Search for the $H \rightarrow b\bar{b}$ decay mode of the Standard Model Higgs boson in associated production with a vector boson in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$ with the ATLAS experiment

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Jan Therhaag
aus
Bonn

Bonn, 2014

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

1. Gutachter:	Prof. Dr. Norbert Wermes
2. Gutachter:	Prof. Dr. Klaus Desch
Tag der Promotion:	14.03.2014
Erscheinungsjahr:	2014

Acknowledgements

I would like to avail myself of the opportunity of thanking a few people who helped me in the preparation of this thesis and supported me during the last years.

First, I would like to thank Prof. Dr. Norbert Wermes for giving me the opportunity to take part in the work of the ATLAS collaboration during these exciting times and to spend two years at CERN. He provided a lot of support throughout the last years, in particular during the continuous strategic discussions concerning the $H \to b\bar{b}$ analysis in the HSG5 group.

Furthermore I would like to thank Dr. Eckhard von Toerne, who supervised me during my diploma thesis and tought me a lot in my first year in high energy physics. Even though he was not personally involved with the $H \to b\bar{b}$ analysis, he constantly provided advice and guidance.

It is a pleasure to thank my colleagues Dr. Goetz Gaycken, Stephan Hageboeck and Dr. Vadim Kostyukhin who successfully strove to establish this demanding analysis with me - all of them have contributed outstanding ideas and worked hard in the last two years. It is only fair to single out Dr. Goetz Gaycken, who not only implemented a fabuluous data analysis framework for this project and proof-read most chapters of this thesis, but also held the group at CERN together by organizing various sparetime activities, most prominently the famous movie-nights at “the palace”.

I would also like to thank the members of the TMVA developer team, Dr. Andreas Hoecker, Dr. Peter Speckmayer, Dr. Joerg Stelzer, Dr. Eckard von Toerne and Dr. Helge Voss. I benefitted a lot from being part of this interesting project and I enjoyed our weekly discussion over lunch during my time at CERN.

Many members of the Bonn group have become friends over the years and we have enjoyed countless group activities - be it hiking, skiing, cooking, wine tasting or movie nights. I would like to thank all of them for making my time in Bonn and Geneva a very enjoyable one.

I would also like to thank my long-term friends who accompanied me through the years and who reminded me from time to time that there are other things in life than physics.

I’m particular endebted to my parents, who supported me without condition, encouraged me with all my decisions, taught me so much in all the years and showed me that a livable society can only be created if every person assumes responsibility.

Last but not least I would like to thank my partner Eva who has shared joy and grief with me for many years now - I’m looking forward to all the adventures that still lie ahead of us and I’m happy to have her by my side.

Contents

"For if each star is little more a mathematical point [...], then all the stars taken together, tho' innumerable, must [...] in turn represent some single gigantick equation [...], to us unreadable, incalculable. A lonely, uncompensated, perhaps even impossible task - yet some of us must ever be seeking, I suppose."

THOMAS PYNCHON, Mason & Dixon

CHAPTER 1

Introduction

The quest to understand the world we are living in is probably as old as mankind itself. A central part of this quest has always been the search for the smallest building blocks of matter. In trying to identify those elementary building blocks, mankind has come a long way: from the ancient idea of the four primal elements being fire, water, earth and air to what is now called the Standard Model of elementary particle physics. But the most fundamental changes in our picture of matter did only emerge recently, in the 20th century. Not only was the theory of atoms finally confirmed - matter being composed of building blocks which would not allow further decomposition without loss of the characteristical chemical properties of the respective element - but a much more revolutionary insight was gained: entities at the scale of atoms or below behave fundamentally different from what commonplace physical experience suggests. Quantum physics dared to put chance right next to determinism. Greeted with a lot of skepticism at first, quantum physics and in particular quantum field theory is now accepted as the most precise description of nature ever formulated and countless experiments have verified its claims to tremendous precision. The atoms, once believed to be impartible, have long lost their roll to what we call elementary particles today, but quantum field theory stands firmly and forms the theoretical basis on which our current understanding of the physical world is built. This understanding is captured mathematically in the Standard Model of particle physics and will be formally introduced in chapter 2.

The Standard Model has succeeded in answering countless questions about the structure of our universe, but with every question answered, new questions have arisen and some of them remain unanswered yet. One of these long-standing questions is about the origin of mass: while the Standard Model predicts that all constituents of matter should be massless, simple observation proves that they are not. A possible solution to this dilemma known as the Higgs mechanism was already put forward by a group of theorists including Peter Higgs in the 1960s. A central claim of this elegant theory is the extension of the Standard Model by an additional particle which is referred to as the Higgs boson. Proving or excluding the existence of the Higgs boson has been a major goal for experiments in high energy physics ever since. The discovery of a new particle with a mass of 126GeV at the Large Hadron Collider in 2012 was hailed as a major breakthrough in this matter, because this particle could indeed be the sought-after Higgs boson and its discovery could thus finally confirm the theory formulated almost 50 years ago. Identifying and comparing the properties of the recently discovered particle to the theoretical predictions is therefore a task of utmost importance in the course of probing the Higgs mechanism.

A particularly striking property of the Higgs boson is its highly mass-dependent decay spectrum which comprises both fermionic and bosonic decay modes. Since the discovery of 2012 was entirely driven by observations in bosonic channels, the focus is now on the technically more involved detection of fermionic decays. In my thesis I will discuss the search for a Higgs boson decaying to a pair of b-quarks, which is expected to be the most abundant decay at a Higgs boson mass of 126GeV. The thesis is organized as follows: as mentioned above, I will review the basics of the Standard Model and the Higgs mechanism in chapter 2. Chapter 3 discusses some relevant properties of the Higgs boson in more detail and reviews important results obtained in former Higgs searches. It also serves as an additional motivation for the investigation of the $VH, H \to b\bar{b}$ channel. The LHC and the ATLAS experiment which delivered the data employed for this analysis are described in chapter 4, while essential aspects of event simulation for hadron collider physics are discussed in chapter 5. A detailed account of the analysis techniques and their application is given in chapters 7 and 8 and the results of the search are presented in chapter 9. The thesis concludes with a discussion of the results and an outlook on possible future improvements in chapters 10 and 11, respectively.

"Sweet is by convention, bitter by convention, hot by convention, cold by convention, color by convention; in truth there are but atoms and the void."

DEMOCRITUS

CHAPTER 2

The Standard Model of particle physics

The Standard Model (SM) of elementary particle physics provides the framework in which our current understanding about the fundamental building blocks of matter is formulated. Its mathematical structure is imposed by the gauge principle which will be introduced in Section 2.1. The particle content of the SM is listed in Table 2.1 and consists of fermions (particles with half-integral spin) and bosons (particles with integer spin).

Fermions					Bosons		
	I	II	III	charge		Name	Interaction
Leptons	ν_e	ν_μ	ν_τ	0	γ	photon	electromagnetic
	e	μ	τ	-1	$W^\pm, Z$	heavy gauge bosons	weak
Quarks	u	c	t	$+2/3$	g	gluon	strong
	d	s	b	$-1/3$	H	Higgs	

Table 2.1: The particle content of the Standard Model

The fermions can be subdivided further into leptons and quarks and are organized in three so-called families which have identical properties except for the particle masses. While the fermions constitute all known matter in our universe, the bosons act as the mediators of forces between them. But only the quarks take part in all interactions, leptons do not interact strongly and neutrinos only interact weakly. As dictated by the gauge principle, the interactions and thus also the bosons are closely linked to the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ of the Standard Model: the eight massless gluons, mediators of the strong force, correspond to the eight gauge fields associated with the generators of $SU(3)_C$, while the photon and the $W^\pm$ and Z bosons correspond to mixtures of the gauge fields associated with $SU(2)_L \otimes U(1)_Y$, the symmetry group of the electroweak interaction. The fact that the $W^\pm$ and Z bosons are massive while the photon is massless can be explained by the fact that the $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken. This mechanism also gives rise to the existence of the Higgs boson and will be discussed in detail in Section 2.3

2.1 The gauge principle

The Standard Model is a gauge theory, which means that the Lagrangian of the Standard Model is invariant under local gauge transformations. The emphasis is on "local" here, since the requirement that the invariance should not only remain intact for global transformations but also for transformations which depend on the space-time coordinates automatically gives rise to the presence of so-called gauge fields. These fields can be identified with the mediators of the physical forces which is to say that merely imposing a symmetry requirement on the Lagrangian is enough to define the dynamics of the system. Since this concept is essential for the formulation of the Standard Model it will be illustrated in the case of quantum electrodynamics (QED), where the allowed transformations are commutative and can be identified with the group $U(1)$. The symmetry group underlying the full Standard Model Lagrangian is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ and will be discussed in the following.

2.1.1 Gauge invariance in quantum electrodynamics

Let us begin our discussion of gauge invariance with the Dirac Lagrangian for a free fermion:

$$\mathcal{L}_0 = i\bar{\psi}(x)\gamma^\mu\partial_\mu\psi(x) - m\bar{\psi}(x)\psi(x) \tag{2.1}$$

It can easily be checked that $\mathcal{L}_0$ is invariant under global $U(1)$ transformations of the form

$$\psi(x) \rightarrow \psi'(x) = \exp\{-i\alpha\}\psi(x) \tag{2.2}$$

where $\alpha \in \mathbb{R}$ is a constant. This is to be expected since the phase of ψ does not bear any physical meaning. However $\mathcal{L}_0$ is *not* invariant if we promote the phase to a function of the space-time coordinate x, whereupon $\mathcal{L}_0$ transforms like

$$\mathcal{L}_0 \rightarrow \mathcal{L}_0' = \mathcal{L}_0 + \bar{\psi}(x)\gamma_\mu\psi(x)\partial^\mu\alpha(x) \tag{2.3}$$

The gauge principle now states that the invariance of the Lagrangian should also be intact under *local* phase transformations, implying that the physics should not depend on the local phase convention adopted. This makes sense if one considers that any instantaneous propagation of global phase information would violate the principle of causality. To restore invariance under local phase transformations in Eqn. 2.1, one has to introduce a gauge field A_μ which cancels the extra term in the Lagrangian when transformed. In fact, this requirement is sufficient to determine the behavior of the field under the transformation completely:

$$A_\mu(x) \rightarrow A_\mu'(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\alpha(x) \tag{2.4}$$

Defining the covariant derivative

$$D_\mu\psi(x) \equiv [\partial_\mu + ieA_\mu(x)]\psi(x) \tag{2.5}$$

one now obtains a modified version of the Lagrangian which is indeed invariant under local $U(1)$ transformations:

$$\mathcal{L} = \mathcal{L}_0 - eA_\mu(x)\bar{\psi}(x)\gamma^\mu\psi(x) = \bar{\psi}(x)[i\gamma^\mu D_\mu - m]\psi(x) \tag{2.6}$$

The new term with respect to $\mathcal{L}_0$ represents the interaction between the Dirac fermion and the gauge field and has arisen naturally from the requirement of local gauge invariance. It can be identified with the familiar electron-photon vertex of QED. Since the free-field Lagrangian

$$\mathcal{L}_A = -\frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{2.7}$$

for the gauge field A_μ shares the same invariance properties, one can simply add it to the modified Dirac Lagrangian to obtain the complete Lagrangian of QED

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(x)[i\gamma^\mu D_\mu - m]\psi(x) - \frac{1}{4}F_{\mu\nu}(x)F^{\mu\nu}(x) \tag{2.8}$$

from which the Maxwell equations can be deduced easily. Note that the derivation of this Lagrangian was entirely guided by the gauge principle and that the invariance requirement prohibits the addition of a mass term for the gauge field, predicting that the photon should be massless, just in agreement with experimental observation. The power of the gauge principle is now evident: requiring that the phase of the free fermion wave function should be irrelevant when considering any physical observables and then taking into account causality constraints which forced us to consider not only global but also local phase transformation, we were automatically led to a complete description of QED.

2.1.2 Non-Abelian gauge symmetry - quantum chromodynamics

The application of the gauge principle to quantum chromodynamics (QCD), the theory of the strong interaction, works in exactly the same way as demonstrated for quantum electrodynamics - except that the mathematics involved is more complicated due to the non-commutative nature of the underlying symmetry group $SU(3)_C$. The invariance underlying the strong interaction is not just simple phase invariance but invariance with respect to rotations in color space. Mathematically, this can be described by the group $SU(3)_C$, where the subscript C refers to “color”. Adopting a convenient vector notation for objects $q_f^T \equiv (q_f^1, q_f^2, q_f^3)$ in the color space with q_f^α representing a quark field of color α and flavor f, the resulting Lagrangian is given by

$$\begin{aligned}\mathcal{L}_{\text{QCD}} = &-\frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a) + \sum_f \bar{q}_f^\alpha (i\gamma^\mu \partial_\mu - m_f) q_f^\alpha \\ &- g_s G_a^\mu \sum_f \bar{q}_f^\alpha \gamma_\mu \left(\frac{\lambda^a}{2}\right)_{\alpha\beta} q_f^\beta \\ &+ \frac{g_s}{2} f^{abc}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G_c^\nu G_\mu^d G_\nu^e\end{aligned} \tag{2.9}$$

where λ^a, $a = 1, \ldots, 8$, denotes the generators of $SU(3)_C$ satisfying the commutation relation

$$\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = i f^{abc}\frac{\lambda^c}{2} \tag{2.10}$$

and f^{abc} are the structure constants of $SU(3)_C$. The gauge fields G_a^μ can be identified with the gluons of which there are eight. While it may be instructive to consult a textbook or an introductory article (see for example [1]) for details on the derivation of the QCD Lagrangian, one can already learn a lot about the nature of the strong interaction from a careful observation of the resulting terms:

1. The interaction strength is described by a single universal coupling constant g_s - the so-called strong coupling constant.

2. Quarks can emit gluons, the corresponding vertex is described by the second line of Eqn. 2.9.
3. The terms in the last line of Eqn. 2.9 which arise due to the non-Abelian nature of $SU(3)_C$ describe self-interactions between the gluons - a feature that was not present in QED.
4. Just as in QED, a mass term $\frac{1}{2}m_G^2 G_a^\mu G_\mu^a$ for the gauge fields is prohibited by gauge invariance. Gluons are massless.

Again, the gauge principle led to a complete description of the dynamics and thus to profound insights into the phenomenology of the strong interaction. Before we move on to the discussion of the theory of electroweak interactions, which is of specific interest in the context of this thesis, let us raise an important question: if the gauge fields are generally required to be massless, how can the heavy vector bosons of the Standard Model acquire their mass? Clearly there must be an additional theoretical concept which allows to preserve symmetry in the presence of massive fields. We shall address this question in Section 2.3.

2.2 The theory of electroweak interactions

Weak interactions are fundamentally different from the phenomena discussed so far. Unlike for QED and QCD, we cannot simply follow the Gauge principle to derive a Lagrangian for the theory of weak interactions since experimental observation reveals several features that call for a more elaborate treatment. Among these observations are:

- The force-mediators associated with the weak interaction are massive, contrary to the observation that only massless gauge-fields preserve symmetry.
- Only left-handed fermions and right-handed anti-fermions take part in charged current weak interactions. Parity is thus maximally violated.

