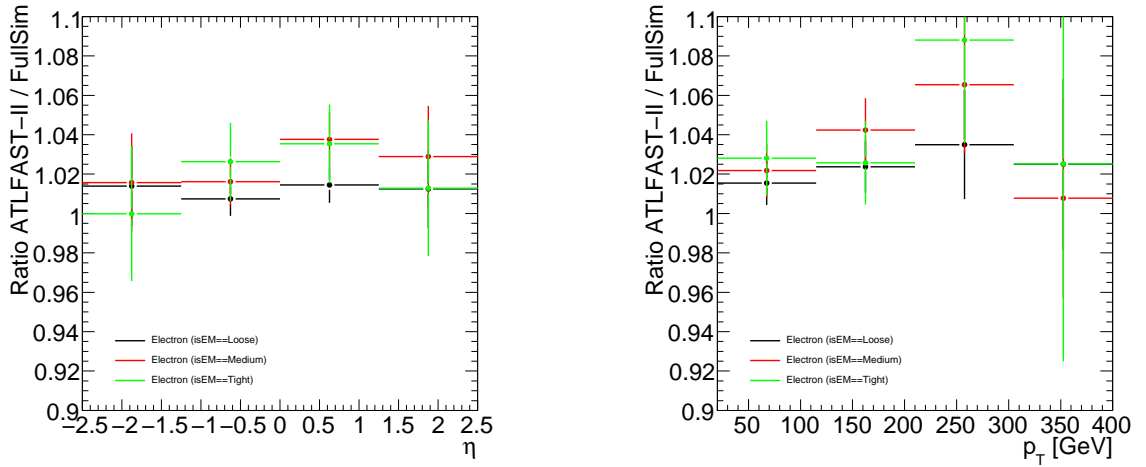


Figure C.6: Ratio between electron efficiency in ATLFAS-II and full simulation as a function η (left) and of p_T (right).

The same ratio is shown for muons in Fig. C.7.

For the muons, ATLFAS-II and the complete full simulation of the ATLAS detector are expected to provide the same result, since in both cases the muons pass through the full simulation of the ATLAS detector.

The overall shifts (ATLFAS-II versus complete full simulation) are shown in Table C.2: ATLFAS-II reproduces the muon efficiencies well (as expected), while a small shift to higher efficiency values is seen for the electrons. This is at most a 2 – 4% effect, but it is accounted for in the fit based analysis of Chapter 8, where the systematic uncertainties connected with this study are considered.

Since in the analysis a cut on p_T is applied, it is worth checking whether ATLFAS-II

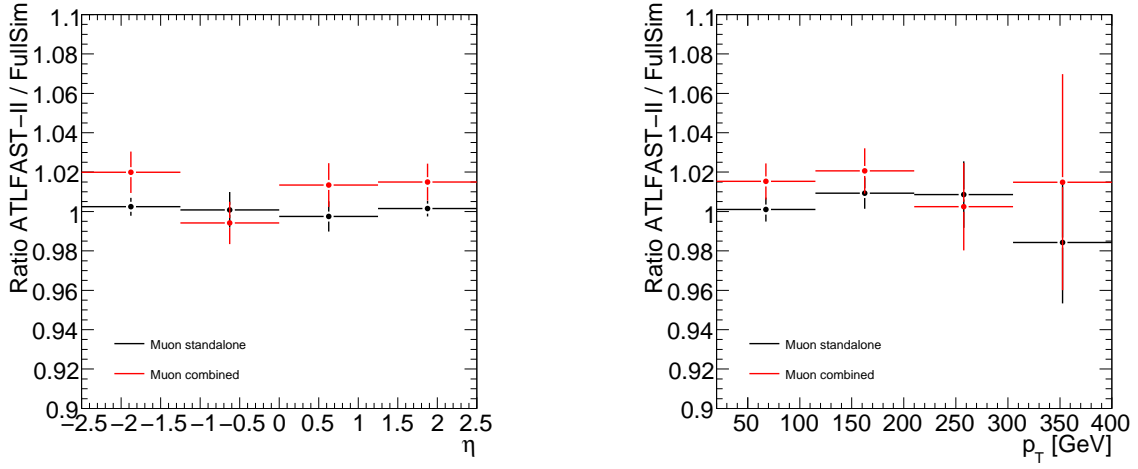


Figure C.7: Ratio between muon efficiency in ATLFast-II and full simulation as a function η (left) and of p_T (right).

Object	Definition	Shift ATLFast-II/ FullSim (%)
Muon	standalone	-0.1 ± 0.4
Muon	combined	0.8 ± 0.5
Electron	isEM==loose	2.0 ± 0.8
Electron	isEM==medium	3.2 ± 1.0
Electron	isEM==tight	3.1 ± 1.4

Table C.2: Overall shift of efficiencies for ATLFast-II with respect to full simulation, for both electrons and muons.

reproduces the p_T resolution of a muon or electron correctly. The resolution in p_T for both muons and electrons is shown in Fig. C.8.

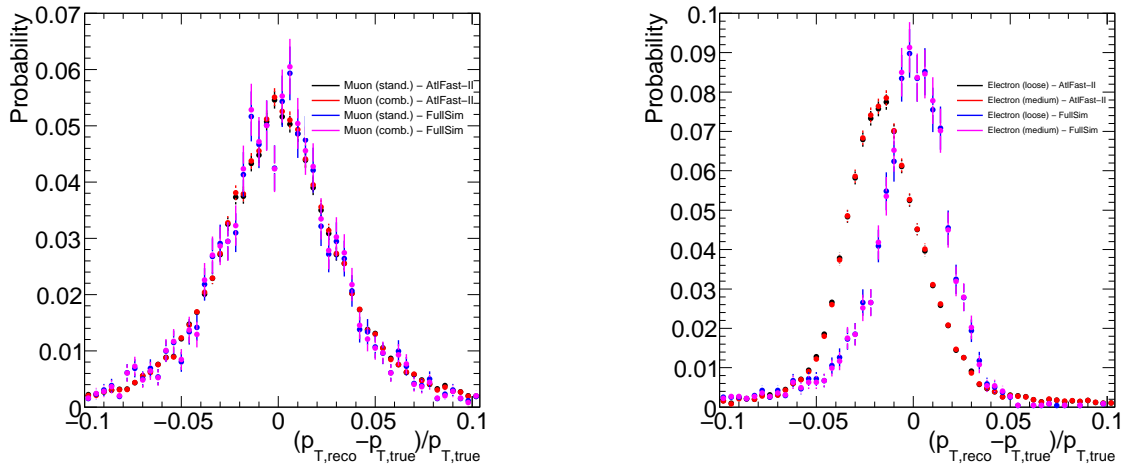


Figure C.8: Lepton p_T resolution in ATLFast-II and in full simulation for muons (left) and electrons (right).

While the muon p_T variable is perfectly reproduced (as expected), the electron p_T shows a shift of around 2% to lower p_T values. The resolution is slightly worse in ATLFast-II with respect to full simulation (the reason for this being that the parameterization of the EM shower shapes and energy profile of single particles in the calorimeter is based on an old version of the ATLAS GEANT4 simulation software). Both effects are again taken into account in the likelihood fit based analysis discussed in Chapter 8.

It is also worth considering the analysis selection in more detail looking separately for events with a W boson candidate decaying into a muon and into an electron. In Table C.3 and C.4 the analysis selection is shown separately in term of these two components, starting from the $p_T(e/\mu)$ cut.

	WH(120)		WZ		$t\bar{t}(p_T^{min})$		Wt		W+jet	
	e	μ	e	μ	e	μ	e	μ	e	μ
$p_T(e/\mu) > 30$ GeV	71	74	470	526	59269	55783	7916	7809	87800	90204
$p_T(\text{additional } \mu) < 10$ GeV	71	74	462	480	54642	52194	7534	7458	87605	89937
$p_T(\text{additional } e) < 10$ GeV	70	73	423	462	49585	47719	6997	6884	86283	88658
$\Delta\phi(W,H) > \frac{2}{3}\pi$	70	72	394	448	38570	46203	6282	6716	80597	87108
no additional b -jets $p_T > 15$ GeV	65	66	371	418	15029	15576	3858	3946	77281	83328
jets on W side $p_T < 60$ GeV	57	59	301	336	9642	9779	2924	2947	58532	62906
jets on H side $p_T < 60$ GeV	51	52	249	277	6858	6983	2170	2200	45407	48648
one subjet b -tagged	45	46	60	66	4344	4294	1217	1204	3405	3559
both subjets b -tagged	23	22	21	23	295	281	78	83	139	127
loose fit cuts	23	22	21	22	290	275	76	81	135	122
$112 \text{ GeV} < \text{mass}(H) < 136 \text{ GeV}$	15.3	14.5	1.4	2.2	60	69	20	18	17	15

Table C.3: The table shows the expected events going through the selection at each stage for the signal and the main backgrounds, starting from the $p_T(e/\mu)$ cut, separately for electrons and muons from the W boson. The numbers are projected to 30 fb^{-1} of collected data.

	ZH(120)		WW		ZZ		$t\bar{t}(p_T^{max})$		Z+jet	
	e	μ	e	μ	e	μ	e	μ	e	μ
$p_T(e/\mu) > 30$ GeV	3.6	5.4	1693	1771	106	159	1545	1354	4407	5728
$p_T(\text{additional } \mu) < 10$ GeV	3.3	2.4	1682	1759	98	80	1351	1157	4209	2965
$p_T(\text{additional } e) < 10$ GeV	1.5	2.1	1642	1719	49	72	1101	936	1991	2750
$\Delta\phi(W,H) > \frac{2}{3}\pi$	1.4	2.0	1571	1687	43	69	746	892	1411	2671
no additional b -jets $p_T > 15$ GeV	1.2	1.8	1507	1622	38	64	249	246	1283	2468
jets on W side $p_T < 60$ GeV	0.8	1.5	1220	1305	30	52	168	156	921	1825
jets on H side $p_T < 60$ GeV	0.8	1.4	1004	1079	25	44	120	113	712	1434
one subjet b -tagged	0.7	1.3	132	138	5.5	11	73	61	67	140
both subjets b -tagged	0.35	0.6	2.2	3.2	1.5	3.8	5	3	6	6
$112 \text{ GeV} < \text{mass}(H) < 136 \text{ GeV}$	0.19	0.46	0	0.4	0	0.4	3	0	0	0.8

Table C.4: The table shows the expected events going through the selection at each stage for the reference signal and the remaining backgrounds, starting from the $p_T(e/\mu)$ cut, separately for electrons and muons from the W boson. The numbers are projected to 30 fb^{-1} of collected data.

It can be noticed that around half the signal events is based on the identification of a high p_T muon from the W boson, while the other half is based on a high p_T electron. The pattern is very similar in all the main backgrounds.

C.4.2 Comparison of analysis selection on signal events

In this subsection, the ability of ATLFast-II in describing the Higgs boson candidate kinematic and efficiencies correctly will be analysed, by using event samples based on the WH signal ($m_H = 120 \text{ GeV}$). Table C.5 shows the projected number of events in 30 fb^{-1} passing the single selection cuts, in both the ATLFast-II and full simulation event samples, together with a ratio illustrating their compatibility.