Ignoring the first point for the moment, and concentrating on the second instead, one finds that a fermionic current which involves only left-handed states ψ_L can be written as

$$\bar{\psi}_L \gamma^\mu \psi_L = \frac{1}{2}\bar{\psi}\gamma^\mu(1-\gamma_5)\psi \tag{2.11}$$

which reveals the $V-A$ (vector – axial vector) structure of the charged weak current. Since the charged current interactions raise or lower the charge by one unit, the left-handed fermions taking part in these interactions should appear in doublets with a charge difference of one while the right-handed fermions should appear in singlets:

$$\begin{pmatrix} q^u \\ q^d \end{pmatrix}_L, \quad \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \quad l_R^-,\ q_R^u,\ q_R^d \tag{2.12}$$

2.2.1 Unification and the symmetry group of electroweak theory

With $SU(2)_L$ (the subscript L referring to the left-handed nature of the interaction) being the simplest group with a doublet representation, one might anticipate that this is indeed the symmetry group associated with the weak interactions and that the charged currents and the neutral current should form a weak isospin triplet when combined. The charges T^i associated with these currents should then generate the $SU(2)_L$ algebra:

$$[T^i, T^j] = i\epsilon_{ijk}T^k \tag{2.13}$$

However, measurements have proven that the weak neutral current j_μ^{NC} has a small but non-vanishing right-handed component and thus does not respect the $SU(2)_L$ symmetry. This problem can be resolved if one incorporates the electromagnetic current j_μ^{em} into the picture: it provides both left- and right-handed components and can be combined with the weak neutral current j_μ^{NC} to form two orthogonal combinations with well-defined transformation properties under $SU(2)_L$. This way, one obtains j_μ^3 and j_μ^Y, the former completing the weak isospin triplet and the latter being a singlet under $SU(2)_L$. The charge Y associated with j_μ^Y now generates its own symmetry group $U(1)_Y$ and is called the weak hypercharge. It is defined by

$$Q = T^3 + \frac{Y}{2} \tag{2.14}$$

where Q is the electromagnetic charge and T^3 is the z-component of the weak isospin. Due to this composite nature of weak and electromagnetic interactions, we may refer to both of them as *electroweak* interactions. The corresponding symmetry group is

$$G_{EW} = SU(2)_L \otimes U(1)_Y \tag{2.15}$$

and one may now impose the familiar gauge principle to obtain the complete Lagrangian. The calculation will not be outlined here, but one should note that four gauge fields $W_\mu^1, W_\mu^2, W_\mu^3, B_\mu$ have to be introduced. The interaction of these gauge fields with the fermions is then described by the Lagrangian

$$\mathcal{L}_{Int} = -g\bar{\psi}_1\gamma^\mu\frac{\sigma^i}{2}W_\mu^i\psi_1 - g'B_\mu\sum_{j=1}^{3}Y_j\bar{\psi}_j\gamma^\mu\psi_j \tag{2.16}$$

where we have introduced a shorthand notation with ψ_1 representing a left-handed fermion doublet and ψ_2, ψ_3 representing up-/down-type singlets. The $SU(2)_L$ coupling strength is given by g and the coupling strength associated with Y is given by g'. Expanding the terms in Eqn. 2.16 one obtains a description of the charged current interactions involving the fields

$$W_\mu \equiv (W_\mu^1 + iW_\mu^2)/\sqrt{2}, \quad W_\mu^\dagger \equiv (W_\mu^1 - iW_\mu^2)/\sqrt{2} \tag{2.17}$$

plus additional terms containing interactions with the neutral fields W_μ^3 and B_μ. While the former two can directly be identified with the $W^\pm$ bosons, the latter two are mixtures of Z and γ as was already pointed out in the discussion of the electroweak symmetry group. Making this mixing explicit by writing

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \equiv \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} \tag{2.18}$$

we obtain the Lagrangian describing neutral current interactions:

$$\mathcal{L}_{Int}^{NC} = -\sum_j \bar{\psi}_j\gamma^\mu\left\{A_\mu\left[g\frac{\sigma_3}{2}\sin\theta_W + g'y_j\cos\theta_W\right] + Z_\mu\left[g\frac{\sigma_3}{2}\cos\theta_W - g'y_j\sin\theta_W\right]\right\}\psi_j \tag{2.19}$$

Using Eqn. 2.14 one can immediately read of the condition under which the familiar QED Lagrangian can be restored from the first term:

$$g \sin\theta_W = g' \cos\theta_W = e \tag{2.20}$$

The weak mixing angle θ_W is thus defined by the ratio of the group coupling constants of $U(1)_Y$ and $SU(2)_L$

$$\tan\theta_W = \frac{g'}{g} \tag{2.21}$$

and the neutral current Lagrangian can finally be decomposed:

$$\begin{aligned}
\mathcal{L}_{Int}^{NC} &= \mathcal{L}_{Int}^{QED} + \mathcal{L}_{Int}^{Z} \\
\mathcal{L}_{Int}^{QED} &= -eA_\mu j_{em}^\mu = -eA^\mu (j_\mu^3 + \frac{1}{2} j_\mu^Y) \\
\mathcal{L}_{Int}^{Z} &= -\frac{e}{2\sin\theta_W \cos\theta_W} Z^\mu (j_\mu^3 - \sin^2\theta_W j_\mu^{em})
\end{aligned} \tag{2.22}$$

In addition to the terms describing the interactions of the electroweak gauge fields with the fermions, a multitude of cubic and quartic terms which represent self-interactions among the gauge fields is obtained. We will not discuss those in detail but only note that each of the terms contains at least one pair of W bosons which means that a neutral vertex containing only Z bosons or photons is not permitted.

At this point one might be tempted to say that careful reasoning about experimental observations, symmetry groups and the gauge principle has finally led to a complete description of the electroweak interactions, but we still have to address the severe problem mentioned at the beginning of this section: the gauge bosons of the electroweak interaction are massive, despite the fact that an explicit mass term is not permitted if the gauge symmetry is to be preserved. The following discussion of spontaneous symmetry breaking will not only help to solve this problem, but will also set the stage for the introduction of the Higgs boson, which is the central topic of this thesis.

2.3 Spontaneous symmetry breaking

The concept of spontaneous symmetry breaking (SSB) provides an elegant explanation for the origin of the gauge boson masses in electroweak theory. For a physical system to exhibit SSB the associated Lagrangian must be invariant under a given group of transformations G and at the same time allow for a degenerate set of minimal-energy states which form a G-multiplet. The selection of one of these states as the ground state of the system means that the symmetry is spontaneously broken.

2.3.1 The Goldstone theorem

The degrees of freedom that are associated with transitions between the degenerate states of minimal energy along flat directions of the potential can be identified with additional massless excitations of the system. To see this, consider the Lagrangian

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - V(\phi), \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{2.23}$$

where ϕ is a complex scalar field. Assuming that the potential $V(\phi)$ is bounded ($\lambda > 0$), two distinct situations are possible:

1. $\mu^2 > 0$: This describes a massive scalar particle with mass μ. The potential has a single minimum at $\phi = 0$, see Figure 2.1a.

2. $\mu^2 < 0$: All configurations satisfying

$$|\phi_0| = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} > 0 \tag{2.24}$$

are degenerate states of minimal energy, see Figure 2.1b.

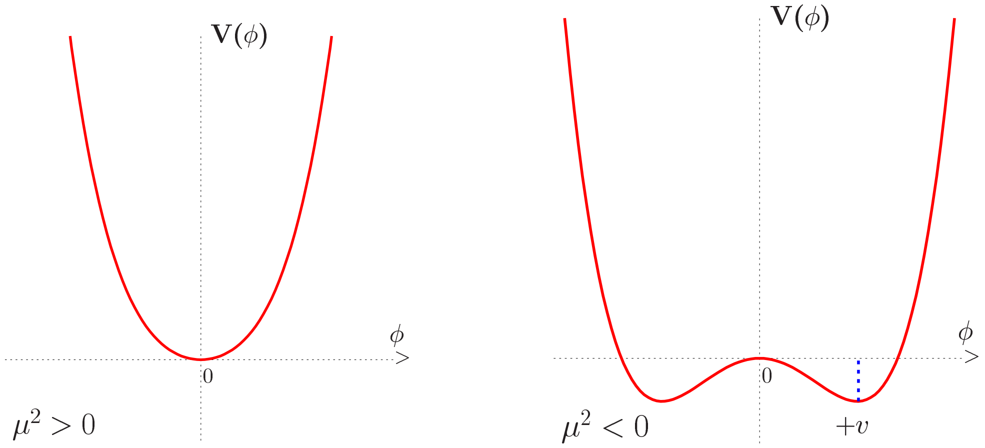


Figure 2.1: The scalar potential $V(\phi)$ for $\mu^2 > 0$ (left, only one minimum) and $\mu^2 < 0$ (right, infinitely many degenerate minima) [2].

The second case deserves more attention: assuming that the Lagrangian is invariant under global phase transformations

$$\phi(x) \to \phi' = \exp\{i\theta\}\phi(x) \tag{2.25}$$

there exist infinitely many states of minimal energy. Promoting any particular choice of θ to the ground state of the system will break the symmetry. Once a ground state is chosen, it makes sense to parametrize the field in a way that is suitable to describe small perturbations of this state. Choosing $\theta = 0$ such a parametrization can be written as

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \phi_1(x) + i\phi_2(x)) \tag{2.26}$$

with real fields ϕ_1 and ϕ_2. The potential now takes the form:

$$V(\phi) = V(\phi_0) - \mu^2\phi_1^2 + \lambda v\phi_1(\phi_1^2 + \phi_2^2) + \frac{\lambda}{4}(\phi_1^2 + \phi_2^2)^2 \tag{2.27}$$

While ϕ_1 has a mass of $-2\mu^2$, there is no mass term for ϕ_2. The massless particle ϕ_2 can be identified with excitations along a flat direction in the potential which connects states with the same minimal energy. This is schematically shown in Figure 2.1b. The appearance of such massless states due to SSB is a general result known as the Goldstone theorem [3].

2.4 The Higgs mechanism

The Higgs mechanism [4], [5], [6], [7] (also referred to as the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism due to the fact that other researchers proposed it independently of Peter Higgs at about the same time) combines the ideas of SSB and local gauge symmetry to finally achieve what has been missing from the theory outlined so far: the generation of masses for the gauge bosons associated with weak interactions. To see this, let us introduce a scalar $SU(2)_L$ doublet of complex fields

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.28}$$

along with a Lagrangian that features the covariant derivative associated with the $SU(2)_L \otimes U(1)_Y$ symmetry of the electroweak interaction and a potential like the one described in Section 2.3.1:

$$\begin{aligned} \mathcal{L}_S &= D_\mu \phi^\dagger D^\mu \phi - V(\phi), \quad V(\Phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \\ D_\mu &= \partial_\mu + ig \frac{\sigma_i}{2} W^i_\mu + i \frac{g'}{2} Y B_\mu \end{aligned} \tag{2.29}$$

As discussed before we can pick a vacuum expectation value for ϕ:

$$< \phi >_0 = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \quad v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{2.30}$$

While a particular choice of the ground state breaks the $SU(2)_L \otimes U(1)_Y$ symmetry, it still preserves a symmetry under $U(1)_{em}$ as can be easily seen from

$$Q < \phi >_0 = (T_3 + \frac{Y}{2}) < \phi >_0 = 0 \Rightarrow \exp\{i\alpha Q\} < \phi >_0 = < \phi >_0 \tag{2.31}$$

This guarantees that the electric charge is conserved in our model, in agreement with experimental observation. Moreover it allows the corresponding gauge boson, the photon, to remain massless. With foresight, the scalar doublet can also be parametrized by four real fields $\theta^1, \theta^2, \theta^3, H$, where the θ^i correspond to the three generators that are spontaneously broken by the choice of the ground state:

$$\phi(x) = \exp\{i \frac{\sigma_i}{2} \theta^i(x)\} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.32}$$

Since the original Lagrangian is still invariant under $SU(2)_L$, one can now apply an appropriate transformation to eliminate any dependence on θ^i. This is known as the unitary gauge. In this gauge, the scalar Lagrangian 2.29 takes the form:

$$\mathcal{L}_S = \left| \left(\partial_\mu + ig \frac{\sigma_i}{2} W^i_\mu + i \frac{g'}{2} Y B_\mu \right) \frac{v + H}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right|^2 - \mu^2 \frac{(v+H)^2}{2} - \lambda \frac{(v+H)^4}{4} \tag{2.33}$$

which in terms of the $W^\pm$ and Z^0 fields introduced in Section 2.2.1 can be written as:

$$\mathcal{L}_S = \frac{1}{2} \partial_\mu H \partial^\mu H + \frac{g^2}{4} (v + H)^2 \left(W^+_\mu W^{-\mu} + \frac{1}{\cos^2 \theta_W} Z_\mu Z^\mu \right) - \mu^2 \frac{(v+H)^2}{2} - \lambda \frac{(v+H)^4}{4} \tag{2.34}$$

Something interesting has happened: quadratic terms in the vector fields have appeared, rendering the $W^\pm$ and Z^0 bosons massive:

$$m_W = \frac{gv}{2}, \quad m_Z = \frac{gv}{2\cos\theta_W} = \frac{m_W}{\cos\theta_W} \tag{2.35}$$

That means one can generate masses for the gauge bosons of the weak interaction by adding $\mathcal{L}_S$ to the $SU(2)_L \otimes U(1)_Y$ model discussed in Section 2.2.1 without breaking the symmetry of the total Lagrangian. The key to this seemingly contradictory result is the occurrence of spontaneous symmetry breaking: three of the four generators of $SU(2)_L \otimes U(1)_Y$ are broken by the choice of a particular ground state for ϕ. The resulting massless Goldstone bosons can in turn be eliminated from the Lagrangian by choosing an appropriate gauge, as a result of which the $W^\pm$ and Z^0 acquire mass. The total number of degrees of freedom remains of course unchanged: the original massless $W^\pm$ and Z^0 bosons have two possible polarizations each, accounting for 6 d.o.f. in total, plus four real scalar fields with one d.o.f. each. Through SSB three gauge bosons become massive and acquire an additional d.o.f. for the longitudinal polarization, while three of the four scalar fields vanish. The remaining scalar field H is associated with a new particle: the Higgs boson, named after the British physicist Peter Higgs. Although Higgs was not the only one to propose the specific mechanism for symmetry breaking introduced here, he was, according to *Nature* [8],

> *"the first to postulate the existence of a massive particle",*

which now bears his name.

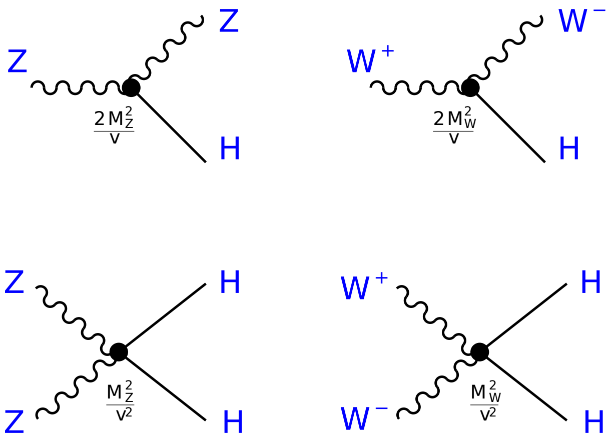


Figure 2.2: The interactions of the Higgs boson with the heavy gauge bosons. Note that the couplings are proportional to the squared boson masses.

Let us investigate the properties of the Higgs boson a bit more by using Eqns. 2.30 and 2.35 to rewrite the Lagrangian 2.29 in terms of the gauge boson masses:

$$\begin{aligned}
\mathcal{L}_S &= \frac{1}{4}\lambda v^4 + \mathcal{L}_H + \mathcal{L}_{HVV} \\
\mathcal{L}_H &= \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}(-2\mu^2)H^2 - \frac{(-2\mu^2)}{2v}H^3 - \frac{(-2\mu^2)}{8v^2}H^4 \\
\mathcal{L}_{HVV} &= m_W^2 W_\mu^+ W^{-\mu}\left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right) + \frac{1}{2}m_Z^2 Z_\mu Z^\mu \left(1 + \frac{2}{v}H + \frac{H^2}{v^2}\right)
\end{aligned} \tag{2.36}$$

We can now identify the mass of the Higgs boson as

$$m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v \tag{2.37}$$

and see that the interactions between the gauge bosons and the Higgs boson bear a very characteristic form in the sense that the couplings are proportional to the squared boson masses. The fact that the couplings are entirely determined by v, m_H, m_W and m_Z allows very precise predictions about Higgs production and decay (see Section 3.2). The interaction vertices and couplings are depicted in Figure 2.2.

2.4.1 Generation of fermion masses

With the scalar Higgs doublet at hand, one can also introduce additional Yukawa-type couplings of the Higgs field to fermions:

$$\mathcal{L}_Y = -c_1(\bar{u},\bar{d})_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_R - c_2(\bar{u},\bar{d})_L \begin{pmatrix} \phi^0 \\ -\phi^- \end{pmatrix} - c_3(\bar{\nu}_e,\bar{e}) \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R + \text{h.c.} \tag{2.38}$$

Writing $\mathcal{L}_Y$ in unitary gauge reveals that this produces mass terms for the fermions after SSB:

$$\begin{aligned} \mathcal{L}_Y &= -\frac{1}{\sqrt{2}}(v+H)\left(c_1\bar{d}d + c_2\bar{u}u + c_3\bar{e}e\right) \\ m_d &= c_1\frac{v}{\sqrt{2}}, \qquad m_u = c_2\frac{v}{\sqrt{2}}, \qquad m_e = c_3\frac{v}{\sqrt{2}} \end{aligned} \tag{2.39}$$

Since the parameters c_i are unknown, the fermion masses are not fixed by theory. Measuring them however immediately determines the couplings of the Higgs boson to fermions which can be expressed in terms of the masses:

$$\mathcal{L}_Y = -(1+\frac{H}{v})\left(m_d\bar{d}d + m_u\bar{u}u + m_e\bar{e}e\right) \tag{2.40}$$

We have seen that the Higgs mechanism provides an elegant solution to the problem of symmetry conservation in the presence of massive gauge bosons and can also account for massive fermions, in agreement with experimental observation. But even far beyond mass generation, the Higgs mechanism plays a crucial role for the consistency of the Standard Model as will be discussed in Chapter 3.1. It is thus not surprising that considerable experimental efforts have been made to verify or disprove the proposal ever since it was put forward in the 1960s. The focus of this work has always been the search for the Higgs boson, which is the most important fingerprint of the theory. The recent discovery of a Higgs-like boson at the LHC [9],[10],[11],[12] (see also Section 3.3) was thus met with great interest in the scientific community and further studies are currently being undertaken to investigate whether this boson is indeed the sought-after Higgs boson. A more detailed description of the properties of the Higgs boson and the most important experimental results related to the search for it can be found in Chapter 3.

"For more than three decades, the Higgs has been physicists' version of King Arthur's Holy Grail, Ponce de Leon's Fountain of Youth, Captain Ahab's Moby Dick. It's been an obsession, a fixation, an addiction to an idea that almost every expert believed just had to be true."

SCIENCE NEWS 2012

CHAPTER 3

Higgs physics

The Higgs mechanism introduced in Section 2.4 provides an elegant way to introduce spontaneous symmetry breaking to the theory of electroweak interactions and thus generate mass terms for the gauge bosons without giving up on the gauge invariance of the full Lagrangian. No statement is made however about the mass of the Higgs boson - the parameter that determines most of its properties and is therefore of great importance in any search for the Higgs boson. On closer examination though, several theoretical considerations lead to constraints on the Higgs boson mass. I will review some of these arguments in Section 3.1. In addition, a multitude of even stronger constraints on the Higgs mass has been derived from higher order corrections to electroweak precision measurements, let alone the direct Higgs searches at LEP and the Tevatron. A detailed summary of these measurements is clearly beyond the scope of this thesis, but I may refer to some of the results when it is appropriate in the discussion of the search presented here. No up-to-date account of Higgs boson physics would be complete however without a short presentation of the discovery of a Higgs boson at the LHC in 2012. I will address this exciting matter in Section 3.3.

3.1 Theoretical constraints

In this section I will briefly review some theoretical considerations known under the terms "unitarity", "perturbativity", "triviality" and "stability". All of them impose constraints on the a-priori unknown Higgs boson mass. Moreover, some constraints can also be interpreted in the context of new physical phenomena which would be required to save the predictive power of the Standard Model in the absence of a Higgs boson. The discussion follows the presentation in [2].

3.1.1 Unitarity and perturbativity

Consider the scattering of W bosons $W^+W^- \to W^+W^-$ at high energies ($s \gg m_W^2$). In this regime the main contribution to the amplitude for this process comes from longitudinal W bosons and can - owing to the electroweak equivalence theorem [13] - be related to the amplitude of the associated Goldstone bosons w:

$$A(W_L^+W_L^- \to W_L^+W_L^-) = A(w^+w^- \to w^+w^-) + O(\frac{m_W^2}{s}) \tag{3.1}$$

The latter can be written as

$$A(w^+w^- \to w^+w^-) = -\left(\frac{2m_H^2}{v^2} + \left(\frac{m_H^2}{v}\right)^2 \frac{1}{s - m_H^2} + \left(\frac{m_H^2}{v}\right)^2 \frac{1}{t - m_H^2}\right) \tag{3.2}$$

where s, t are the Mandelstam variables and $v = 246\,\mathrm{GeV}$ is the vacuum expectation value of the Higgs field introduced in Section 2.4. To see that this expression yields a violation of unitarity for large m_H^2, one has to expand A in partial waves a_l of angular momentum l (given by Legendre polynomials) and exploit that the cross-section for a $2 \to 2$ process is related to this expansion via

$$\sigma = \frac{16\pi}{s} \sum_{l=0}^{\infty} (2l+1)|a_l|^2 \tag{3.3}$$

For the discussion of unitarity, it is sufficient to take a closer look at the first ($l = 0$) term of the expansion, which can be derived from Eqn. 3.2:

$$a_0 = \frac{1}{16\pi s} \int_s^0 dt|A| = -\frac{m_H^2}{16\pi v^2}\left(2 + \frac{m_H^2}{s - m_H^2} - \frac{m_H^2}{s} \log\left(1 + \frac{s}{m_H^2}\right)\right) \tag{3.4}$$

For energies $\sqrt{s} \gg m_H$, a_0 behaves like

$$a_0 \to -\frac{m_H^2}{8\pi v^2} \tag{3.5}$$

Owing to the optical theorem on the other hand, σ can be expressed as

$$\sigma = \frac{1}{s}\Im(A(\theta = 0)) \tag{3.6}$$

which, using Eqn. 3.3, leads to the unitarity condition

$$|a_l|^2 = \Re(a_l)^2 + \Im(a_l)^2 = \Im(a_l) \quad \Rightarrow \quad |\Re(a_l)| \leq \frac{1}{2} \tag{3.7}$$

For a_0, this condition can only be fulfilled if

$$\frac{m_H^2}{4\pi v^2} \leq 1 \tag{3.8}$$

from which an upper bound $m_H \lesssim 870\,\mathrm{GeV}$ for the Higgs boson mass is obtained. Including other vector boson scattering channels, the bound can even be tightened to

$$m_H \lesssim 710\,\mathrm{GeV} \tag{3.9}$$

However, this result has to be taken with care since the Higgs self-coupling $\lambda = m_H^2/(2v^2)$ grows with the Higgs mass and thus leads to large radiative corrections which break perturbativity at some point, rendering the argument given above invalid.

Turning the argument around by going to the limit of very large Higgs masses $m_H \gg \sqrt{s}$ and applying the unitarity condition from Eqn. 3.7 also leads to a unitarity-violating high-energy behavior of a_0. This can be formulated as an upper limit for $\sqrt{s}$:

$$\sqrt{s} \lesssim 1.2\,\mathrm{TeV} \tag{3.10}$$

Since a very heavy Higgs boson is equivalent to a theory *without* a Higgs boson, this result can be interpreted in the following way: unitarity in vector-boson scattering processes must either be restored by the existence of a Higgs bosons with $m_H \lesssim 710\,\mathrm{GeV}$ or by the appearance of new physics contributions at a scale of $\sim 1\,\mathrm{TeV}$. The latter point is of great importance with respect to the experiments at the LHC: even if the Higgs boson would not have been discovered (see Section 3.3), the reachable energies would have guaranteed the observation of any new phenomena which could possibly restore unitarity.

The warning about large corrections arising from a strong self-coupling λ of the Higgs boson stated above can also be turned into a further constraint on the Higgs boson mass. Consider the partial decay width Γ_{WZ} of the Higgs boson into massive gauge bosons. One can include loop corrections involving the Higgs bosons by expanding the decay width in terms of λ [14]:

$$\Gamma_{WZ} \simeq \Gamma_{WZ}^{Born}\left(1 + 3\left(\frac{\lambda}{16\pi^2}\right) + 62\left(\frac{\lambda}{16\pi^2}\right)^2 + (O)(\lambda^3)\right) \tag{3.11}$$

Since $\lambda \propto m_H^2$, the higher order terms quickly surpass the size of the Born term for large m_H and perturbativity is lost. Taking into account other processes and criteria [15], the perturbativity argument yields an upper bound of $\sim 700\,\mathrm{GeV}$ for the Higgs mass, in good agreement with the unitarity argument already discussed.

3.1.2 Triviality and stability

A different bound for the Higgs boson mass comes from loop corrections to the quartic Higgs coupling λ. Like other couplings in the Standard Model, λ depends on the energy scale Q. Considering only loop corrections which involve the Higgs boson itself, the evolution of λ as a function of Q can be described by the renormalization group equation [16]

$$\frac{d\lambda(Q)}{dQ^2} = \frac{3}{4\pi^2}\lambda^2(Q^2) \tag{3.12}$$

Choosing $Q_0 = v$ this differential equation is solved by

$$\lambda(Q^2) = \lambda(v^2)\left(1 - \frac{3}{4\pi^2}\lambda(v^2)\log\frac{Q^2}{v^2}\right)^{-1} \tag{3.13}$$

which goes to infinity if the denominator goes to zero, that is at

$$\Lambda_C = v\exp\left(\frac{4\pi^2}{3\lambda}\right) = v\exp\left(\frac{8\pi^2 v^2}{3m_H^2}\right) \tag{3.14}$$

where the cut-off scale Λ_C is called the "Landau pole". One can interpret this result in the following way: assuming that new physical phenomena are not encountered below a given scale Λ_C, the Higgs mass must obey

$$m_H^2 < \frac{8\pi^2 v^2}{3\log\frac{\Lambda_C}{v}} \tag{3.15}$$

Setting Λ_C to the Higgs boson mass itself one obtains $m_H \lesssim 700\,\mathrm{GeV}$. This is in good agreement with a result obtained from a more rigorous treatment involving simulations on the lattice [17] which yields $m_H \lesssim 640\,\mathrm{GeV}$. If there is no new physics at Λ_C, the only way to avoid the Landau pole at all scales is to impose triviality ($\lambda = 0$) [18]. However, this would render the Higgs boson massless and prohibit spontaneous symmetry breaking - the triviality argument thus ensures that there is an upper bound to the energy domain where the Standard Model is valid.

Turning to a more complete description of the energy dependence of λ which also takes into account loop contributions involving top quarks and heavy gauge bosons, one can derive yet another constraint on the Higgs boson mass. For small values of the quartic coupling $\lambda \ll \lambda_t, g_1, g_2$, where $\lambda_t = m_t/v$ denotes the Yukawa coupling to the top quark and g_1, g_2 are the couplings to the weak bosons, one obtains an approximate solution to the regularization group equation for the evolution of λ [2]:

$$\lambda(Q^2) = \lambda(v^2) + \frac{1}{16\pi^2}\left(-12\frac{m_t^4}{v^4} + \frac{3}{16}(2g_2^4 + (g_2^2 + g_1^2)^2)\right)\log\frac{Q^2}{v^2} \tag{3.16}$$

If $\lambda(v^2)$ is too small, this expression becomes negative by virtue of the large negative contribution from the top quark. In order to avoid a scalar potential $V(Q^2) < V(v)$ which would destabilize the vacuum [19], the Higgs boson mass must thus be bounded from below:

$$m_H^2 > \frac{v^2}{8\pi^2}\left(-12\frac{m_t^4}{v^4} + \frac{3}{16}(2g_2^4 + (g_2^2 + g_1^2)^2)\right)\log\frac{Q^2}{v^2} \tag{3.17}$$

Again, the constraint depends on the choice of the cut-off scale Λ_C. For a low $\Lambda_C \sim 10^3$ GeV one obtains

$$m_H \gtrsim 70\,\text{GeV} \tag{3.18}$$

while a high $\Lambda_C \sim 10^{16}$ GeV yields

$$m_H \gtrsim 130\,\text{GeV} \tag{3.19}$$

A graphical representation of the boundaries obtained from triviality and vacuum stability is shown in Figure 3.1.

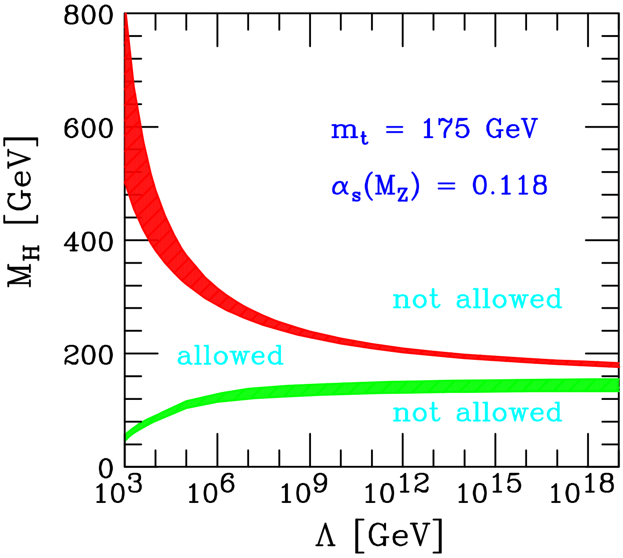


Figure 3.1: Constraints on the Higgs boson mass imposed by triviality and vacuum stability considerations. [2]

3.2 Higgs phenomenology

With its charge (0), spin (0), and parity (+) fixed by theory, an interesting property of the Standard Model Higgs boson is the fact that its behavior in terms of couplings, production mechanisms and decay channels is entirely determined by a single free parameter: its mass. The mass can be expressed in terms of the vacuum expectation value of the Higgs field v and the Higgs self-coupling parameter λ:

$$m_H = \sqrt{2\lambda}v \tag{3.20}$$

While v can be related to the Fermi coupling G_F via $v = (\sqrt{2}G_F)^{-1/2} \approx 246\,\text{GeV}$, which in turn was measured to high precision in muon decays [20], the self-coupling is unknown and thus no direct prediction for the Higgs mass is possible. One can however calculate the production rates and branching ratios of the Higgs boson as a function of m_H. The results provide important input to any search for the Higgs boson as they reveal the most promising decay channels in any given mass range and thus dictate what decay products to look for. I will shortly review some of these results in the following, focusing on the mass range $100\,\text{GeV} \lesssim m_H \lesssim 200\,\text{GeV}$. A more detailed discussion of the production mechanism/decay channel studied in this thesis is given in Section 3.4.