The overall signal efficiency in ATLFast-II is $(11 \pm 7)\%$ higher than in full simulation, after the full *loose* selection, except for the mass window cut, is applied. Starting from the events which passed the generator filter selection, the special jet clustering procedure outlined in Section 7.4 is applied, where exactly one jet candidate with $p_T > 200 \text{ GeV}$ splitting in a

	WH(120) ATLFAST-II	WH(120) FullSim	Ratio
After filter cuts	1253 \pm 8	1253 \pm 12	1.00 \pm 0.01
1 Higgs candidate	569.7 \pm 3.0	542.8 \pm 4.6	1.05 \pm 0.01
filtered $p_T > 200$ GeV	512.7 \pm 3.2	490.2 \pm 4.8	1.05 \pm 0.01
$E_T^{\text{miss}} > 30$ GeV	362.4 \pm 3.2	353.5 \pm 4.8	1.03 \pm 0.02
$p_T(W) > 200$ GeV	171.0 \pm 2.6	164.4 \pm 3.8	1.04 \pm 0.03
$p_T(e/\mu) > 30$ GeV	145.6 \pm 2.4	137.3 \pm 3.6	1.06 \pm 0.03
$p_T(\text{additional } \mu) < 10$ GeV	144.6 \pm 2.4	136.6 \pm 3.6	1.06 \pm 0.03
$p_T(\text{additional } e) < 10$ GeV	142.9 \pm 2.4	134.1 \pm 3.5	1.07 \pm 0.03
$\Delta\phi(W,H) > \frac{2}{3}\pi$	142.2 \pm 2.4	133.1 \pm 3.5	1.07 \pm 0.03
no additional b -jets $p_T > 15$ GeV	130.6 \pm 2.3	119.3 \pm 3.4	1.09 \pm 0.04
jets on W side $p_T < 60$ GeV	115.7 \pm 2.2	107.0 \pm 3.2	1.08 \pm 0.04
jets on H side $p_T < 60$ GeV	102.7 \pm 2.1	95.1 \pm 3.0	1.08 \pm 0.04
one subjet b -tagged	91.4 \pm 2.0	85.9 \pm 2.9	1.06 \pm 0.04
both subjets b -tagged	45.6 \pm 1.4	41.2 \pm 2.1	1.11 \pm 0.07
loose fit cuts	45.4 \pm 1.4	41.0 \pm 2.1	1.11 \pm 0.07
112 GeV $< \text{mass}(H) < 136$ GeV	29.8 \pm 1.2	27.0 \pm 1.7	1.10 \pm 0.08

Table C.5: The table shows the expected number of events going through the selection at each stage for the signal produced through ATLFAST-II compared to full simulation. The numbers are projected to 30 fb⁻¹ of collected data.

sufficiently symmetric way in two nearby lying jets is required to exist in the event. This yields the filtered mass distribution of Fig. C.9. No lepton in the event is required yet.

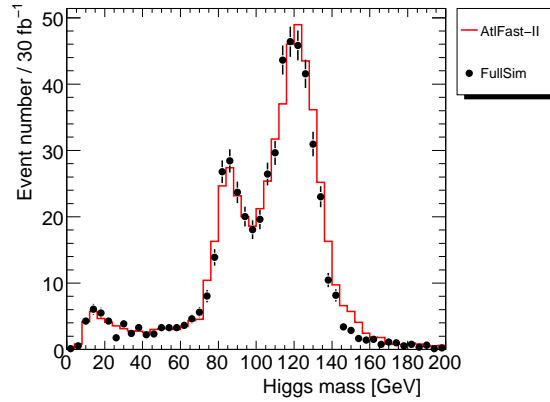


Figure C.9: Comparison of filtered mass of Higgs boson candidate (in GeV) for the WH signal produced with ATLFAST-II and full simulation. The distributions are normalized to an integrated luminosity of 30 fb⁻¹.

Since no constraint is applied at generator level to the signal events to force the W boson to decay into a lepton and a neutrino, a second mass peak is visible on the left of the Higgs boson mass peak, corresponding to a W boson decaying hadronically. These events are, however, removed in a later stage of the analysis selection.

In ATLFAST-II $5 \pm 1\%$ more mono-jets representing Higgs boson candidates are found. However, the invariant mass distribution is reproduced fairly well; ATLFAST-II shows a slightly increased right shoulder in the Higgs boson mass, which is not completely trivial to explain only by the slightly different jet energy scale calibration in ATLFAST-II with respect to full simulation, which was illustrated in Section 7.9. In order to understand to what kinematic region for the Higgs boson candidate this corresponds to, Fig. C.10 shows

the filtered p_T distribution in case the Higgs boson mass is between 100 and 150 GeV, while Fig. C.11 shows the distribution for the ΔR between the two highest p_T subjets and the asymmetry parameter y —which triggers the splitting condition during jet clustering, in the same Higgs boson mass window. Again, no requirement on the W boson side is applied yet. The difference is mainly in candidates close to the 200 GeV p_T threshold, corresponding to a larger aperture between the two Higgs subjets ΔR and a more symmetric configuration of the two subjets. This is however also the region which is most affected by small changes to the jet energy scale, due to the p_T cut at 200 GeV: it is therefore not excluded that a proper jet energy calibration of the subjets in ATLFAST-II would solve this inconsistency.

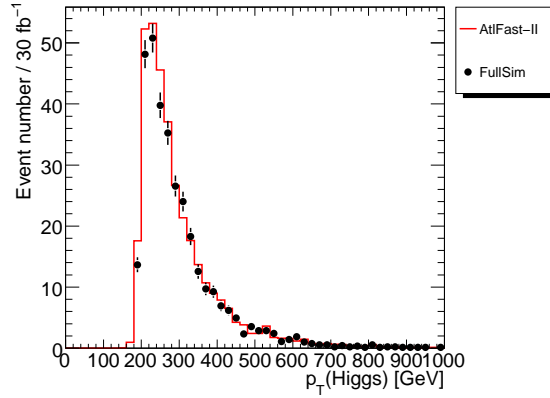


Figure C.10: Comparison of filtered p_T of Higgs boson candidate (in GeV) for the WH signal produced with ATLFAST-II and full simulation. The distributions are normalized to an integrated luminosity of 30 fb^{-1} .

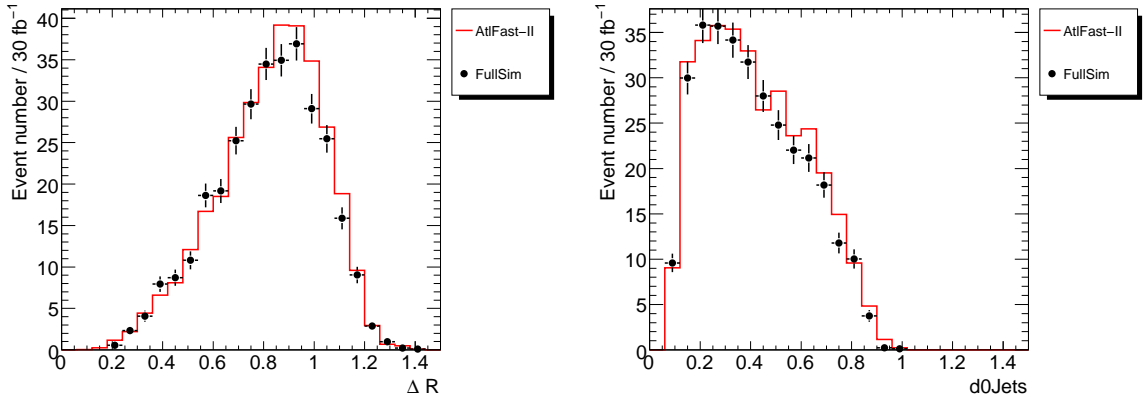


Figure C.11: Distributions for ΔR (left) and d_0 (y_{clust}) between the two highest p_T subjets of the Higgs boson candidate. The distributions are normalized to an integrated luminosity of 30 fb^{-1} .

The E_T^{miss} distribution, normalized to the number of entries in the histogram, is shown in Fig. C.12, after applying the Higgs boson candidate selection: for ATLFAST-II it has more entries at lower E_T^{miss} values, most probably because a lower p_T Higgs boson corresponds to a lower p_T back to back W boson.

Fig. C.13 shows the p_T distribution for the p_T of the hardest lepton, for electrons (on the left) and for muons (on the right), again straight after the Higgs boson selection. Since these

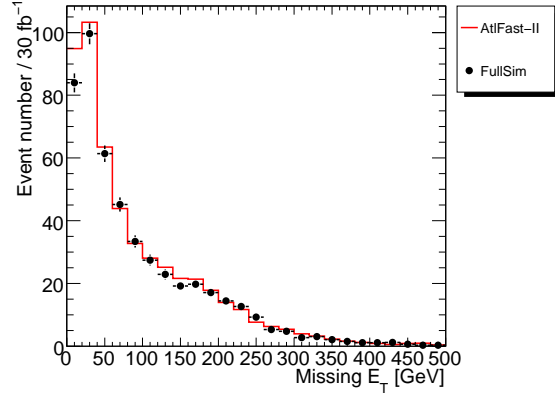


Figure C.12: Comparison of E_T^{miss} in the event for the WH signal sample produced with ATLFast-II and the corresponding one produced with full simulation. The distributions are normalized to an integrated luminosity of 30 fb^{-1} .

distributions are again normalized to the number of their respective entries, they hide the effect of the different efficiency for the two different lepton families: at the point where the cut on the p_T of the leptons is applied, in the ATLFast-II sample $(49.2 \pm 0.9)\%$ of the events contain an electron, the remaining ones a muon, while in the full simulation sample $(48.7 \pm 1.4)\%$ of the events contains an electron, the remaining ones a muon. Applying the complete Higgs boson candidate selection decreases the statistics significantly, so that a proper comparison of the electron and muon efficiencies is not possible. However this comparison was performed already in Section C.4.1. According to it, the muon efficiency is expected to be comparable in ATLFast-II and in full simulation, while the electron efficiency (medium quality electron identification requirement) is expected to be higher in ATLFast-II by a factor $(3.2 \pm 1.0)\%$.

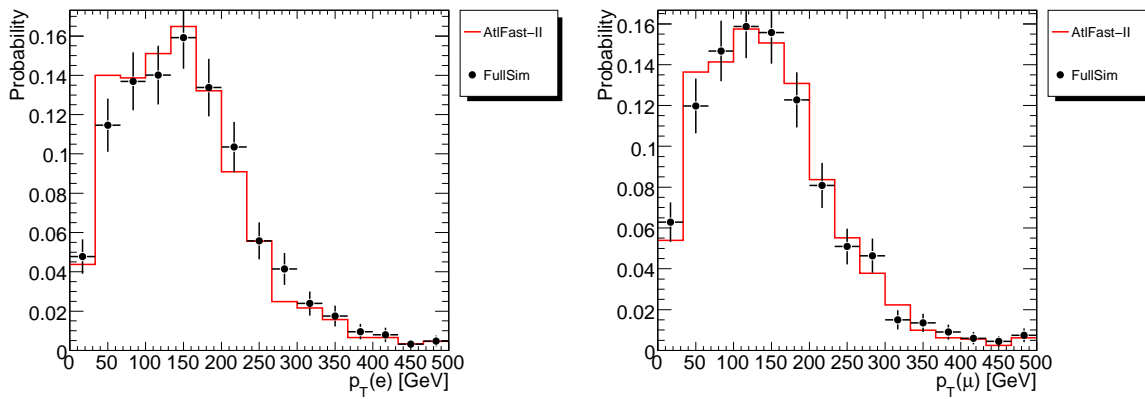


Figure C.13: Distributions for the p_T of the hardest lepton in the events, for electrons (left) and for muons (right), after the Higgs boson selection cuts have already been applied. The distributions are normalized to an area of one.

Fig. C.14 shows the $\Delta\phi(\text{Higgs}, W)$ variable, after both Higgs and W bosons selections are applied. The two distributions agree well in the limit of the available statistics. Fig. C.15 shows the comparison for the p_T of additional b-jets in the event.

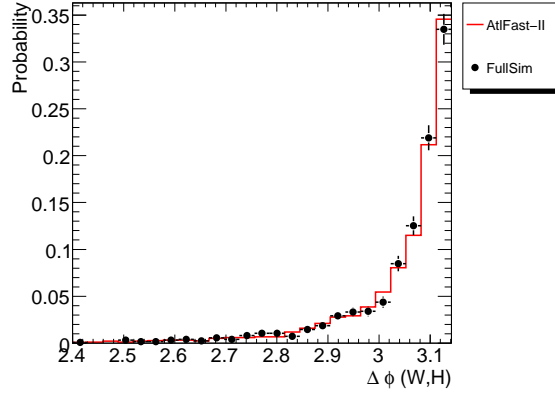


Figure C.14: Comparison of $\Delta\phi(\text{Higgs}, W)$ in ATLFAST-II and full simulation, after the Higgs and W bosons selection cuts have been applied. The distributions are normalized to an area of one.

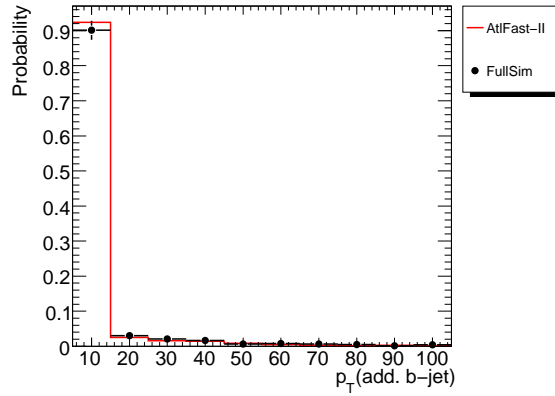


Figure C.15: Comparison of $p_T(\text{add } b\text{-jet})$ variable (in GeV) in ATLFAST-II and full simulation, after all previous selection cuts have been applied. The distributions are normalized to an area of one.

Fig. C.16 shows the comparison for the $p_T(\text{additional jet})$ distribution, where the additional jet is on the side of the W boson (left) or on the side of the Higgs boson (right). All selection cuts before the light jet veto are applied for both distributions. The agreement is reasonable in both cases.

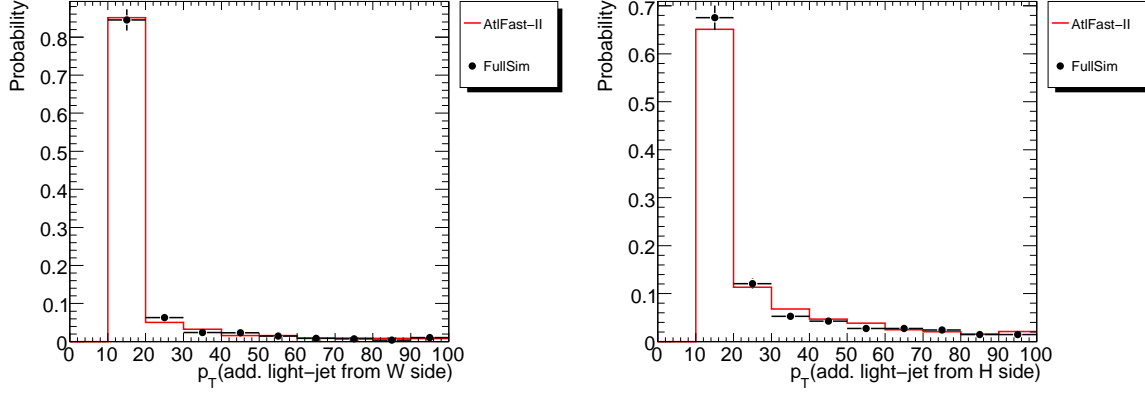


Figure C.16: Distributions for p_T of the hardest additional jet (in GeV) on the W boson side (left) and on the Higgs boson side (right) for ATLFAST-II and full simulation. Distributions are normalized to the number of entries.

Fig. C.17 shows the comparison for the b -weight distribution, after initial Higgs and W bosons selection and after the $\Delta\phi > \frac{2}{3}$ requirement. The jet vetoes are not applied in order to increase the available statistics. The distributions agree very well, as is expected since both simulations make use of the full GEANT4 simulation of the inner detector.

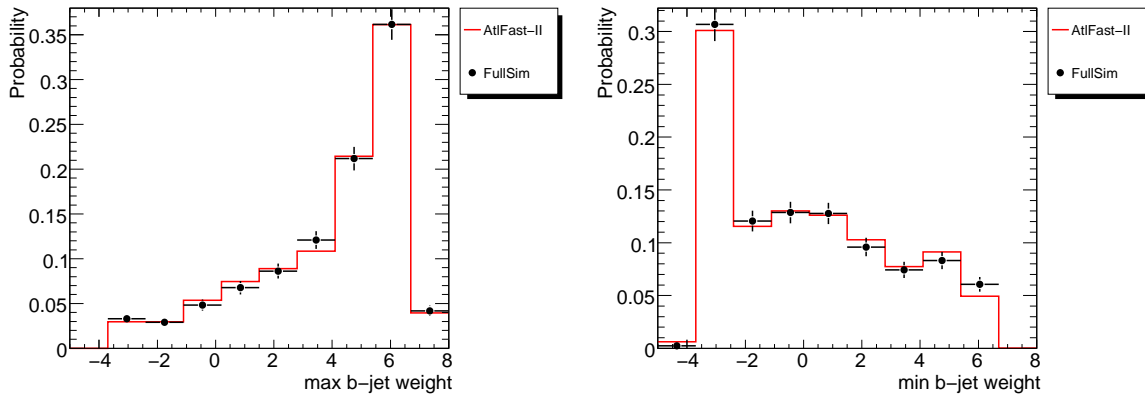


Figure C.17: The distributions for maximum and minimum b -weight of the two subsets, for both ATLFAST-II and full simulation. Distributions are normalized to the number of entries.

Fig. C.18 shows the b -weight distribution of the third subset, after initial Higgs and W bosons selection, after the $\Delta\phi > \frac{2}{3}$ cut and after the b -tagging requirement is applied on the first two subsets.

Finally Fig. C.19 shows the filtered mass distribution just after the initial W and H boson selection and the $\Delta\phi(W,H)$ cut (left) and after all selection cuts of the *loose* selection. The expected number of events in 30 fb^{-1} is shown.

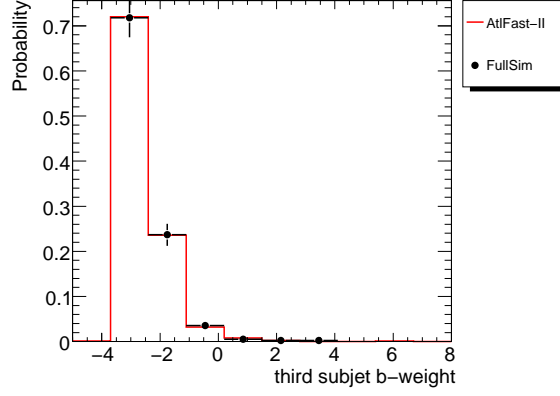


Figure C.18: Comparison of the b -weight of the third subject in ATLFast-II and in full simulation. The distributions are normalized to the number of entries.

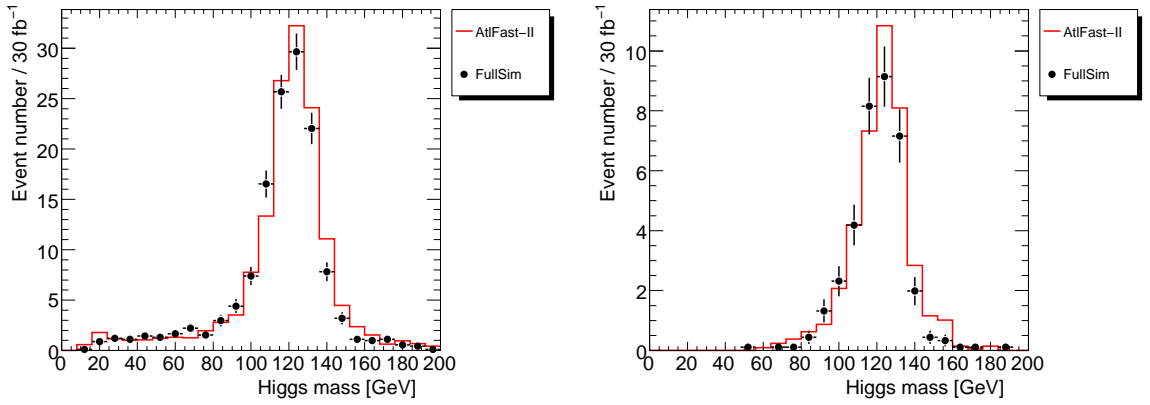


Figure C.19: The distributions for the filtered mass before b -Tagging and jet veto cuts (left) and afterwards (right). The distributions are normalized to an integrated luminosity of 30 fb^{-1} .

In order to compare the shape of the invariant mass distribution in ATLFAST-II and full simulation, the final mass distribution is also shown after normalizing it to the total number of events in Fig. C.20. In addition a Gaussian fit is performed iteratively, until the $\pm 1.5\sigma$ region around the fitted mean stabilizes and remains equal to the interval chosen for the fit. Apart from a small overall shift to higher values, due to the different calibration of the jet energy scale (the means of the Gaussian fit turn out to be $\mu(\text{ATLFAST-II}) = 123.0 \pm 0.4$, $\mu(\text{full simulation}) = 121.5 \pm 0.7$), the shape of the Higgs boson invariant mass is reproduced very well in ATLFAST-II and the resolutions are compatible.

In the limit of the available statistics of the signal samples used for the present comparison, the overall analysis efficiency for signal events produced with ATLFAST-II agrees reasonably well with full simulation. The small discrepancy between ATLFAST-II and full simulation seen in the earlier steps of the analysis selection can be corrected for taking into account the following discrepancies:

- A reduced Higgs mono-jet reconstruction efficiency by $(-5 \pm 1)\%$
- A reduced electron efficiency by $(-3.1 \pm 1.4)\%$, corresponding to an overall reduced efficiency by $\approx -1.5\%$

These correction factors are taken into account in the likelihood fit based analysis described in the Chapter 8. In the first case, the very conservative hypothesis will be made that this inefficiency affects only the signal sample. It is however worth mentioning that a proper jet energy scale calibration in ATLFAST-II is not unlikely to solve the first of the two inconsistencies.

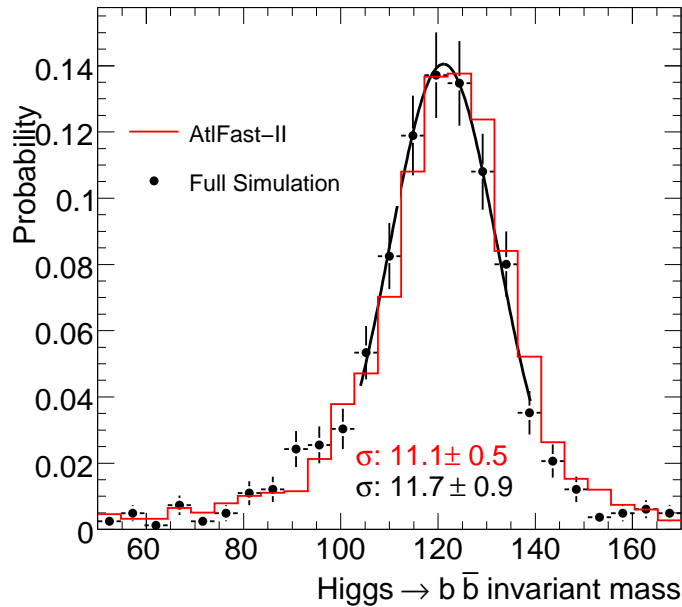


Figure C.20: Comparison between the final mass distribution (in GeV) for signal as produced with ATLFAST-II compared to full simulation.