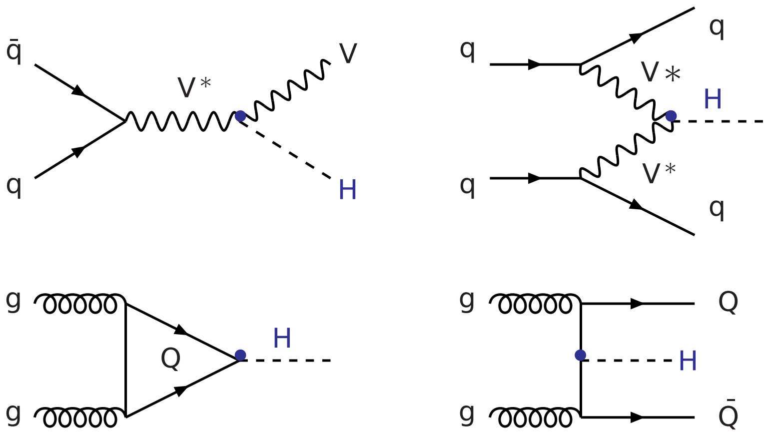
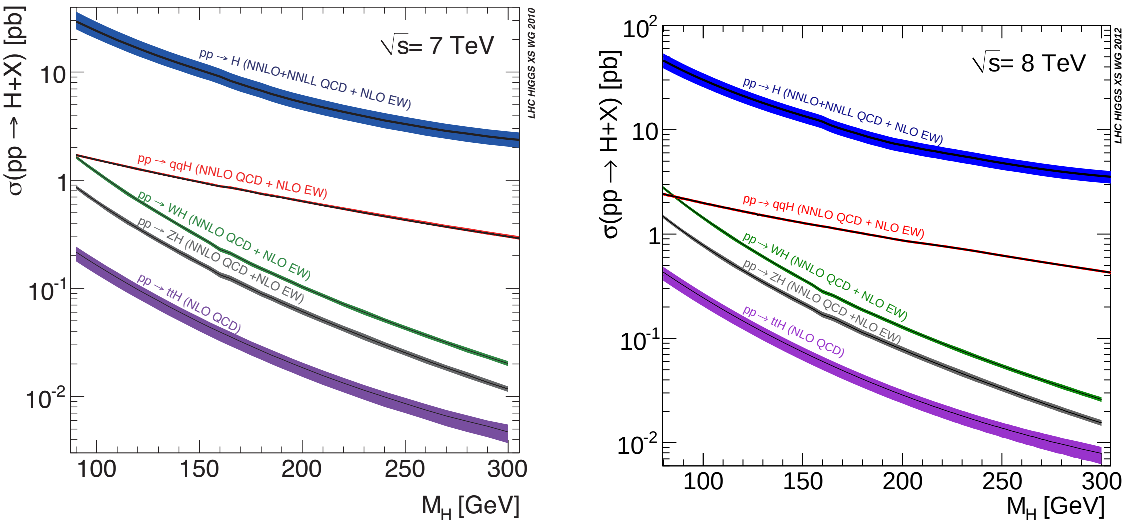


Figure 3.2: The Feynman diagrams of the four dominant Higgs production mechanisms at the LHC: gluon-gluon fusion (bottom left), vector boson fusion (top right), Higgs-strahlung (top left), associated production with $t\bar{t}$ (bottom right) [2].

3.2.1 Higgs production mechanisms at the LHC

The dominant Higgs production mechanism at the LHC is gluon-gluon fusion $gg \rightarrow H + X$. Although it can only be induced through a quark loop, it is dramatically enhanced by the fact that small parton momentum fractions x are being most often probed at the LHC's center of mass energy of $\sqrt{s} = 7\,\text{TeV}$ (2011) or $\sqrt{s} = 8\,\text{TeV}$ (2012), respectively. Since the Higgs couples to mass, the gluon-gluon fusion production process is dominated by top and bottom quark loops and has a cross-section which is an order of magnitude larger than that of vector boson fusion $qq \rightarrow qqH + X$, the second most abundant production process. Other important processes, in particular for a light Higgs with $m_H \lesssim 130\,\text{GeV}$, are the associated production $q\bar{q} \rightarrow V + H, V = W, Z$ with a W or a Z boson - often refered to as "Higgs-strahlung" - and the associated production $gg \rightarrow t\bar{t} + H$ with a $t\bar{t}$-pair. The Feynman diagrams of all four processes are displayed in Figure 3.2 while their cross-sections as a function of m_H are shown in Figure 3.3.



(a) (b)

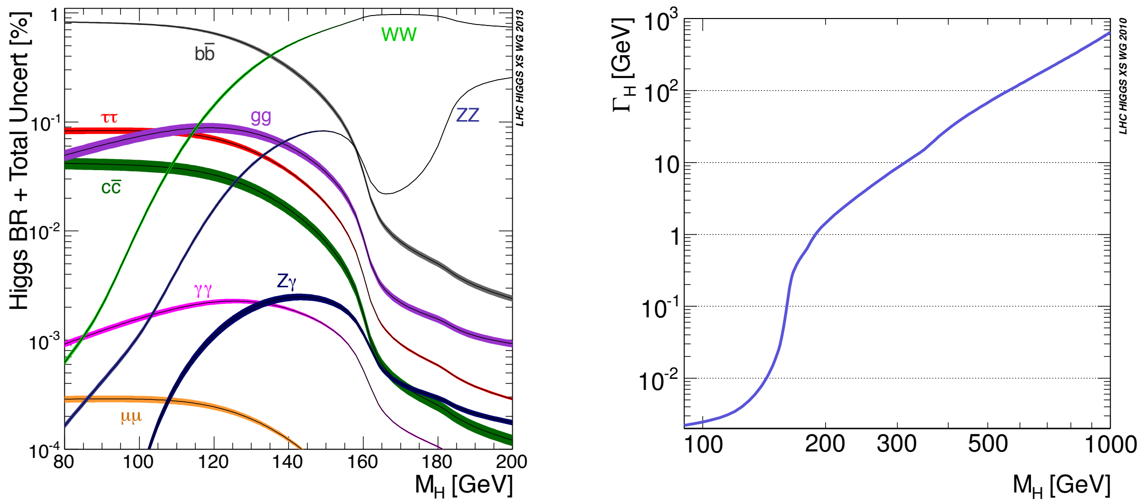
Figure 3.3: The cross-sections of the most important Higgs production mechanisms at the LHC as a function of m_H for center of mass energies of $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, respectively [21].

3.2.2 Decay channels of the Higgs boson

The short lifetime of the Higgs boson renders a direct observation impossible, which means that it can only be detected through its decay products. Since the Higgs boson couples to mass, it decays predominantly to the heaviest particles allowed by the disposable phase space - its partial branching ratios and its total width are thus highly mass dependent. As can be seen in Figure 3.4, the low mass region $100\,\text{GeV} \lesssim m_H \lesssim 130\,\text{GeV}$ is strongly dominated by the decay $H \to b\bar{b}$, followed by decays to $\tau^+\tau^-$, gg and $c\bar{c}$. The upper half of that mass range also has a significant contribution from $H \to WW^*$ which increases with m_H until it finally surpasses $H \to b\bar{b}$ at $\sim 130\,\text{GeV}$ to become the prevalent decay for higher Higgs masses. The uncertainties of the branching fractions in the low mass range which are shown in Figure 3.4a stem from the uncertainties associated with the quark masses and the strong coupling constant α_s. Note that the relative error on $\text{BR}(H \to b\bar{b})$ is small since the b-quark mass is known more precisely than the c-quark mass and its running is less dramatic.

3.3 First observation of a Higgs boson at the LHC in 2012

On July 4, 2012, both the ATLAS and the CMS collaboration announced the discovery of a new particle with a mass of $\sim 126\,\text{GeV}$ and decay modes indicating that this particle is a boson [9],[10],[11],[12]. The new particle was first observed in decays to pairs of photons and Z bosons with rates compatible to those predicted for the standard model Higgs boson. The invariant mass spectra of the decay products in these channels are shown in Figure 3.5 and the observed p-values (probabilities) for the background-only hypothesis are shown in Figure 3.6.



(a)

(b)

Figure 3.4: The decay branching fractions and the total decay width of the Higgs boson as a function of m_H [21].

The new particle was immediately suspected to be a[1] Higgs boson, although the experimental results still left room for other explanations at that point in time. In order to allow for more detailed studies, the data target for 2012 was increased from 15 fb^{-1} to 20 fb^{-1} and the projected first long shutdown of the LHC was postponed by seven weeks from October to December [22]. The first follow-up results were presented at the Hadron Collider Physics (HCP) Symposium 2012 in Kyoto. The discovery of a new particle was confirmed with great confidence and the inclusion of different decay channels showed no significant deviation from the properties expected for a Higgs boson [23],[24].

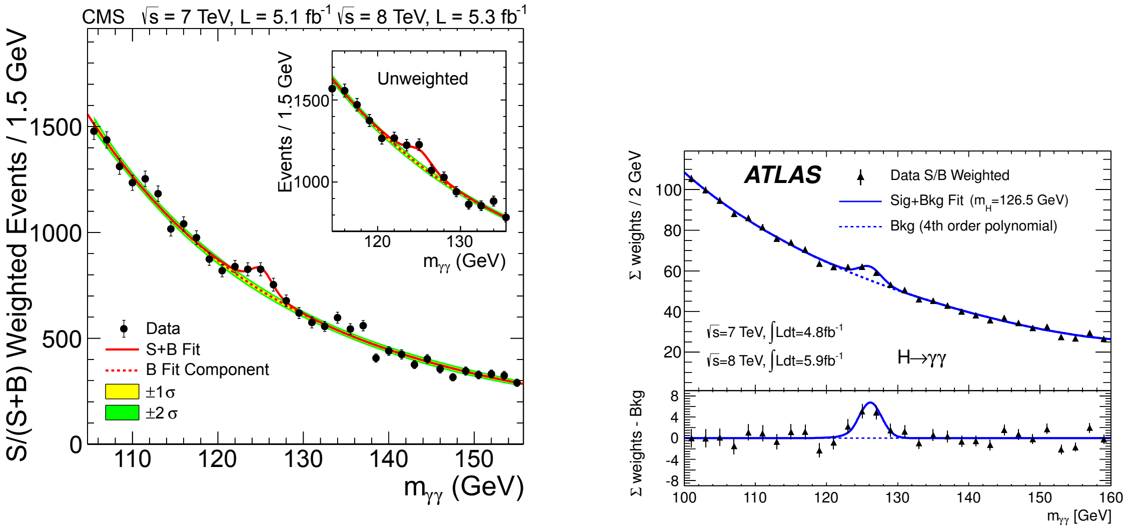
In the meantime, further studies of the particle's spin and parity have been published [25],[26]. They provide strong evidence for the assumption that the observed boson is indeed a Higgs boson as the data favor spin 0 and strongly prefer a pure scalar hypothesis (positive parity) over a pseudo-scalar (negative parity) one. The new boson is thus the first elementary scalar particle to be discovered in nature, a fact that underpins the importance of its discovery. In the words of CERN director general Rolf-Dieter Heuer:

> *"All the matter particles are spin-1/2 fermions. All the force carriers are spin-1 bosons. Higgs particles are spin-0 bosons (scalars). The Higgs is neither matter nor force. The Higgs is just different."*

At the time of this writing, careful analyses of all LHC data collected in 2011 and 2012 (a total of $\sim 25\,\mathrm{fb}^{-1}$) have firmly established observations of a Higgs boson in the decay channels to $\gamma\gamma$, WW and ZZ. The fermionic decay channels to $\tau\tau$ and $b\bar{b}$ on the other hand are more difficult to exploit and could not be confirmed yet. In particular the decay to a pair of $b-$quarks provides an important handle for testing the compatibility of the couplings of the new boson with the expectation for the Standard Model Higgs boson. This decay channel will thus be the subject matter of this thesis and some of its important properties will be discussed further in Section 3.4.

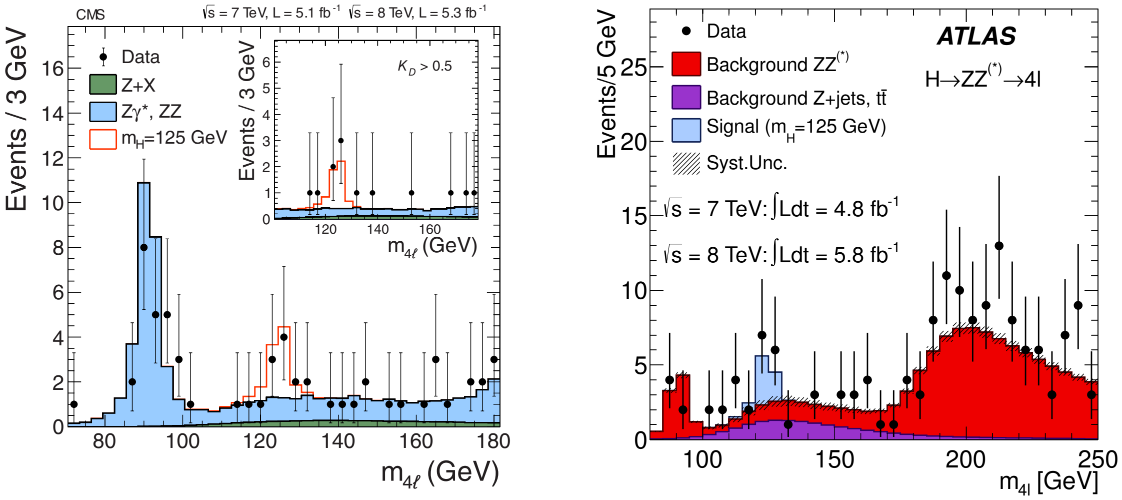
---

[1] In the discussion of the discovery, I deliberately use the indefinite article, i.e. *a* Higgs boson, to emphasize that there are other theories beyond the Standard Model which also predict the existence of one or more Higgs bosons and which cannot be ruled out at the current time. In this thesis however, I will only be concerned with the Standard Model Higgs boson.



(a) Invariant mass spectrum of Higgs candidates decaying to photon pairs in the $H \to \gamma\gamma$ analysis at CMS. The lines represent the fitted background and signal, and the colored bands represent the ±1 and ±2 standard deviation uncertainties in the background estimate. [12]

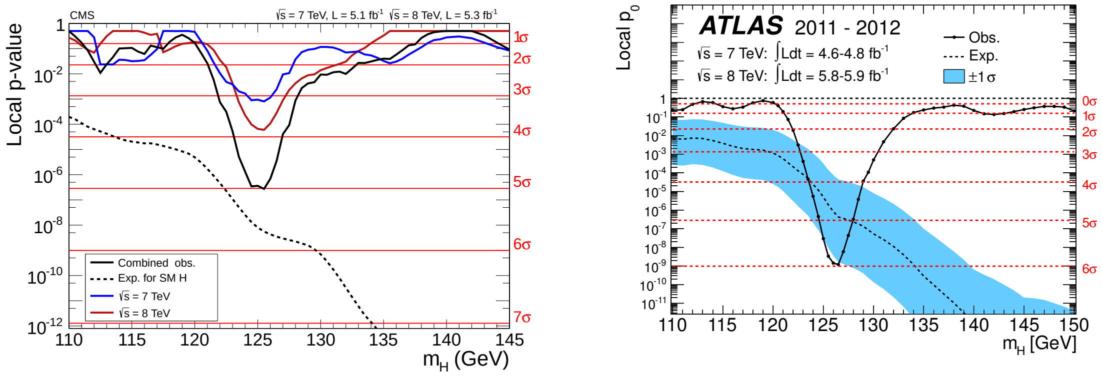
(b) Invariant mass spectrum of Higgs candidates decaying to photon pairs in the $H \to \gamma\gamma$ analysis at ATLAS with the result of a fit to data with signal component fixed to $m_H = 126.5\,\text{GeV}$ superimposed. The bottom inset displays the residuals of the data with respect to the fitted background component. [11]



(c) Distribution of the four-lepton invariant mass for the $ZZ \to 4l$ analysis at CMS. The points represent the data, the filled histograms represent the background expectation, and the open histogram shows the signal expectation for a Higgs boson of mass $m_H = 125\,\text{GeV}$, added to the background expectation. [12]

(d) Distribution of the four-lepton invariant mass for the $ZZ \to 4l$ analysis at ATLAS compared to the background expectation in the mass range 80 GeV – 250 GeV. The signal expectation for a SM Higgs with $m_H = 125\,\text{GeV}$ is also shown. [11]

Figure 3.5: Invariant mass distributions of Higgs candidate decay products in the decay channels first observed at ATLAS and CMS



(a) The observed local background-only p-value for 7 TeV and 8 TeV data, and their combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H. [12]

(b) The observed (solid) local background-only p_0 as a function of m_H. The dashed curve shows the expected local p_0 under the hypothesis of a SM Higgs boson signal at that mass with its plus/minus one sigma band. [11]

Figure 3.6: Background-only hypothesis p-values in the decay channels where the new bosons was first observed at ATLAS and CMS

3.3.1 Addendum: 2013 Nobel Prize in Physics

In the light of the evidence presented in the aforementioned publications, the 2013 Nobel Prize in Physics was awarded to Peter Higgs and Francois Englert

> *"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider."*
>
> - Press release by the Royal Swedish Academy of Sciences, 8 October 2013

3.4 $H \to b\bar{b}$ in associated production with a vector boson

In the light of the Higgs boson discovery in 2012 (see Section 3.3) which was driven by bosonic decay channels, the interest in searches in the decay channel $H \to b\bar{b}$ has risen - after all this is the dominant decay channel for a Higgs boson with a mass of 126 GeV. Yet $H \to b\bar{b}$ did not play a role in the 2012 observation, because this channel is also notoriously difficult to exploit. Although b-jets can be efficiently distinguished from jets originating from other flavors with the ATLAS detector (see Section 7.3.5), the rate of di-jet events produced in QCD processes at the LHC is simply too large to allow for any search in $gg \to H \to b\bar{b}$. A similar argument applies to the purely hadronic final state produced by the vector boson fusion process $qq \to qqH \to qqb\bar{b}$ which is by far outnumbered by the irreducible QCD 4-jet $qqb\bar{b}$-background [27]. One thus has to resort to the associated production with a vector boson which - at least in the leptonic decay channels - can be efficiently triggered on.

3.4.1 The cross-sections of Higgs-strahlung processes

The cross-sections for the Higgs-strahlung processes have been calculated to NNLO for the QCD corrections [28], [29] and to NLO for the electroweak corrections [30] leaving a residual uncertainty of

$\lesssim 5\%$. The radiative QCD corrections to NLO are due to the $q\bar{q}$ vertex being modified by virtual gluon exchange, quark self-energy corrections and the emission of an additional gluon - they are thus identical to the corrections to the Drell-Yan process [31], see Figure 3.7.

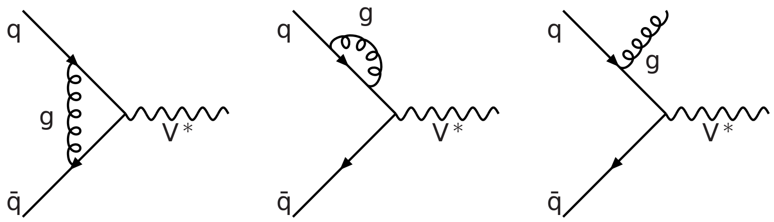


Figure 3.7: Exemplary Feynman diagrams of the NLO QCD corrections to associated Higgs production. They are identical to the NLO corrections to the Drell-Yan process [2].

The NNLO corrections comprise a larger set of contributions: two-loop correction to the $q\bar{q}$ vertex, one-loop corrections to $qg \to qV^*$ and $q\bar{q} \to gV^*$ and all possible tree-level contributions with two additional partons. While the Drell-Yan-like corrections are completely known to NNLO both for integrated and differential observables [21], other contributions have only been computed for the total cross-section [21]. A notable non-Drell-Yan NNLO correction for the ZH channel is a gluon induced production mechanism via quark loops [32] which yields a non-negligible contribution to the total cross-section at the LHC due to the high gluon luminosity. The corresponding diagrams are shown in Figure 3.8.

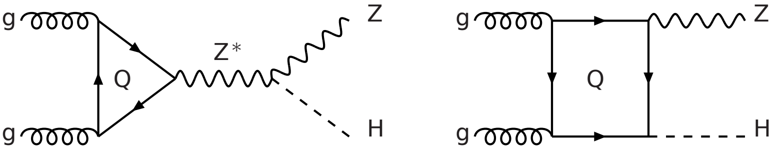


Figure 3.8: Feynman diagrams of the additional gluon induced NNLO QCD production mechanisms for ZH [2].

The NLO electroweak corrections to the total $H \to VH$ cross-section comprise hundreds of Feynman diagrams [30], a few exemplary ones are shown in Figure 3.9. It turns out that the largest contribution comes from bosonic one-loop corrections which outweigh both the fermionic loop corrections and the photonic initial state corrections. Since these bosonic contributions are negative, they reduce the $q\bar{q} \to VH$ production cross-section at the LHC by 5-10% [2]. Nevertheless the total K-factor for the NNLO QCD + NLO electroweak corrections is larger than one since the positive QCD corrections dominate the negative electroweak terms. The total cross-sections and K-factors for both WH and ZH production are shown in Figure 3.10. For a Higgs boson mass of 126 GeV, one obtains [21]:

$$\begin{aligned} \sigma_{WH} &= 563.3\,\text{fb} \pm 3.7\,\text{fb}, & \sqrt{s} &= 7\,\text{TeV} \\ \sigma_{ZH} &= 326.7\,\text{fb} \pm 6.0\,\text{fb}, & \sqrt{s} &= 7\,\text{TeV} \\ \sigma_{WH} &= 686.1\,\text{fb} \pm 3.5\,\text{fb}, & \sqrt{s} &= 8\,\text{TeV} \\ \sigma_{ZH} &= 405.1\,\text{fb} \pm 5.9\,\text{fb}, & \sqrt{s} &= 8\,\text{TeV} \end{aligned} \tag{3.21}$$

The given uncertainties include the residual scale dependence and the PDF uncertainties. The dependence on the renormalization and factorization scales is evaluated by mutually fixing one of them at the central scale $\sqrt{k^2}$ of the calculation, which was chosen to be the mass of the intermediate gauge boson, and then varying the other one between $\sqrt{k^2}/3$ and $3\sqrt{k^2}$. The PDF uncertainties are calculated according to the PDF4LHC recommendation [29].

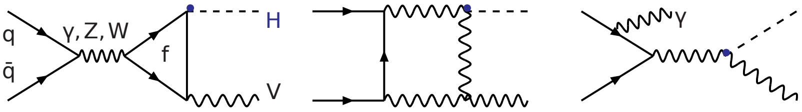


Figure 3.9: Exemplary Feynman diagrams of the NLO electroweak corrections to associated Higgs production [2].

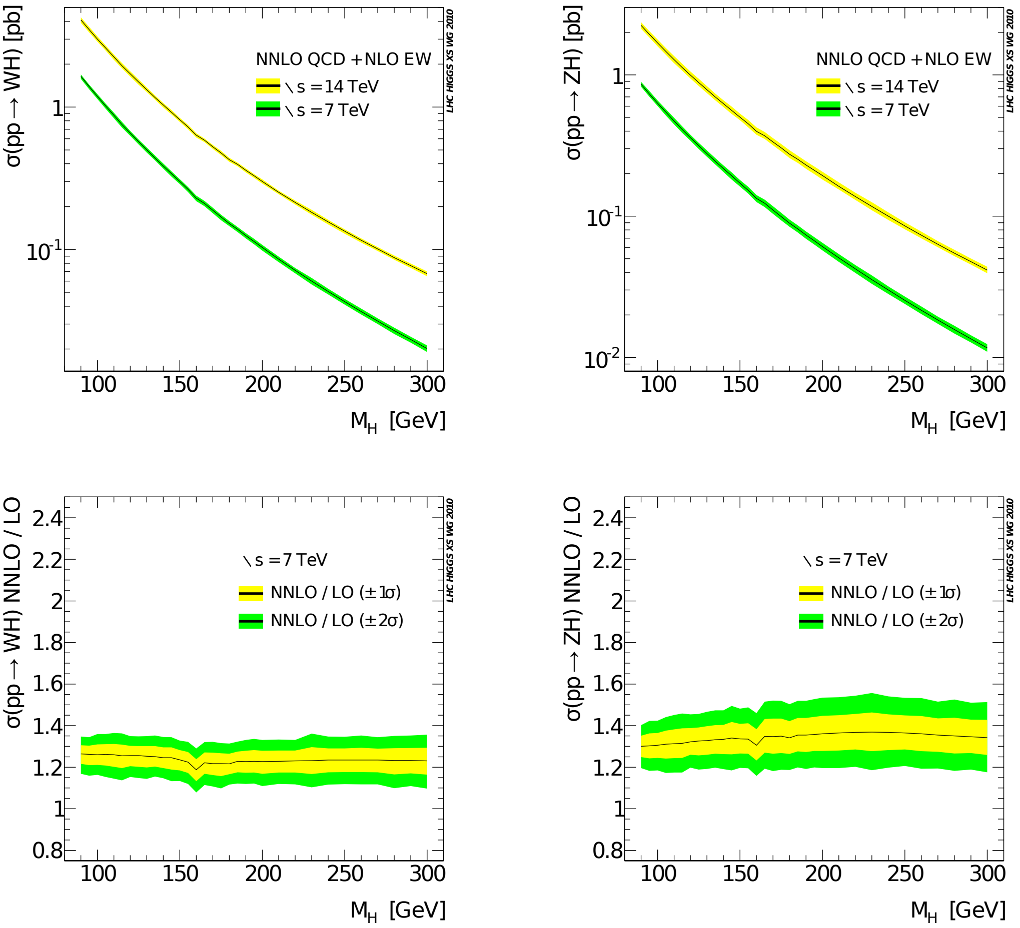
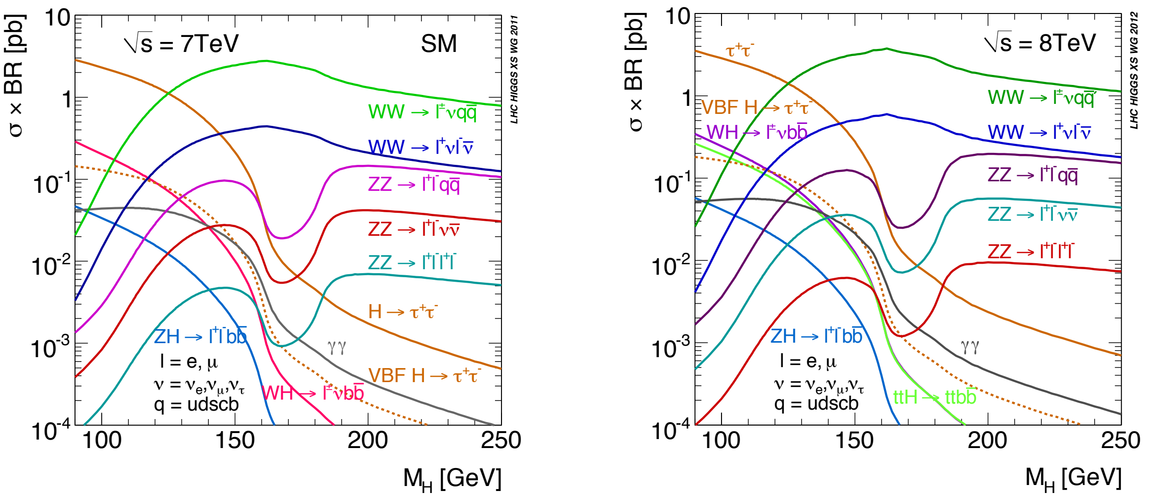


Figure 3.10: Total cross sections and NNLO K-factors for WH and ZH production at the LHC

3.4.2 The $H \to b\bar{b}$ decay channel and its impact on Higgs coupling measurements

As was already pointed out before, the decay $H \to b\bar{b}$ is the most abundant one for a Higgs boson mass of $m_H = 126\,\mathrm{GeV}$, but it can only be accessed via a search in the associated production channels. Taking into account the cross-sections (see Section 3.4.1) of these production mechanisms, one arrives at the plot shown in Figure 3.11 where total cross-sections and branching ratios for various channels have been multiplied.



(a)

(b)

Figure 3.11: Higgs production cross-sections multiplied by Higgs branching ratios for $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$. If not stated otherwise in the plot, the production mechanism is gluon-gluon fusion.