D WH analysis: fit based approach

D.1 Cross-check of correlations

One of the crucial assumptions made in the WH fit based analysis is the requirement that the distributions of the discriminant variables used in the likelihood function, separately for each signal and background component, are uncorrelated. As illustrated in Section 8.2.3, the linear correlation values seem to be all smaller than around $\approx 12\%$, however a more detailed cross-check is needed, since in principle more complex correlations between variables can appear, such that they result in a negligible linear correlation. This is the scope of this appendix.

The main signal and background contributions are considered separately. Given a certain discriminating variable, the likelihood ratio is computed on an event basis, based on the definition contained in Eq. 8.21, where the dependence of the likelihood function on the variable of interest has first been integrated out. This likelihood ratio provides then the optimal discrimination between signal and background, based on all remaining variables: combining on an event basis this likelihood ratio with an additional likelihood ratio based exclusively on the variable of interest would provide the same discriminating power as the original likelihood function. If the above defined likelihood ratio is uncorrelated with the distribution of the variable of interest, then the effect of correlations between this variables and the remaining fit variables can be considered as negligible.

The likelihood function used here does not include the Poisson term and is computed on a single event basis: its parameters and yields are fixed to the values determined in the fit performed on the mixed Monte Carlo sample, as shown in Table 8.8. The correlation between the likelihood ratio and the variable of interest is cross-checked in the following way: given a certain variable of interest and a certain signal or background component, the Monte Carlo sample is subdivided in two equal regions, by applying a cut on the likelihood ratio, one is called *signal like* ($LR > \text{cut value}$), the other is called *background like* ($LR < \text{cut value}$), then the distribution for the variable of interest is analysed and compared in the two different regions. If the correlations are negligible, the two distributions must be compatible within statistics. The ratio between the two distributions is shown in order to evaluate the compatibility of the two distributions: in addition, the result of a Kolmogoroff-Smirnov test of compatibility, applied directly on the original unbinned distributions, is also quoted.

In the following the results are shown. For all signal and background components and all variables of interest, in the limit of the available statistics, the more signal and background like distributions agree fairly well. Clearly for some of the samples a higher Monte Carlo statistics would be highly desirable.

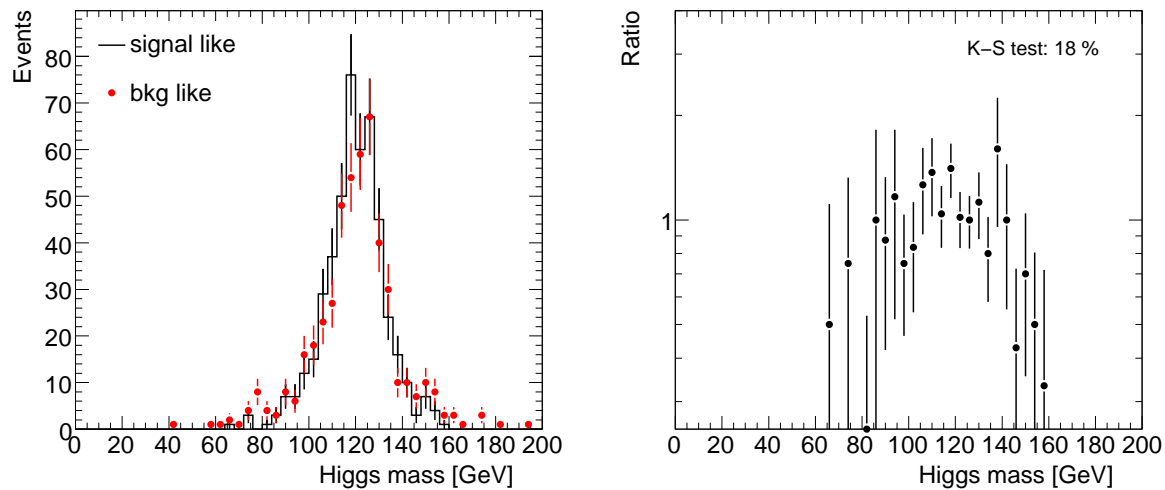


Figure D.1: WH signal: (Left) Distribution for the invariant mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

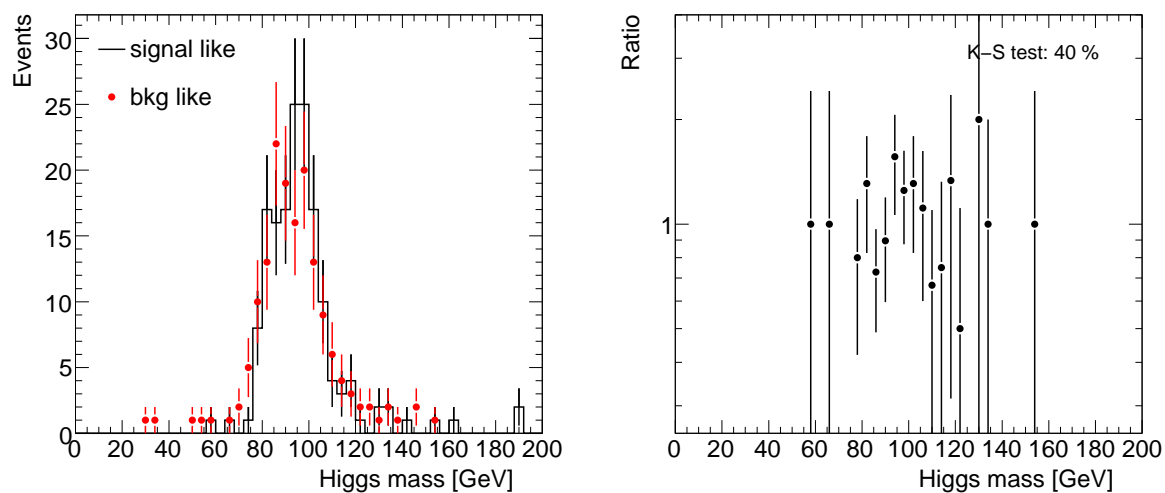


Figure D.2: WZ background: (Left) Distribution for the invariant mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

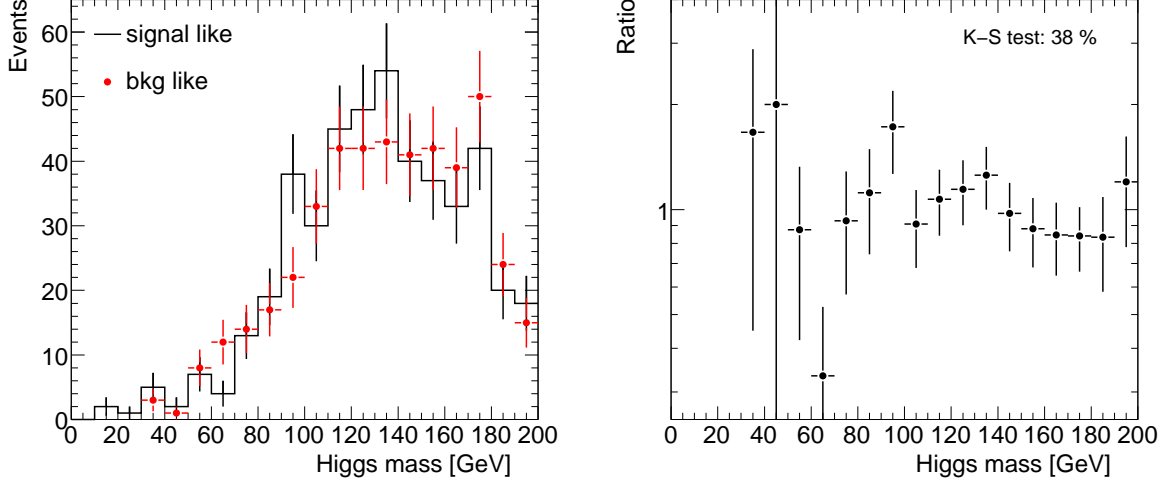


Figure D.3: Top pair background: (Left) Distribution for the invariant mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

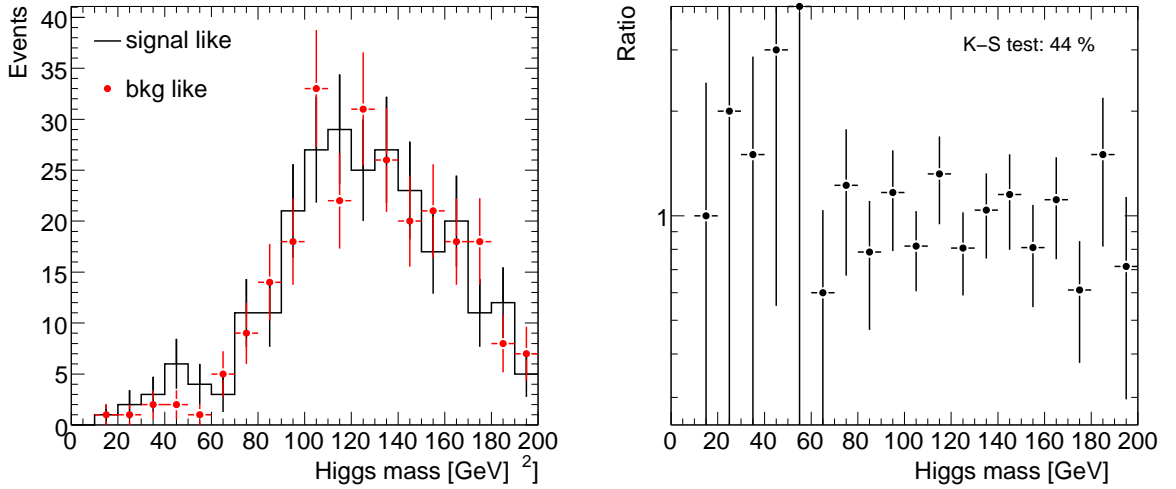


Figure D.4: Single top background: (Left) Distribution for the invariant mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

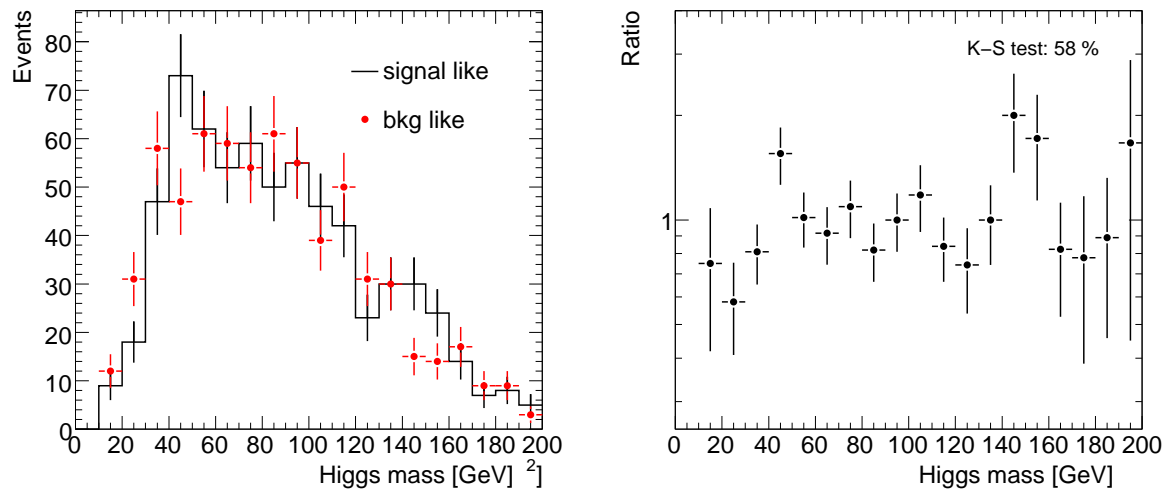


Figure D.5: $Wb\bar{b}$ background: (Left) Distribution for the invariant mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

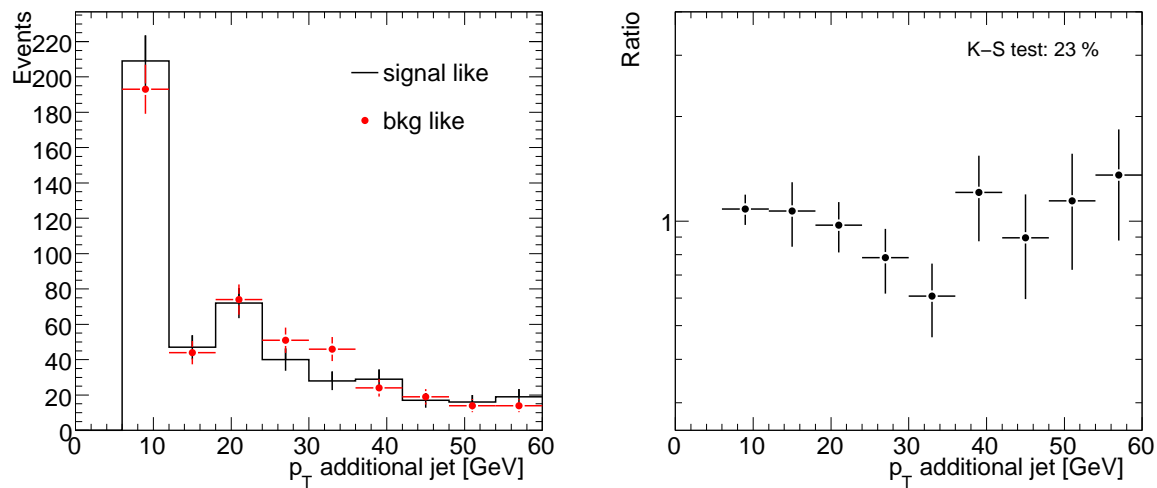


Figure D.6: WH signal: (Left) Distribution for the p_T of additional jets in the event, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

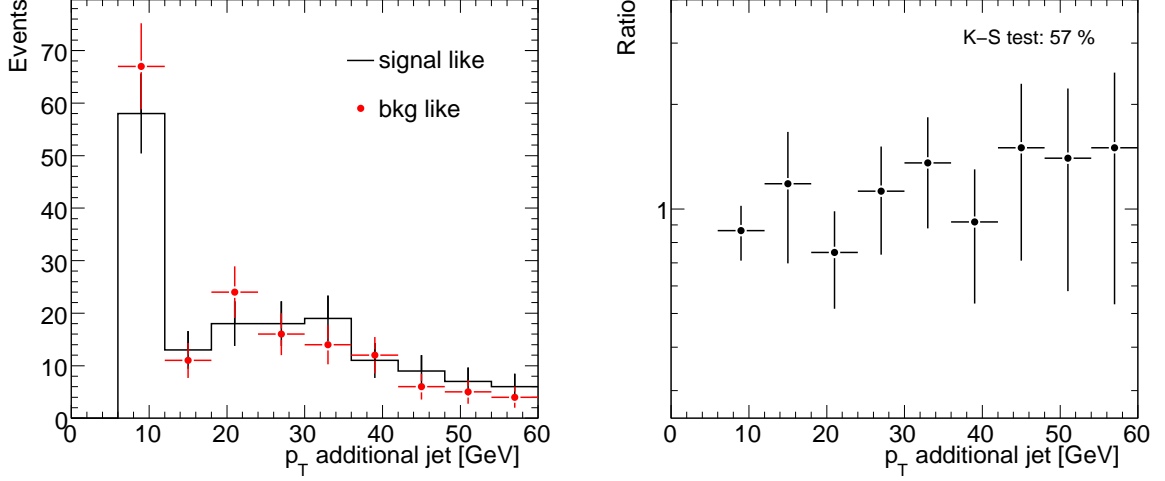


Figure D.7: WZ background: (Left) Distribution for the the p_T of additional jets in the event, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

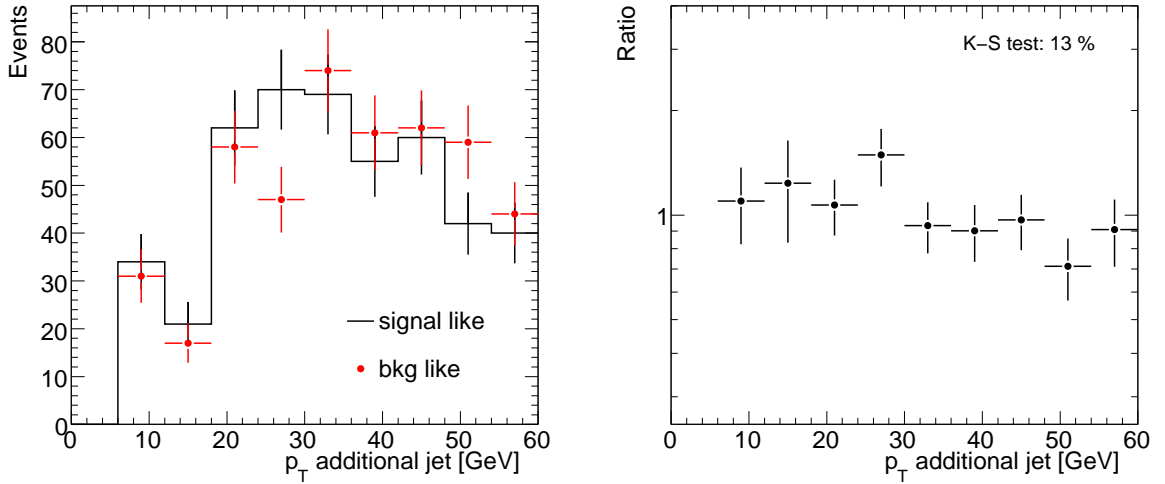


Figure D.8: Top pair background: (Left) Distribution for the the p_T of additional jets in the event, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

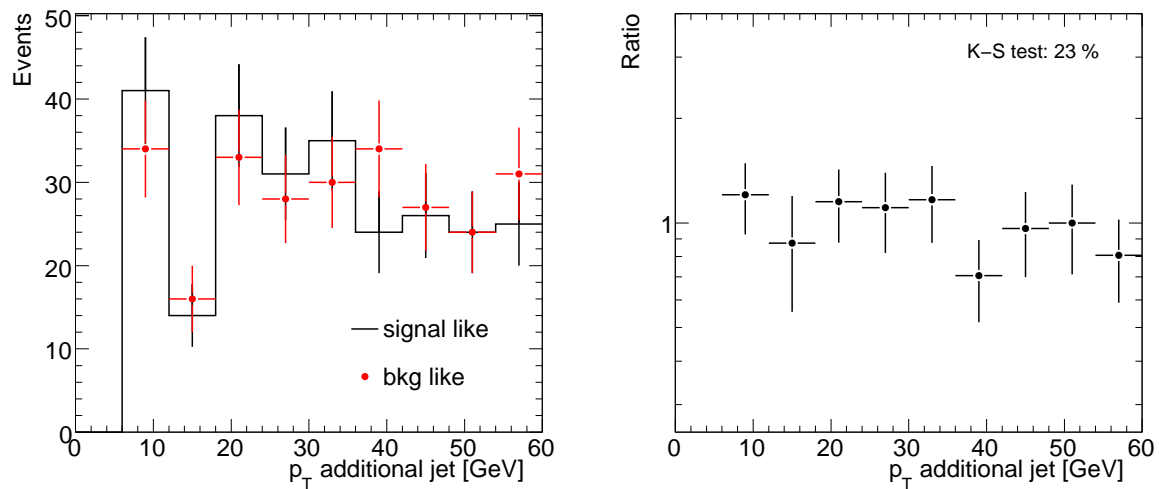


Figure D.9: Single top background: (Left) Distribution for the the p_T of additional jets in the event, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

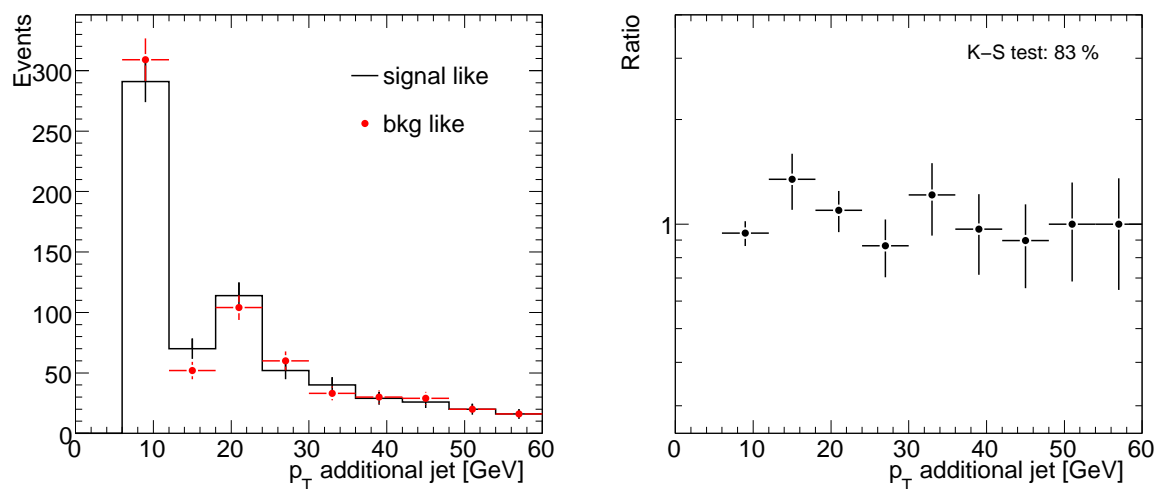


Figure D.10: $Wb\bar{b}$ background: (Left) Distribution for the the p_T of additional jets in the event, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

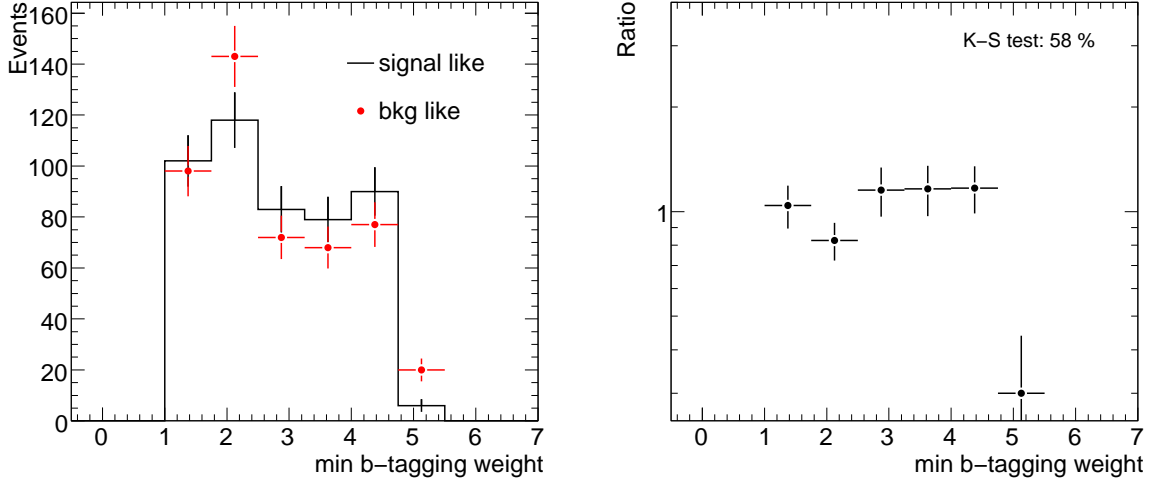


Figure D.11: WH signal: (Left) Distribution for the b -tagging weight, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