Even though the $VH \to Vb\bar{b}$ channel may not be suited for an immediate discovery, it is indispensable for a measurement of the Higgs boson couplings and thus must be studied in order to verify that the Higgs boson discovered at the LHC in 2012 (see Section 3.3) is indeed the Higgs boson of the Standard Model. In fact, the sensitivity of the search in the $VH \to Vb\bar{b}$ channel will not only determine the achievable accuracy for a measurement of the $Hb\bar{b}$ coupling $g_{Hb\bar{b}}$ but will also strongly affect the measurement of other Higgs couplings. It can be shown that

$$\sigma \cdot \mathrm{BR}(H \to XX) \propto \frac{\Gamma_{prod}\Gamma_{H\to XX}}{\Gamma_H^{tot}} \tag{3.22}$$

where Γ_{prod} is the partial decay width associated with the Higgs coupling in the respective production process with cross-section σ, and $\Gamma_{H\to XX}$ is the partial width associated with the coupling involved in the decay $H \to XX$ with branching ratio $\mathrm{BR}(H \to XX)$. That means that any measurement of a given coupling g_{HXX} through $\sigma \cdot \mathrm{BR}$ (which can be obtained from the rate of signal events) will rely on a precise knowledge of the overall normalization given by the total width Γ_H^{tot}. Since the total width is dominated by $\Gamma_{H\to b\bar{b}}$ for a Higgs with a mass of $m_H = 126\,\text{GeV}$, the Higgs coupling to b-quarks will be of particular importance in this context. An interesting study of the importance of the $VH \to Vb\bar{b}$ channel for the determination of the Higgs couplings is presented in [33]: on the basis of a simulation corresponding to $30\,\text{fb}^{-1}$ of data recorded at the LHC at $\sqrt{s} = 14\,\text{TeV}$, the sensitivity to detect any deviation from the Higgs couplings predicted by the Standard Model is evaluated. To do this, the couplings are parametrized according to

$$g_{HXX} = g_{HXX}^{SM}(1 + \Delta_{HXX}) \tag{3.23}$$

where $\Delta_{HXX} \neq 0$ would indicate a deviation from the Standard Model prediction. Using simulated data generated with $g_{HXX} = g_{HXX}^{SM}$, the compatibility of $\Delta_{HXX} \neq 0$ with the signal event rates observed in different analysis channels is evaluated. The measure of compatibility for a given value of Δ_{HXX}

is given by the ratio $\Delta(\chi^2)$ of the (negative log) likelihoods of Δ_{HXX} and the best fit value Δ_{HXX}^{ML}. By construction $\Delta(\chi^2) = 1$ for Δ_{HXX}^{ML}, which is also the minimal value obtainable. The importance of the $VH \to Vb\bar{b}$ channel can clearly be seen in Figure 3.12: the top row shows the $1/\Delta(\chi^2)$ distributions for a measurement which includes information about the coupling $g_{Hb\bar{b}}$ obtained from $VH \to Vb\bar{b}$, the middle row shows the same distributions with the sensitivity of $VH \to Vb\bar{b}$ reduced by 50%, and the bottom row relies on $g_{Hb\bar{b}}$ being solely obtained from $t\bar{t}H \to t\bar{t}b\bar{b}$. Multiple solutions (peaks) arise from the fact that the partial widths depend quadratically on the couplings in some cases, leading to a degeneracy with respect to the sign of Δ_{HXX}. Without the information from $VH \to Vb\bar{b}$, not only the determination of $\Delta_{Hb\bar{b}}$ becomes virtually impossible, but also the uncertainties of Δ_{HWW} and Δ_{HZZ} increase by 50% to 100% [33].

3.5 Important background processes for the search in $VH \to Vb\bar{b}$

The selection of candidate events for the $VH \to Vb\bar{b}$ search is based on the identification of a leptonically decaying heavy vector boson and two b-jets. While the signature required in the ZH channel is comprised of - apart from the b-jets - two high p_{T} leptons of the same flavor, only one high p_{T} lepton is required in the WH channel. Any additional high p_{T} leptons and jets are vetoed. A cleaner set of signal candidates could be obtained in the WH channel by exploiting the fact that the associated neutrino shows up in the detector as missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$, but it has turned out that the analysis in this channel reaches a higher sensitivity if no explicit requirement regarding $E_{\mathrm{T}}^{\mathrm{miss}}$ is made in the pre-selection (see Section 7.7 for details). Before any further selection is applied, the signal region is thus populated by all processes which generate or fake exactly two high p_{T} b-jets and exactly one (WH) and two (ZH) high p_{T} lepton(s), respectively. The most frequent background processes which can result in this signature will be discussed in the following.

top-antitop pairs: The $t\bar{t}$-background contains two real b-jets resulting from the top decays and can also contain one or two real high p_{T} leptons depending on the decay channel of the W bosons. Although a large fraction of the $t\bar{t}$ events is rejected by vetoing a third jet in the signal region, it is still the most abundant background in the WH channel and also represents an important contribution in the ZH channel. The cross-section for this background is $\sim$ 166.8 pb for $\sqrt{s}$ = 7 TeV and $\sim$ 238.1 pb for $\sqrt{s}$ = 8 TeV.

single top: All of the three production mechanisms of single top (t-channel $gq \to qt(b)$, s-channel $qq \to tb$ and Wt $gb \to Wt$) can fake the signal signature. The t-channel however has by far the highest cross-section ($\sim$ 64.6 pb for $\sqrt{s}$ = 7 TeV and $\sim$ 87.8 pb for $\sqrt{s}$ = 8 TeV) and is thus the dominant source of background.

W+jets: With 10.46 nb (7 TeV) and 10.97 nb (8 TeV), respectively, the cross section for inclusive production of W bosons and subsequent leptonic decays is more than four orders of magnitude higher than the cross-section of the WH signal. While most of this background can be rejected by jet multiplicity cuts and b-tagging, there is a sizable amount of events where the W-boson is produced together with two real b-jets, e.g. through $qq \to Wg \to Wb\bar{b}$, resulting in an irreducible background component for the WH channel. Events involving jets which originate from c-quarks are also likely to fake the desired signature. In the analysis presented here, the W+jets background is categorized according to the flavor of the associated jets, and the relative normalization of the different flavor components is determined from dedicated control regions (see Section 7.8).

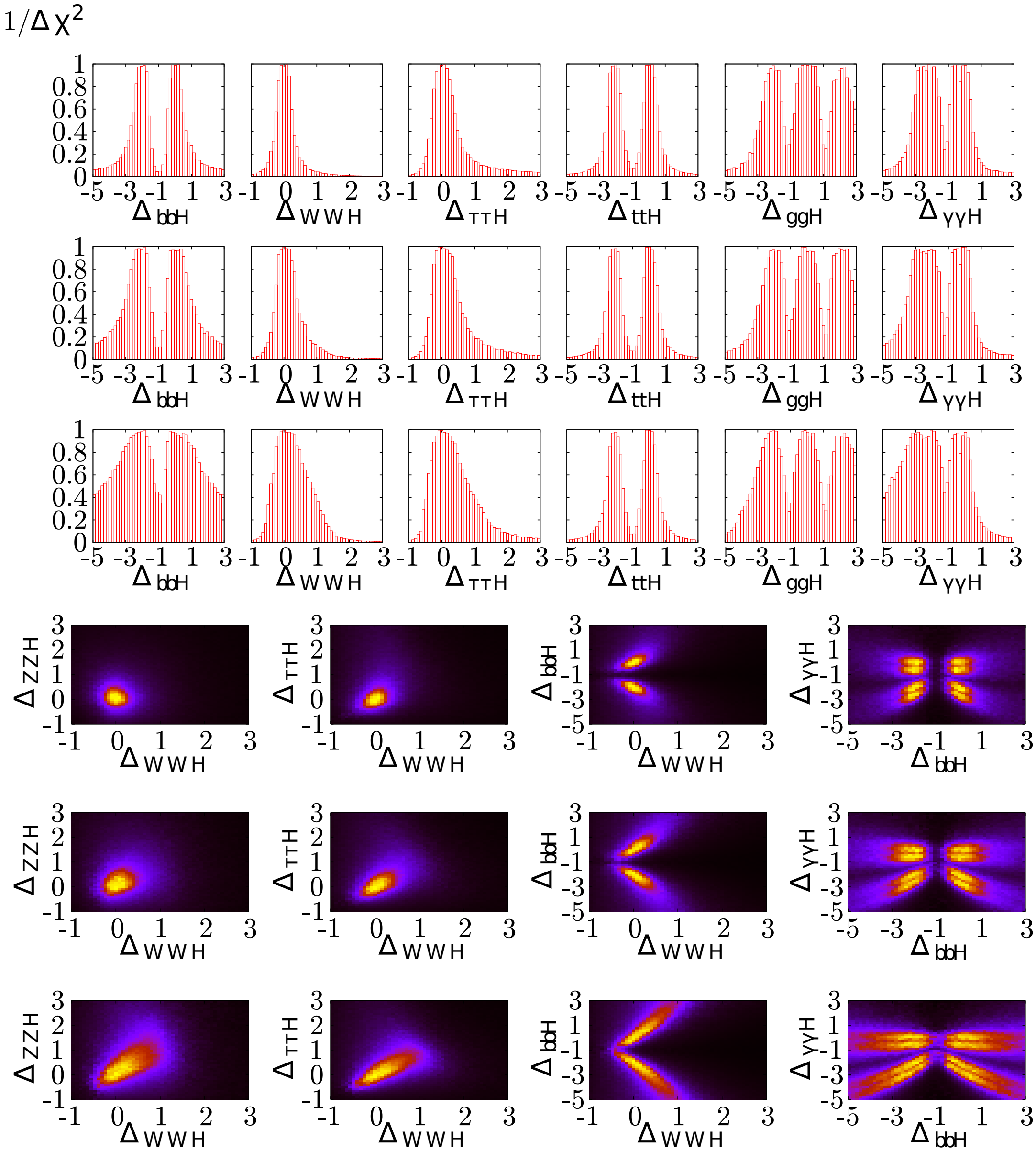


Figure 3.12: Sensitivity to deviations Δ_{HXX} of the Higgs couplings from the Standard Model prediction. Top row: Including information from all analysis channels, in particular $VH \to Vb\bar{b}$. Middle row: Impact of the sensitivity in $VH \to Vb\bar{b}$ being reduced by 50%. Bottom row: $VH \to Vb\bar{b}$ excluded from the fit, information about the coupling $g_{Hb\bar{b}}$ solely obtained from $t\bar{t}H \to t\bar{t}b\bar{b}$ [33].

Z+jets: The Z+jets background plays a similar role for the ZH search as W+jets does for the WH search. The cross-section for inclusive Z production followed by a decay to a pair of charged leptons with an invariant mass of $> 40\,\text{GeV}$ is $\sim 1.1\,\text{nb}$ ($\sqrt{s} = 7\,\text{TeV}$) and $\sim 1.2\,\text{nb}$ ($\sqrt{s} = 8\,\text{TeV}$), respectively. The heavy flavor component $Z + b\bar{b}$ constitutes the most important source of background for the ZH channel and is very difficult to discriminate against. As done for W+jets in the WH channel, the flavor composition of the Z+jets background is measured using control regions with different b-tag multiplicities in the ZH channel.

Diboson: The production of heavy gauge boson pairs (WW, WZ, ZZ) forms an important background source for both the WH and the ZH channel. In particular the decays $WZ \rightarrow Wb\bar{b}$ ($\sigma^{66<m_{ll}<116\,\text{GeV}}_{WZ,8\,\text{TeV}} \approx 22.7\,\text{pb}$) and $ZZ \rightarrow Zb\bar{b}$ ($\sigma^{66<m_{ll}<116\,\text{GeV}}_{ZZ,8\,\text{TeV}} \approx 7.7\,\text{pb}$) yield irreducible background components for the Higgs boson search in associated production.

QCD multijet production: The sheer abundance of QCD-induced multijet events produced in proton-proton collisions renders these processes an important background source. Real production of heavy flavors followed by semi-leptonic decays, e.g. from $gb \rightarrow b\bar{b}b$ or $gc \rightarrow b\bar{b}c$, contributes as well as fake leptons and mistagged light flavor jets. Due to the immense cross section and the low acceptance of these processes, no reliable simulation is available for the multijet background. In this analysis, it is modeled using shape templates obtained from data which are selected in a dedicated multijet control region. See Section 7.4 for a detailed description.

The description of all the background processes - except multijet production - relies on simulated events which are generated with so called "Monte Carlo generators". Since these simulations play a crucial role for the analysis presented here, I will discuss some important technical aspects of event generation in Chapter 5. A complete list of the Monte Carlo samples used in this analysis can be found in Section 5.5.

*"Any sufficiently advanced technology is
indistinguishable from magic."*
ARTHUR C. CLARKE

CHAPTER 4

The Large Hadron Collider and the ATLAS experiment

The search for new particles like the Higgs boson requires the use of particle accelerators, where heavy particles can be produced resonantly from two colliding particles if the center-of-mass energy $\sqrt{s}$ in the collision exceeds the rest mass of the particle in question. Apart from the energy, also the rate at which the collisions take place is an important characteristic to describe the capacity of a particle accelerator, since the observation of particles with small couplings (and thus small production cross-sections) requires a large number of recorded collisions. On the other hand, a high event rate poses a challenge for the detectors, requiring very fast and radiation-hard electronics. The Large Hadron Collider (LHC) and the ATLAS[1] detector are designed to meet the requirements that arise in the search for new physics up to the range of several TeV. They define the experimental frame in which the data for this thesis have been recorded. I will shortly discuss the technical key features of the LHC and the ATLAS detector in the following.

4.1 The Large Hadron Collider

The LHC, situated at the *European Laboratory for Particle Physics* (CERN) near Geneva, Switzerland, is a circular proton-proton[2] collider with a circumference of 26.7 km. It is installed in an underground tunnel which previously hosted the Large Electron Positron Collider (LEP). The LHC provides a rich physics potential, which is exploited by four major experiments at four different sites where the particle beams are brought to collision: ALICE, a detector designed to study the properties of quark-gluon plasma, and LHCb, investigating b-quark physics and matter-antimatter asymmetries, are two experiments with dedicated tasks. ATLAS and CMS on the other hand are so-called multi-purpose detectors which are intended for an extensive investigation of physics phenomena in the TeV range, including the search for the Higgs boson. A schematic view of the LHC and the four major experiments is shown in Figure 4.1.

The LHC was designed to reach a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$, which makes it the most powerful particle accelerator ever constructed, surpassing the second most powerful accelerator (the

---

[1] **A T**orodial **L**HC **A**pparatu**S**

[2] There also exists the possibility to collide lead ions for specific experiments which are not of interest in the context of this thesis

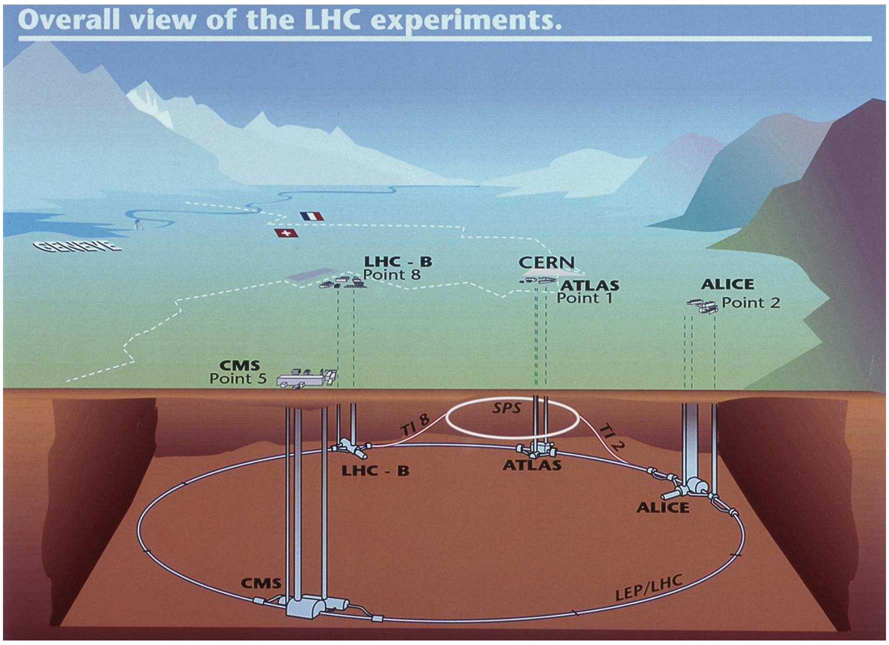


Figure 4.1: Overview of the LHC and the four major experiments [34].