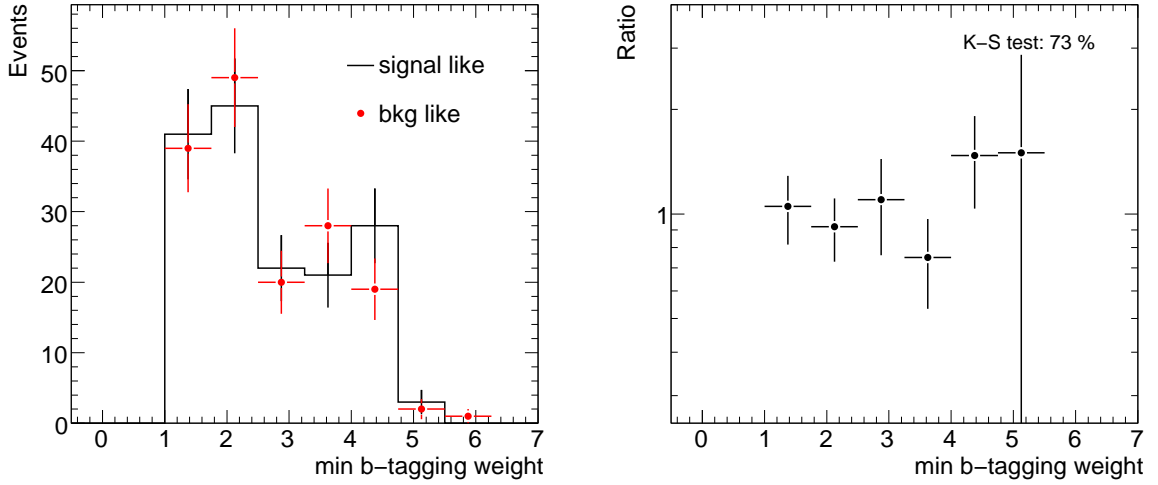


Figure D.12: WZ background: (Left) Distribution for the b -tagging weight, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

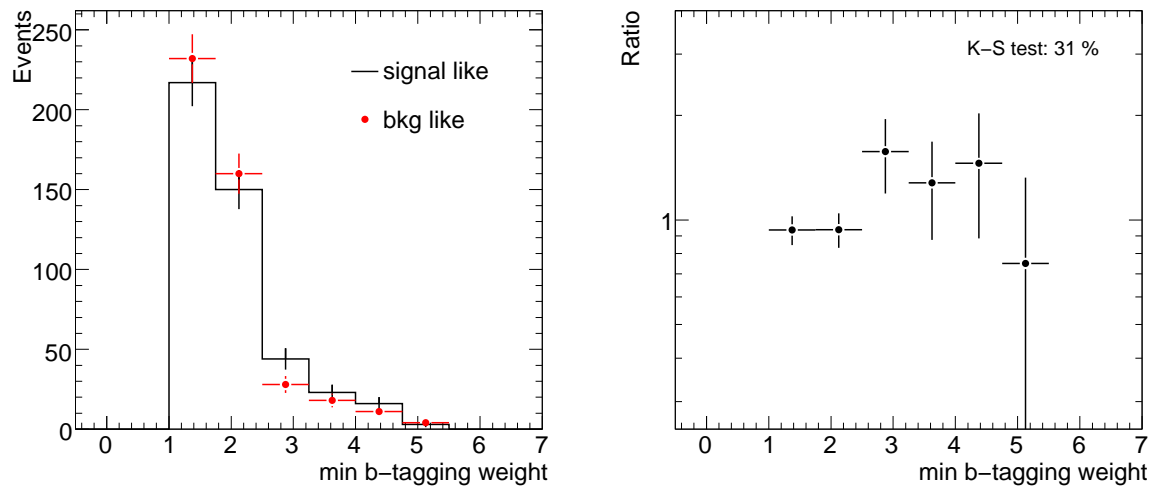


Figure D.13: Top pair background: (Left) Distribution for the b -tagging weight, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

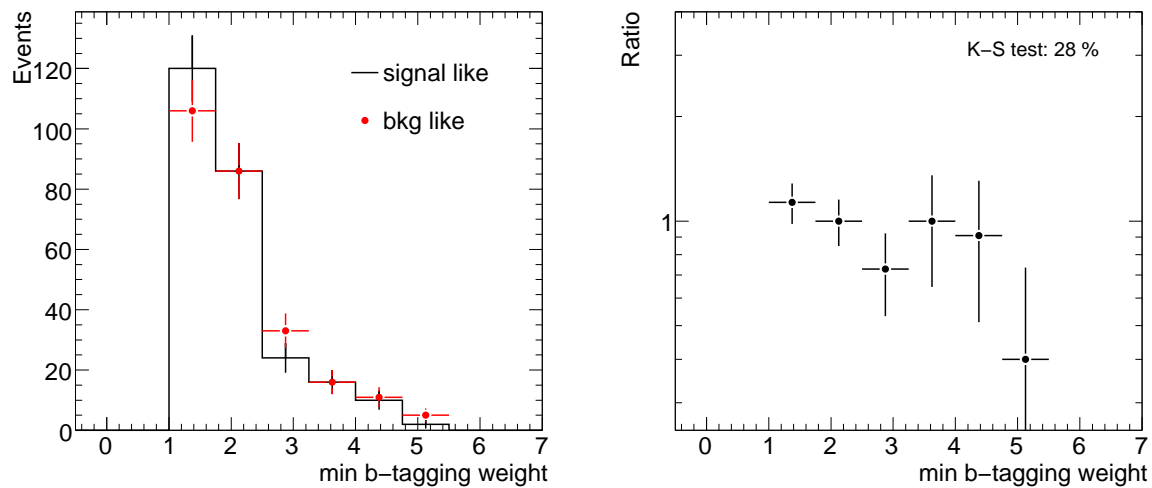


Figure D.14: Single top background: (Left) Distribution for the b -tagging weight, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

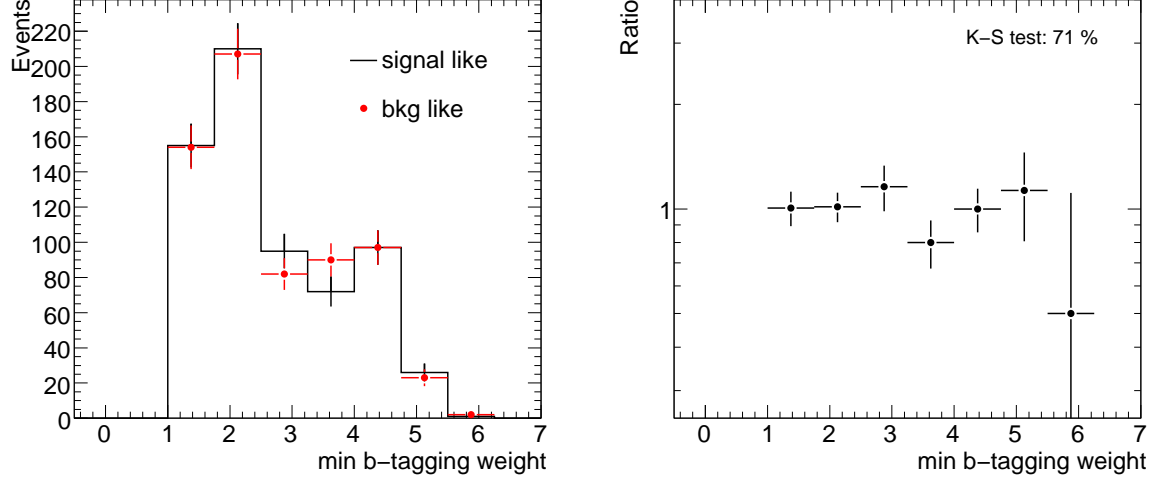


Figure D.15: $Wb\bar{b}$ background: (Left) Distribution for the b -tagging weight, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

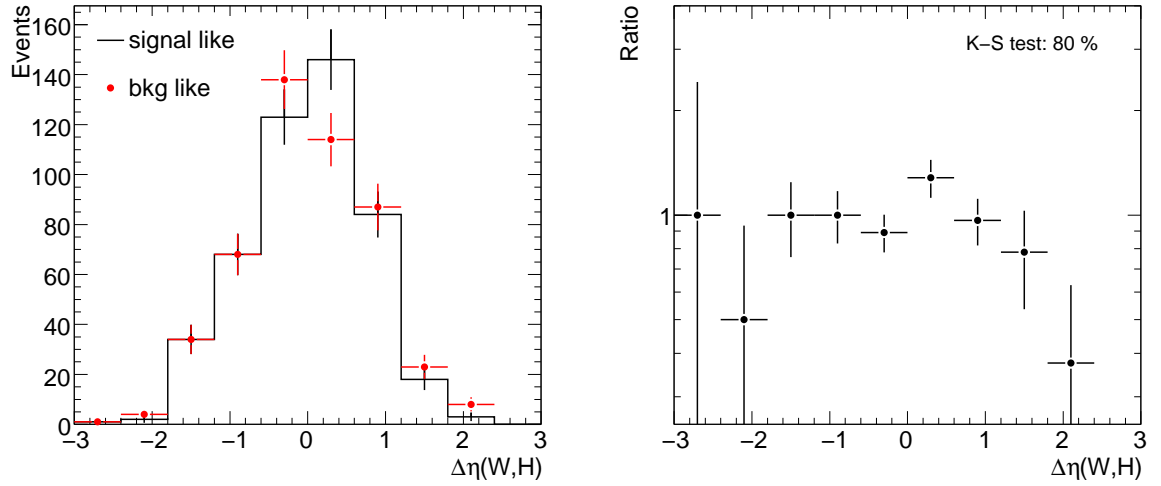


Figure D.16: WH signal: (Left) Distribution for the $\Delta\eta(W,H)$ variable, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

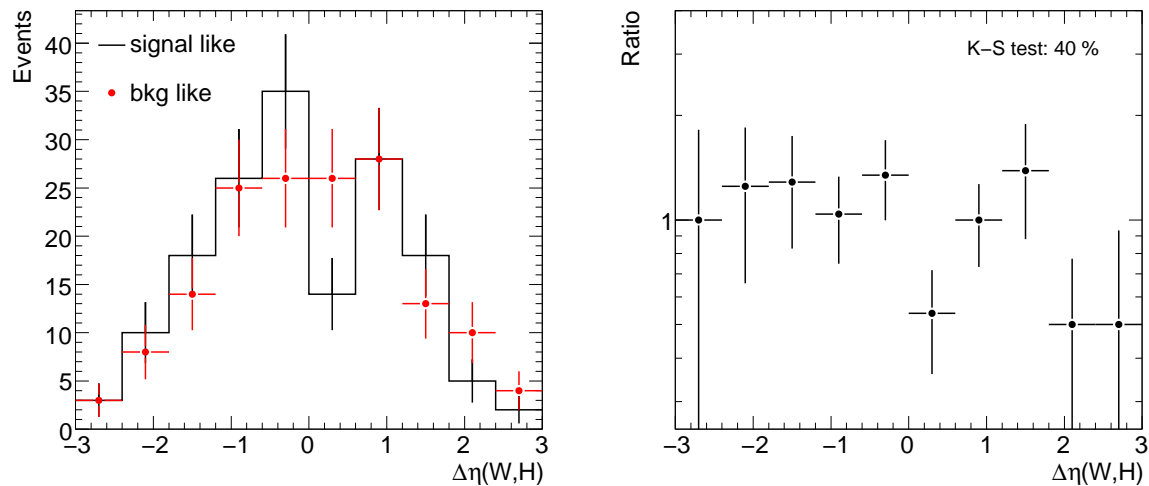


Figure D.17: WZ background: (Left) Distribution for the $\Delta\eta(W,H)$ variable, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

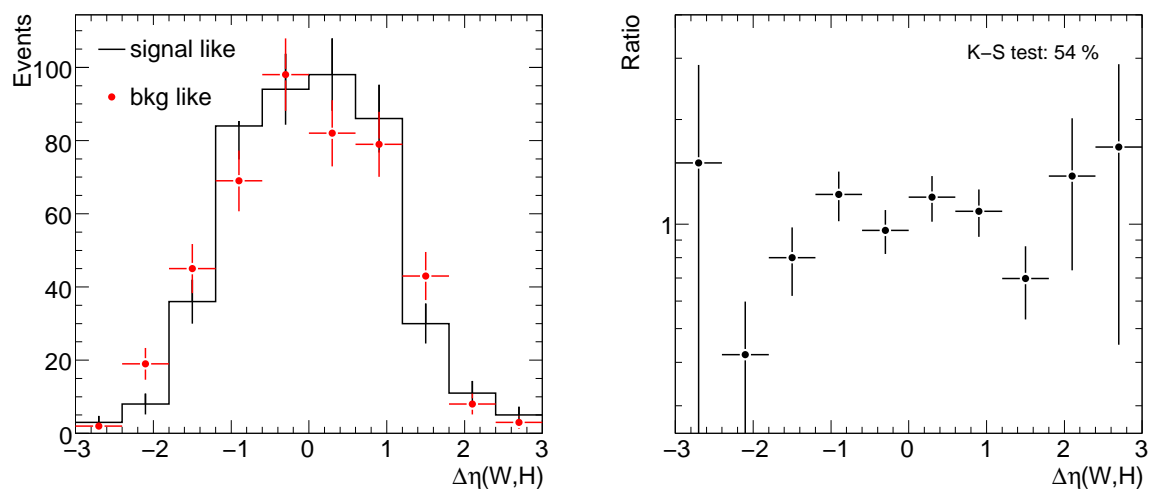


Figure D.18: Top pair background: (Left) Distribution for the $\Delta\eta(W,H)$ variable mass of the Higgs candidate, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

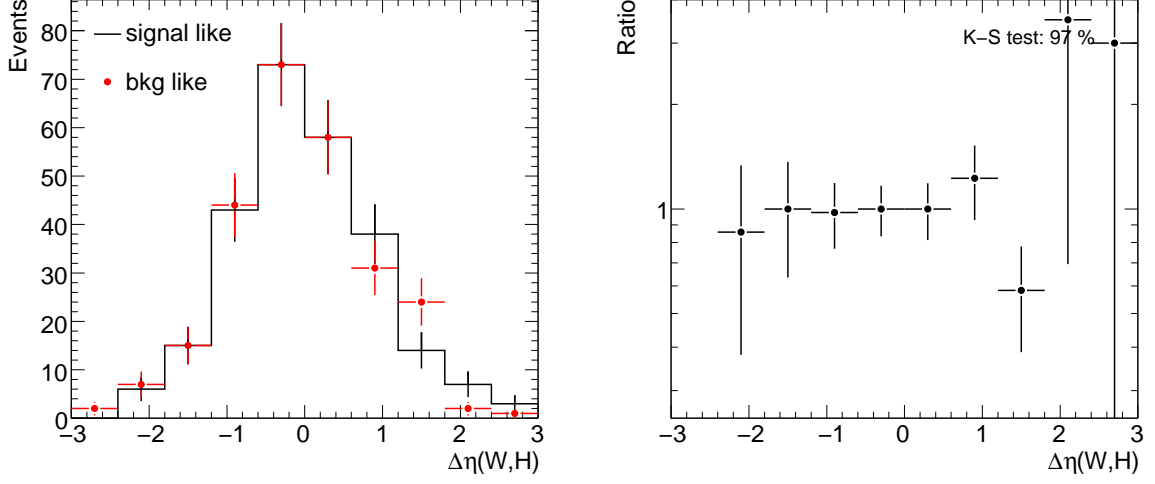


Figure D.19: Single top background: (Left) Distribution for the $\Delta\eta(W,H)$ variable, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

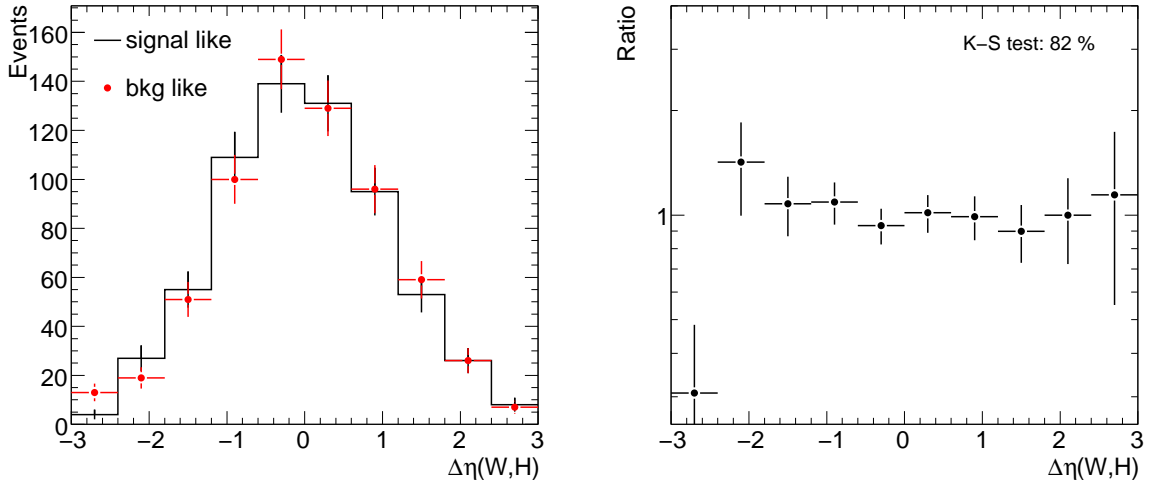


Figure D.20: $Wb\bar{b}$ background: (Left) Distribution for the $\Delta\eta(W,H)$ variable, in the most signal and background like regions of the fit. (Right) Corresponding ratio of distributions.

Bibliography

- [1] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008), [0802.2470].
- [2] S. L. Glashow, Nucl. Phys. **22**, 579 (1961).
- [3] A. Salam and J. C. Ward, Phys. Lett. **13**, 168 (1964).
- [4] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).
- [5] H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. **B47**, 365 (1973).
- [6] D. J. Gross and F. Wilczek, Phys. Rev. **D8**, 3633 (1973).
- [7] H. D. Politzer, Phys. Rept. **14**, 129 (1974).
- [8] P. W. Higgs, Phys. Lett. **12**, 132 (1964).
- [9] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).
- [10] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory (Frontiers in Physics)* (Westview Press, 1995).
- [11] C. F. Kolda and H. Murayama, JHEP **07**, 035 (2000), [hep-ph/0003170].
- [12] N. Cabibbo, L. Maiani, G. Parisi and R. Petronzio, Nucl. Phys. **B158**, 295 (1979).
- [13] A. Djouadi, Phys. Rept. **457**, 1 (2008), [hep-ph/0503172].
- [14] ALEPH, 0811.4682.
- [15] LEP Working Group for Higgs boson searches, R. Barate *et al.*, Phys. Lett. **B565**, 61 (2003), [hep-ex/0306033].
- [16] CDF, 0903.4001.
- [17] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).
- [18] J. C. Collins, D. E. Soper and G. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1 (1988), [hep-ph/0409313].
- [19] F. Krauss, R. Kuhn and G. Soff, JHEP **02**, 044 (2002), [hep-ph/0109036].
- [20] T. Stelzer and W. F. Long, Comput. Phys. Commun. **81**, 357 (1994), [hep-ph/9401258].
- [21] O. Brein, A. Djouadi and R. Harlander, Physics Letters B **579**, 149 (2004).
- [22] G. Altarelli, R. K. Ellis and G. Martinelli, Nucl. Phys. **B157**, 461 (1979).
- [23] J. Kubar-Andre and F. E. Paige, Phys. Rev. **D19**, 221 (1979).

- [24] R. Hamberg, W. L. van Neerven and T. Matsuura, Nucl. Phys. **B359**, 343 (1991).
- [25] M. L. Ciccolini, S. Dittmaier and M. Kramer, Phys. Rev. **D68**, 073003 (2003), [hep-ph/0306234].
- [26] O. Brein *et al.*, hep-ph/0402003.
- [27] M. Cacciari and M. Greco, Nucl. Phys. **B421**, 530 (1994), [hep-ph/9311260].
- [28] S. Catani, L. Trentadue, G. Turnock and B. R. Webber, Nucl. Phys. **B407**, 3 (1993).
- [29] A. Banfi, G. P. Salam and G. Zanderighi, JHEP **03**, 073 (2005), [hep-ph/0407286].
- [30] G. Altarelli, Phys. Rept. **81**, 1 (1982).
- [31] R. K. Ellis, W. J. Stirling and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996).
- [32] T. Sjostrand, S. Mrenna and P. Skands, JHEP **05**, 026 (2006), [hep-ph/0603175].
- [33] G. Corcella *et al.*, hep-ph/0210213.
- [34] S. Frixione and B. R. Webber, JHEP **06**, 029 (2002), [hep-ph/0204244].
- [35] S. Frixione, P. Nason and C. Oleari, JHEP **11**, 070 (2007), [0709.2092].
- [36] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **11**, 063 (2001), [hep-ph/0109231].
- [37] M. L. Mangano, M. Moretti and R. Pittau, Nucl. Phys. **B632**, 343 (2002), [hep-ph/0108069].
- [38] X. Artru and G. Mennessier, Nucl. Phys. **B70**, 93 (1974).
- [39] T. Sjostrand and M. van Zijl, Phys. Rev. **D36**, 2019 (1987).
- [40] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. **C72**, 637 (1996), [hep-ph/9601371].
- [41] A. Moraes, C. Buttar and I. Dawson, Eur. Phys. J. **C50**, 435 (2007).
- [42] ATLAS, Detector physics performance technical design report, CERN/LHCC/99-14/15, 1999.
- [43] A. Djouadi, J. Kalinowski and M. Spira, Comput. Phys. Commun. **108**, 56 (1998), [hep-ph/9704448].
- [44] The ATLAS Collaboration, G. Aad *et al.*, arXiv:0901.0512.
- [45] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, Physics Letters B **556**, 50 (2003).
- [46] E. Gabrielli *et al.*, Nucl. Phys. **B781**, 64 (2007), [hep-ph/0702119].
- [47] T. Plehn, G. P. Salam and M. Spannowsky, 0910.5472.
- [48] K. Jakobs, Eur. Phys. J. **C59**, 463 (2009).