Tevatron at Fermilab with $\sqrt{s} \approx 2\,\text{TeV}$) by a factor of seven. The LHC started operation in late 2008, but had to be shut down merely a week later due to a severe malfunction. The restart did not happen until late 2009. Since then, the LHC has been running with a reduced center-of-mass energy to avoid any risk of further technical failure. The data analyzed in this thesis were taken with a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $\sqrt{s} = 8\,\text{TeV}$ in 2012, respectively. At the time of this writing, the LHC is undergoing a shutdown for maintenance and upgrades, it will be operational again in 2015, presumably reaching its design center-of-mass energy.

The highly energetic proton beams in the LHC circulate in two ultrahigh vacuum (1×10^{-13} bar) tubes to avoid collisions with gas atoms. They are forced on their circular trajectories by powerful dipole magnets which create field strengths of up to 8 T. This is achieved by superconducting coils whose operation requires the accelerator to be cooled to 1.9 K using liquid helium. The accelerating cavities, which provide a field of $5\,\text{MV}\,\text{m}^{-1}$, are also superconducting. Since the LHC accelerates protons in both directions, two magnet systems with opposite orientation - one for each beam - have to be enclosed in a common cryostat vessel. The complete construction comprising two vacuum tubes, the bending dipole magnets and the surrounding cryostat is often referred to as “a dipole”, each of which is 15 m long and weighs ~ 35 t. The LHC consists of 1232 of these dipoles, a diagram of their cross-section is shown in Figure 4.2.

4.1.1 Luminosity

The LHC is designed as a discovery machine, not only by virtue of its high center-of-mass energy, but also by virtue of its formidable luminosity, which allows for reasonable event rates even for many processes with very low cross-sections. For a given process with cross-section σ, the luminosity $\mathcal{L}$ is defined as

$$\mathcal{L} = \frac{\dot{N}}{\sigma} \tag{4.1}$$

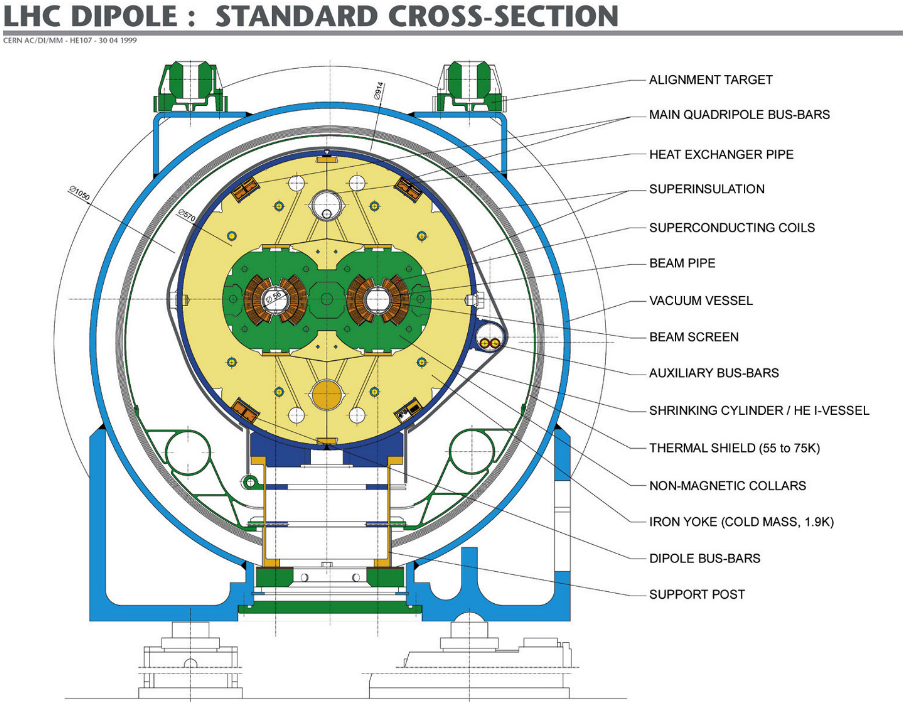


Figure 4.2: Cross-section of a LHC dipole magnet [35].

where $\dot{N}$ is the event rate of that process. While the cross-section is entirely determined by the center-of-mass energy, the luminosity and thus the event rate can be deliberately tuned with an appropriate accelerator design. At the LHC, the protons are not accelerated in a continuous beam but are stored in packets, so called bunches. Labeling the rate at which these bunches collide f, and the number of protons inside a bunch n, the luminosity is given by

$$\mathcal{L} = \frac{fn^2}{A_{int}} \tag{4.2}$$

where A_{int} is the effective area in which the beams are brought to collision. It is obvious that a highly focused beam (i.e. small A_{int}) is desirable in order to reach a high luminosity. The LHC can support up to 2808 bunches being accelerated at the same time, each one containing about 10^{11} protons. This corresponds to a bunch spacing of only 25 ns or 7 m. Operating at its design values, the LHC reaches a luminosity of $\mathcal{L}_{design} = 1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The data analyzed in this thesis however have been collected with different luminosity settings: the bunch spacing was only gradually lowered from 75 ns to 50 ns, reached by mid 2011. On the other hand, the number of protons inside each bunch and the beam focus were constantly improved until both parameters finally exceeded the design values, so that the peak luminosity provided in the second half of 2012 was very close to the LHC design value, see Figure 4.3.

A very high luminosity is not only beneficial in terms of high event rates for rare processes (like Higgs boson production), it also poses a serious challenge when it comes to the detection and selection of interesting events: high bunch crossing frequencies require a short response and processing time, or the following events will be lost. A high number of interactions per bunch crossing requires a robust suppression of pile-up. The ATLAS detector was designed with these requirements in mind, and I will present a short overview of the detector components that are crucial for the Higgs boson search in the $VH \to Vb\bar{b}$ channel in the following.

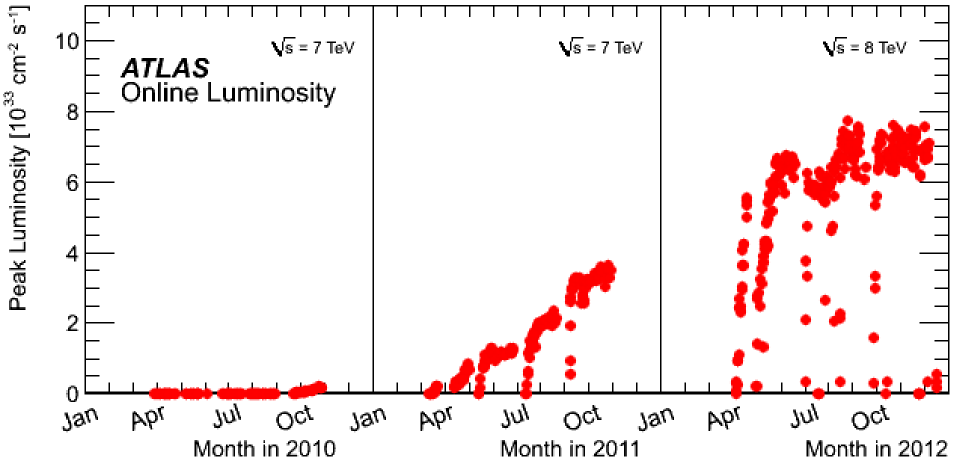


Figure 4.3: Luminosity delivered to the ATLAS experiment by the LHC in 2010-2012.

4.2 The ATLAS experiment

ATLAS is a multi-purpose particle detector at the LHC, designed to operate in the high energy and high particle multiplicity regime provided by the LHC. ATLAS is barrel-shaped, 42 m long and 22 m high, providing almost full coverage of the azimuthal angle and a large acceptance in the polar angle. Its appearance is dominated by the large toroid magnet system which provides the field for the muon system (see Section 4.2.3). The detector encloses the beam-pipe with the interaction point at which the proton bunches are brought to collision. Particles created in the collisions thus traverse the detector from the innermost layer outwards. This is reflected in the design of ATLAS, with different sub-detectors being arranged in concentric layers around the beam-pipe. The subsystem which is closest to the interaction point is the inner detector, followed by the electromagnetic calorimeter, the hadronic calorimeter and finally the muon system. A computer-generated cross-section of ATLAS is shown in Figure 4.4, it illustrates both the concentric design and the immense size of the detector. A more detailed description of the ATLAS detector and its subsystems can be found in [36].

The right-handed coordinate system in which physical processes in the ATLAS detector are described is defined as follows: the z-axis is aligned with the beam-pipe, the y-axis points towards the center of the LHC ring and the x-axis points upwards. The origin of the coordinate system is defined as the nominal interaction point. The azimuthal angle ϕ is the angle to the x-axis, measured around the beam, and the polar angle θ is measured to the beam-axis. The latter is usually replaced by the so-called pseudo-rapidity

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{4.3}$$

which has the advantage that differences in η are invariant under longitudinal Lorentz-boosts for massless objects, an approximation that typically holds for most objects produced at LHC energies. Particles that traverse the detector perpendicular to the beam have $\eta = 0$, while the beam itself is at $\eta = \pm\infty$. Using ϕ and η, three-dimensional distances in the ATLAS coordinate system can be defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \tag{4.4}$$

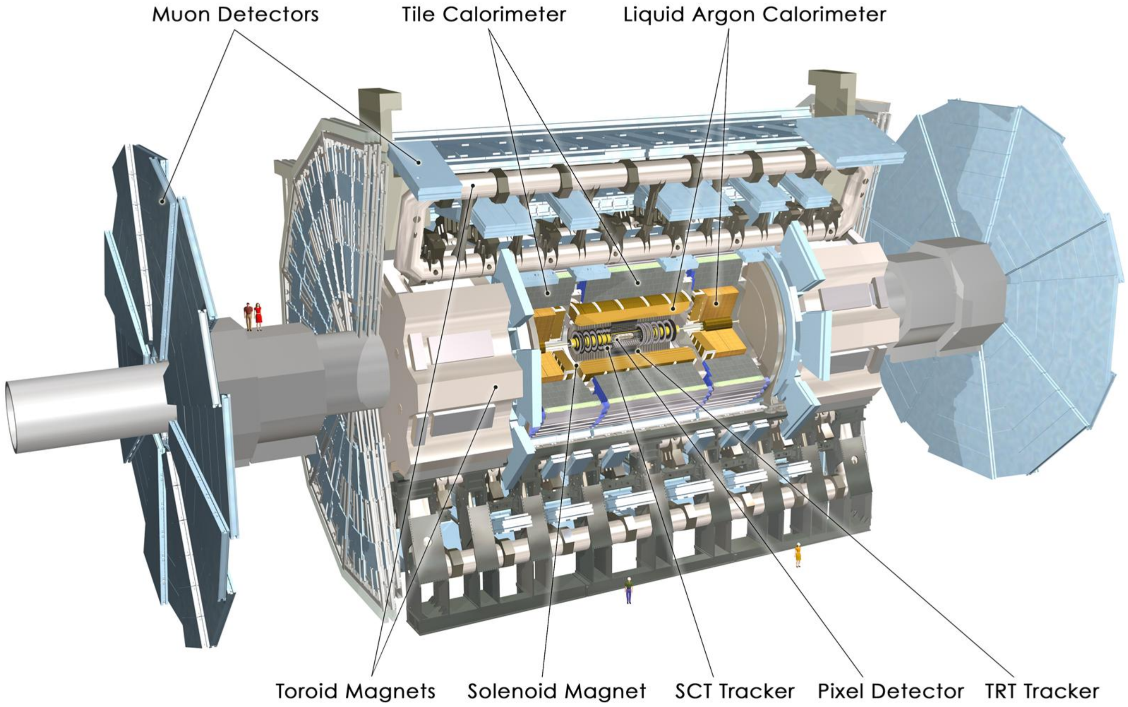


Figure 4.4: Computer generated cross-section of the ATLAS detector, illustrating the concentric arrangement of the sub-detectors [37].

4.2.1 The inner detector

The inner detector (ID) is a so-called tracking device which serves to measure the trajectories of electrically charged particles, called *tracks*. It is ~6.2 m long and ~2.1 m in diameter and covers the range $|\eta| < 2.5$. The inner detector is surrounded by a solenoid magnet which immerses it in a 2 T magnetic field, causing the tracks of charged particles to bend. This allows to measure the momentum and the sign of the electric charge of the particles via the curvature of the track. Tracks can also be used to determine the position of the interaction vertex - this is particularly important if one aims to identify long-lived particles (like hadrons with b-quarks, as in this analysis), since their decay may lead to the presence of secondary vertices which are displaced from the primary one.

The inner detector itself consists of three subsystems with different granularities: the closer to the interaction point, the higher the resolution of a given detector component has to be in order to resolve single tracks. A computer generated cross-section of the inner detector components is shown in Figure 4.5, while Figure 4.6 displays a more detailed diagram of the subsystems.

Pixel detector

The innermost sub-detector is the pixel detector, surrounding the beam-line at a distance of only ~5 cm. It provides the highest granularity of all ATLAS systems, consisting of three layers of silicon sensors ("pixels") with a total of 80 million readout channels. The first layer, which is particular important for precise impact parameter measurements which in turn play a crucial role in the identification of long-lived particles, is called the b-layer. The pixel sensors are segmented in $R-\phi$ and z with individual pixel cells as small as $50\,\mu\text{m} \times 400\,\mu\text{m}$. This allows for a measurement accuracy of $10\,\mu\text{m}$ ($R-\phi$) and $115\,\mu\text{m}$

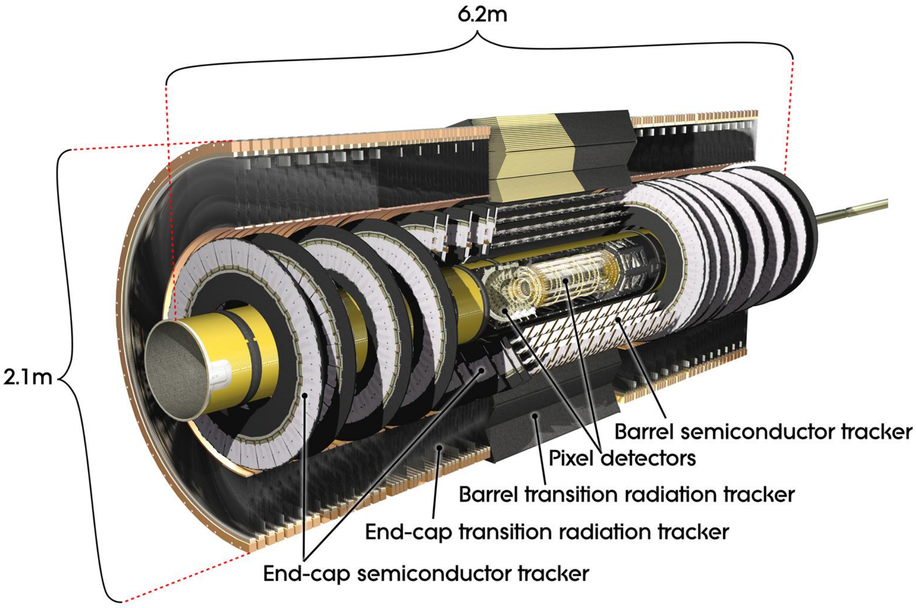


Figure 4.5: Computer generated cross-section of the inner detector [38].

(z), respectively. The cells are diodes which are depleted by a high voltage when in operation. Charged particles passing through the cells create free charge carriers in the form of electron-hole pairs, which are then collected to generate an electric signal.