- [49] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duehrssen, arXiv:0904.3866 [hep-ph].
- [50] M. Duhrssen *et al.*, Phys. Rev. **D70**, 113009 (2004), [hep-ph/0406323].
- [51] M. Duhrssen, ATL-PHYS-2003-030.
- [52] O. Buning, (Ed.) *et al.*, CERN-2004-003.
- [53] ATLAS Collaboration, G. Aad *et al.*, JINST **3**, S08003 (2008).
- [54] T. Cornelissen *et al.*, Concepts, Design and Implementation of the ATLAS New Tracking (NEWT).
- [55] R. Frühwirth, Comp. Phys. Comm. **100**, 1 (1997).
- [56] V. Kartvelishvili, Nucl. Phys. B (Proc. Suppl.) **172**, 208 (2007).
- [57] G. P. Salam, 0906.1833.
- [58] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, Phys. Rev. Lett. **101**, 142001 (2008), [0806.0848].
- [59] H1 Collaboration, I. Abt *et al.*, Nucl. Instrum. Meth. **A386**, 348 (1997).
- [60] T. G. Cornelissen *et al.*, ATL-SOFT-PUB-2007-003.
- [61] P. Billoir and S. Qian, Nucl. Instrum. Meth. **A311**, 139 (1992).
- [62] R. Fruhwirth, Nucl. Instrum. Meth. **A262**, 444 (1987).
- [63] P. Billoir, Nucl. Instrum. Meth. **A225**, 352 (1984).
- [64] W. D. Hulsbergen, Nucl. Instrum. Meth. **A552**, 566 (2005), [hep-ph/0503191].
- [65] A. Salzburger, ATL-SOFT-PUB-2007-005.
- [66] E. Lund, L. Bugge, I. Gavrilenko and A. Strandlie, JINST **4**, P04001 (2009).
- [67] R. Fruhwirth, W. Waltenberger and P. Vanlaer, J. Phys. **G34**, N343 (2007).
- [68] W. Waltenberger and R. Fruhwirth, Adaptive multi-vertex fitting (Proceedings for CHEP 2004).
- [69] J. Bronner, Freiburg diploma thesis, 2009.
- [70] V. Kostyukhin, ATL-PHYS-2003-031.
- [71] Particle Data Group, C. Amsler *et al.*, Phys. Lett. **B667**, 1 (2008).
- [72] N. Cabibbo, Phys. Rev. Lett. **12**, 62 (1964).
- [73] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
- [74] SLD, K. Abe *et al.*, Contributed to 19th International Symposium on Lepton and Photon Interactions at High-Energies (LP 99), Stanford, CA, 9-14 Aug 1999.
- [75] V. Kostyukhin, ATL-PHYS-2003-033.

-
- [76] C. Peterson, T. Rognvaldsson and L. Lonnblad, *Comput. Phys. Commun.* **81**, 185 (1994).
 - [77] C. Weiser, Karlsruhe diploma thesis, 1994.
 - [78] E. Richter-Was, ATL-PHYS-2000-018.
 - [79] K. Assamagan *et al.*, ATL-COM-SOFT-2008-024.
 - [80] M. Duerhssen, ATLAS-PHYS-INT-2008-043.
 - [81] G. Corcella *et al.*, *JHEP* **01**, 010 (2001), [hep-ph/0011363].
 - [82] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, *Z. Phys.* **C72**, 637 (1996), [hep-ph/9601371].
 - [83] B. P. Kersevan and E. Richter-Was, hep-ph/0405247.
 - [84] J. Pumplin *et al.*, *JHEP* **07**, 012 (2002), [hep-ph/0201195].
 - [85] CDF, F. Abe *et al.*, *Phys. Rev.* **D54**, 4221 (1996), [hep-ex/9605004].
 - [86] ZEUS, J. Breitweg *et al.*, *Phys. Lett.* **B443**, 394 (1998), [hep-ex/9810046].
 - [87] R. K. Ellis and S. Veseli, *Phys. Rev.* **D60**, 011501 (1999), [hep-ph/9810489].
 - [88] F. Febres Cordero, L. Reina and D. Wackerroth, *Phys. Rev.* **D74**, 034007 (2006), [hep-ph/0606102].
 - [89] J. Campbell, R. K. Ellis and D. L. Rainwater, *Phys. Rev.* **D68**, 094021 (2003), [hep-ph/0308195].
 - [90] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, *JHEP* **08**, 001 (1997), [hep-ph/9707323].
 - [91] M. Wobisch and T. Wengler, hep-ph/9907280.
 - [92] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, *Nucl. Phys.* **B406**, 187 (1993).
 - [93] S. D. Ellis and D. E. Soper, *Phys. Rev.* **D48**, 3160 (1993), [hep-ph/9305266].
 - [94] A. H. Mueller, *Phys. Lett.* **B104**, 161 (1981).
 - [95] B. I. Ermolaev and V. S. Fadin, *JETP Lett.* **33**, 269 (1981).
 - [96] A. Bassetto, M. Ciafaloni and G. Marchesini, *Phys. Rept.* **100**, 201 (1983).
 - [97] D. Cavalli, W. Davey, S. Dean and M. Duerhssen, ATL-PHYS-INT-2009-073.
 - [98] J. M. Butterworth *et al.*, ATL-PHYS-PUB-2009-088.
 - [99] G. Punzi, hep-ph/0401045.
 - [100] P. Billoir and S. Qian, *Nucl. Instrum. Meth.* **A350**, 624 (1994).
 - [101] A. Wildauer and F. Akesson, 2004, Vertex finding and B-tagging algorithms for the ATLAS inner detector (Proceedings for CHEP 2004).
 - [102] F. Maltoni and T. Stelzer, *JHEP* **02**, 027 (2003), [hep-ph/0208156].

Acknowledgements

I'm particularly grateful to my supervisor Karl Jakobs for introducing me to the ATLAS experiment and for the unique opportunity to work in Freiburg in such a stimulating and friendly team of people. It is four years ago, more or less when this thesis work started, that Karl took the decision to join the b -tagging effort in ATLAS. And, just few months later, it was certainly not a case that Christian Weiser, the former b -tagging convener in CMS, joined the team. I'm particularly thankful to Christian, for the fruitful collaboration, in particular on so many different aspects connected to b -tagging and beyond, that we've had in these years and for the continuous and invaluable support, which was a key ingredient for the success of this thesis work. The last year of this PhD thesis, working on Higgs boson decays to b -quarks, was certainly the most exciting part of this adventure, and I'm thankful both to Karl and Christian for having supported this choice and, even more, for the many interesting and stimulating discussions we had on the analysis in the past year.

I will remember these years in Freiburg also for the extremely nice and friendly group atmosphere and I would like to thank all people who made life in Freiburg so enjoyable. I would like to thank in particular few people. Robindra, for the nice time we had together in Freiburg, in the first years before he moved to Bonn. Henrik, for his invaluable sense of humour even in the more critical situations and the many interesting discussions on physics and beyond we had in these years. Andrea, for how she was able to create occasions of informal meetings and discussion and, not less important, for letting me pass on the other, not-isolated, side of the bridge! Michael Heldmann, for all the long discussions on physics and beyond at university, the nice occasions in the evenings, in particular with Natalia, and the great time we had together in Freiburg. Uli, for his good mood and his joking with Christian. Michael Dührssen, for the help in introducing me into the computing administration business, for the challenging discussions on physics, for extending the limits of my physics knowledge and for the strong cooperation we had in the few past years. And Inga, for the nice occasions of discussion and for the pleasant working atmosphere, and for all the great help and support in the last months of writing my thesis. Susanne, for all her help in so many occasions and for being always so active in everything she does. And Johanna, during this last year, for the nice cooperation on primary vertex finding and for being so patient about my lack of time in these last months. I would also like to thank Jochen, for the nice period we had together at CERN, the many occasions of discussions and the nice trips by bike in the Schwarzwald and around the Lake of Geneva. It was also a great pleasure to cooperate with Stan on b -tagging and to learn more about the Tevatron from Ralf, who was also always ready to provide useful suggestions. I spent many nice days biking with Ingo, whose way to approach life is simply unique, and made some wonderful vacations biking with Evelyn, Michael and Susanne. In addition, I would also like to thank Harald, Matti, Markus Schumacher, Markus Warsinsky, Xavier, Mirjam, Stefan, Urban, Jochen D., Liv, Ulav, Peter, Klaus, Frank, Sasha, Matti, Björn, Johannes, Nico, Jörg, Uwe, Simon, Karsten, Andreas, Christoph, Matthew, Michel, Iacopo, Manuela, Kristin and Dieter. Special thanks also to Chris for making administrative requests and problems always easy to solve! And I'm sure I have forgotten someone!

I would also like to thank many colleagues I had the pleasure to work with in ATLAS in these four years. In particular I would like to mention few of them: Kirill Prokofiev and Andi

Wildauer, in the vertexing group, Markus Elsing and Andi Salzburger, in the tracking group, Laurent Vacavant, Vadim Kostyukhyne, Kay Grübel and Marc Lehmacher, in the b -tagging group, and Jon Butterworth and Adam Davison, on highly boosted Higgs boson searches to a pair of b -quarks. I would also like to thank Ketevi Assamagan and Leandro Nisati, for their help and support as Higgs conveners in many occasions.

I had interesting exchanges of ideas in Freiburg with Andrea Ferroglia and Alberto Guffanti, who I would like also to thank for helping in reviewing the first part of this thesis, with Laura Reina and Sally Dawson in Les Houches, discussing on NLO corrections and resummation, and with Tilman Plehn, for the short but very nice time we had in Firenze discussing on $H \rightarrow b\bar{b}$ decays.

I would like also to thank my “old” good friends in Rome, both from the (ex-)BaBar group, where I passed an exciting year working on my master thesis, and from the university courses, where, while attending some of the most stimulating lectures I had in my life, I met several unique people. In particolare vorrei ringraziare i “vecchi” amici dell’università, fra cui in particolare Paolo, Viola, Marcella, Paola, Riccardo, Gigi, Giorgio, Luca, Walter e John, e gli amici di BaBar, tra cui Maurizio, Emanuele, Francesco, Alessia, Luca, Sharam e Daniele. Inoltre, vorrei ringraziare anche i professori Fernando Ferroni e Guido Martinelli, per avermi trasmesso la passione per la fisica delle particelle elementari.

Last but not least, I would like to thank my family. Prima di tutto vorrei ringraziare di cuore mamma e papà, senza i quali raggiungere questo obiettivo sarebbe stato assolutamente impensabile, per l’aiuto e il supporto in questi 28 anni, prima da vicino e, poi, con non poche difficoltà, da sempre più lontano. É sempre bellissimo poter pensare che nelle mie avventure sono accompagnato da quattro fantastici fratelli, Paolo, Marco, Teresa e Carlo Alberto, che spero in futuro di poter andare a trovare un pó più spesso. Voglio anche ringraziare un sacco nonna, zio Dionisio, zio Lino, Giancarlo e Stefania, che mi auguro proprio possano venire a Gennaio a Friburgo. Herzlichen Dank auch für deine Hilfe, Mumi: es war toll ab und zu mal bei Dir in Heilbronn zu sein. Ich bedanke mich auch bei Christof und Petra: ich hoffe ich kann mit euch im nächsten Sommer mal wieder segeln! Herzlichen Dank auch an Gisela: es war einfach toll mit Dir diese Zeit in Freiburg zu verbringen, das werde ich nie vergessen.