Semi conductor tracker

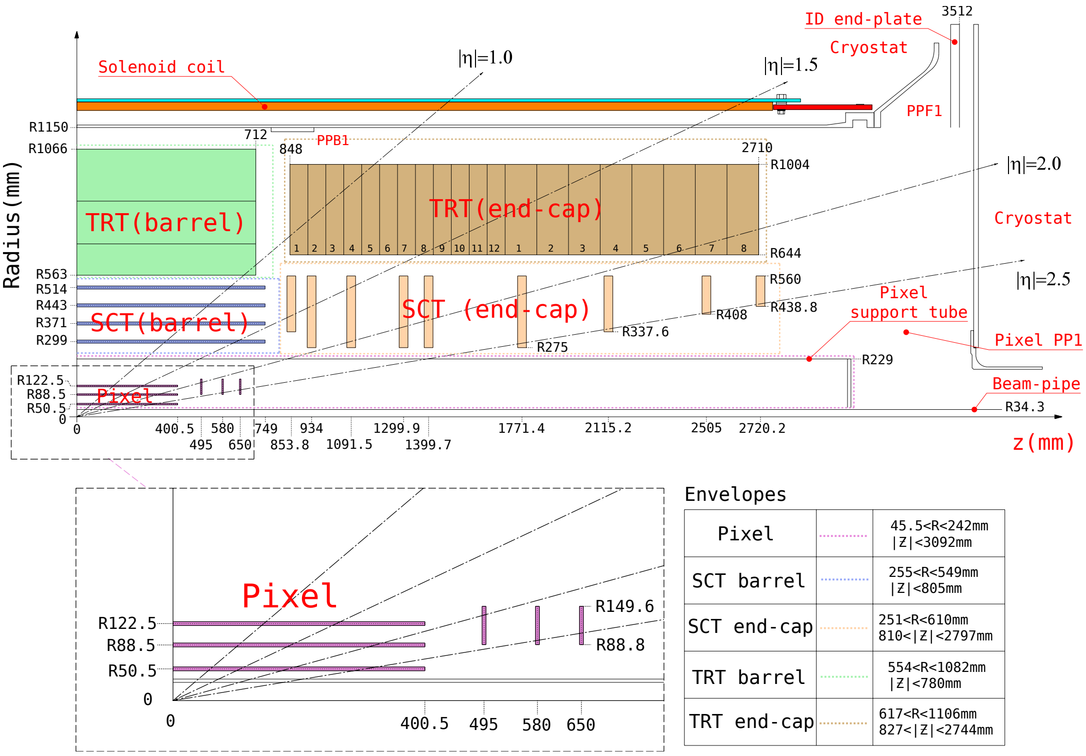
The semi conductor tracker (SCT) is a second silicon-based tracking device which encloses the pixel detector. Consisting of strips with a width of 80 µm and a length of 12.8 cm arranged in four (barrel) and nine (end-cap) double-layers, respectively, it provides ~6.3 million readout channels. The double-layers consist of sensors which are installed back to back at a stereo angle of 40 mrad to provide a two-dimensional position measurement. The SCT delivers a measurement accuracy of 17 µm ($R - \phi$) and 580 µm (z), covering the range $|\eta| < 2.5$.

Transition radiation tracker

The transition radiation tracker (TRT) constitutes the outermost layer of the inner detector. It consists of ~ 300000 so-called straw-tubes with a diameter of 4 mm each. The tubes in the barrel region have a length of 144 cm and are aligned with the beam axis, while the 37 cm long tubes in the end-caps are radially aligned. In contrast to the pixel detector and the SCT, the TRT only covers the range $|\eta| < 2.0$. Each tube of the TRT is filled with a mixture of xenon, carbon-dioxide and oxygen and has a wire running trough it at the center, which is kept at a high positive voltage with respect to the shell of the tube. Charged particles traversing the tubes ionize the gas, creating free electrons which drift to the wire. Due to the high acceleration voltage, an avalanche of secondary electrons is produced, amplifying the original signal.

Compared to the pixel detector and the SCT, the TRT delivers a rather low intrinsic resolution, allowing only measurements in the $R - \phi$-plane with an accuracy of 130 µm per straw. However, the high number of measurements - on average there are 36 hits in the TRT for each charged particle - and the long lever arm render the TRT an important component for the overall momentum resolution of the inner detector.

Moreover, the TRT also facilitates particle identification via so-called transition radiation, which gives the TRT its name: ultra-relativistic particles crossing the boundary between a radiator material surrounding the tubes and the tubes itself emit transition radiation photons due to the different refractive indices of the two materials. These photons produce an additional signal when they ionize the gas in the tubes. This mechanism helps to distinguish electrons from charged pions since electrons have a much smaller mass and thus typically a higher Lorentz γ than pions.



Pixel	⋯	45.5<R<242mm |Z|<3092mm
SCT barrel	⋯	255<R<549mm |Z|<805mm
SCT end-cap	⋯	251<R<610mm 810<|Z|<2797mm
TRT barrel	⋯	554<R<1082mm |Z|<780mm
TRT end-cap	⋯	617<R<1106mm 827<|Z|<2744mm

Figure 4.6: Diagram of a quarter section of the ATLAS inner detector [36].

4.2.2 Calorimetry

The ATLAS calorimeter system is designed to determine the energy of particles and jets and provide an accurate measurement of the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. The latter requires almost complete hermeticity - in fact, the ATLAS calorimeter provides coverage for the range $|\eta| < 4.9$, which corresponds to angles as close as ~ 1 degree to the beam axis. To measure the energies of particles in the TeV range precisely, it is also important that the electromagnetic and hadronic showers resulting from the interaction of the particles with the calorimeter material are contained within the respective subsystem of the calorimeter. Escaping energy would both severely degrade the energy resolution and result in possible punch-throughs to the muon system, generating a false signal. The calorimeter thus has to be sufficiently thick: the material in the ATLAS calorimeter corresponds to ~ 10 interaction lengths λ.

As can be seen in Figure 4.7, the ATLAS calorimeter consists of different types of sampling calorimeters with varying granularity, optimized for either electromagnetic or hadronic showers.

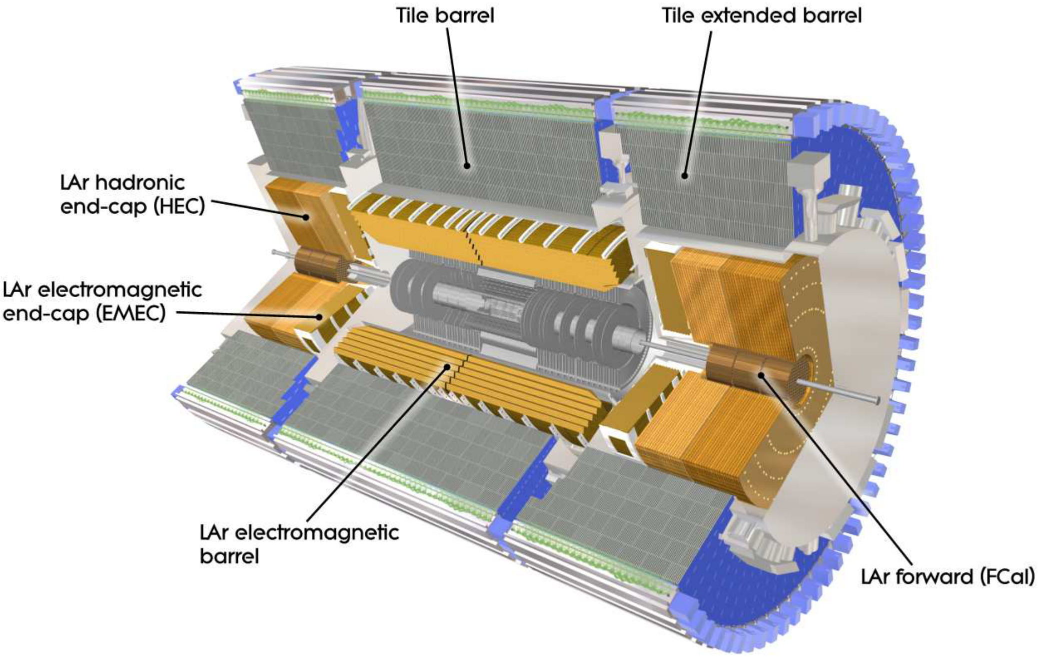


Figure 4.7: Cut-away view of the ATLAS calorimeter components [36].

Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is a sampling calorimeter using liquid argon as the active medium and lead as an absorber. Electrons traversing the absorber emit photons (bremsstrahlung) which in turn convert to electron-positron pairs. An electromagnetic cascade forms and the charged particles deposit energy in the active medium via ionization. The free charge carriers originating from the ionization are then collected and a signal is generated.

The ECAL is subdivided into a barrel segment ($|\eta| < 1.475$) and two end-cap components: an inner wheel covering the region $2.5 < |\eta| < 3.2$ and an outer wheel covering $1.375 < |\eta| < 2.5$. The characteristic accordion-shaped structure of the calorimeter allows for a complete azimuthal coverage without cracks. In the range covered by the inner detector, the EM calorimeter is segmented in three layers in depth, allowing for a precise measurement of electrons and photons. Energy losses of these particles in the inner detector are accounted for by a so-called pre-sampler consisting of an active liquid argon layer which precedes the ECAL.

Hadronic calorimeters

The hadronic calorimeter consists of three different systems in different regions of $|\eta|$:

Tile calorimeter: A sampling calorimeter placed directly outside the ECAL which covers the range $|\eta| < 1.7$. The tile calorimeter uses scintillating tiles as the active material and steel as the absorber. Inelastic hadronic interactions with the material initiate hadronic showers, secondary particles excite the scintillator material which then emits light. The light is guided to photomultipiers by wavelength shifting fibers, where it is converted to an electrical signal. The tile calorimeter is segmented in three layers in depths and divided in 64 modules in azimuthal direction.

LAr hadronic end-cap calorimeter: The hadronic end-cap calorimeter (HEC) is located behind the end-caps of the EM calorimeter. Since it has to withstand higher radiation doses than the tile calorimeter, it uses liquid argon as the active material, sharing the cryostats with the EM end-cap calorimeter. The HEC consists of two wheels per end-cap, each one constructed from 32 identical wedge-shaped modules and segmented in two layers in depth. The HEC covers the range $1.5 < |\eta| < 3.2$, thereby overlapping slightly with both the tile calorimeter and the forward calorimeter (FCal). This way the total material density is almost constant over the entire η-range.

LAr forward calorimeter: The forward calorimeter (FCal) covers the region $3.1 < |\eta| < 4.9$. Like the HEC, it has to withstand considerable radiation doses and thus uses liquid argon as the active material. It is divided into three modules for each end-cap, which makes it a kind of dual-use calorimeter for both electromagnetic and hadronic interactions: the first module, intended for electromagnetic measurements, is made of copper, while the two other modules are made of tungsten, measuring mostly hadronic energy deposition.

4.2.3 Muon spectrometer

Highly energetic muons are among the most important indicators of interesting physics signatures at the LHC. The muon system of the ATLAS detector is thus designed as a completely independent sub-detector, which ensures both the efficient identification of muons and a precise measurement of muon momenta even at high energies when the curvature of the track in the inner detector is small.

The ATLAS muon system forms the outermost shell of the detector - virtually all particles which are not stopped in the preceding calorimeters are muons, which allows for an easy identification. The long lever arm with respect to the inner detector also greatly improves the resolution of momentum measurements, which are - like in the inner detector - performed by measuring the curvature of the muon tracks. The magnetic field which bends the tracks is generated by the huge toroid magnets which dominate the overall appearance of the ATLAS detector. The so-called muon chambers, which provide the position information for the momentum measurement, employ a range of different technologies described below. They are arranged in concentric layers around the beam axis in the barrel and in four wheels perpendicular to the beam in the end-cap regions. The different components of the muon system are shown in Figure 4.8.

Monitored drift tube chambers: MDTs contain tubes with pressurized gas and a central wire at a high positive voltage which is used to collect free charge carriers from ionization. MDTs have a limited count rate of $\sim 150\,\mathrm{Hz\,cm^{-2}}$.

Cathode strip chambers: CSCs are multi-wire proportional chambers which can handle higher counting rates of up to $\sim 1000\,\mathrm{Hz\,cm^{-2}}$. They are thus used instead of MDTs in the first layer of the end-cap at $|\eta| > 2$.

Resistive plate chambers: RPCs consist of two parallel plates with a gas mixture in between. Charged particles traversing the 2 mm wide gap between the plates ionize the gas, initiating an electron avalanche in the strong electric field in the gap. RPCs provide fast tracking information for the muon trigger system in the barrel.

Thin gap chambers: TGCs are a variant of multi-wire proportional chambers with good time resolution and high rate capability. They provide input to the trigger system in the end-cap region.

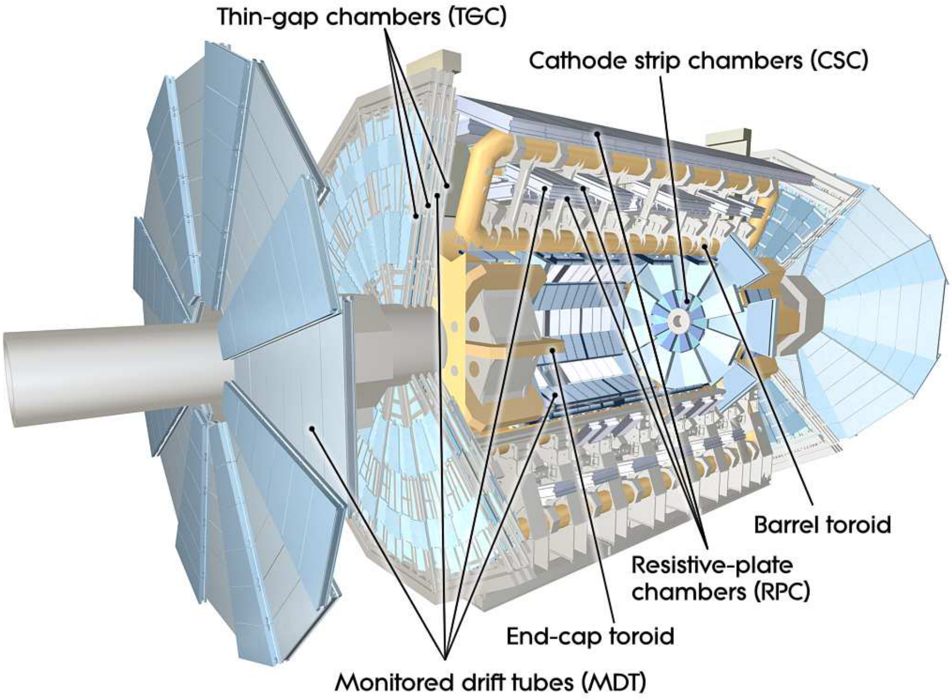


Figure 4.8: Cut-away view of the ATLAS muon system indicating the different muon chamber technologies [36].

4.2.4 Trigger system

The vast majority of bunch crossings does not yield an event that is considered “interesting” in the sense that the particles produced comprise a signature which could indicate a rare physics process. Storing and analyzing all events would thus be a major waste of resources. Apart from that, no readout or storage system available today could possibly cope with the gigantic rate at which data is generated by the ATLAS detector. To deal with this difficulty, a so-called “trigger system” is employed to select interesting signatures on-the-fly, discarding most events without the need to write them to disk.

The ATLAS trigger system has three levels, each one refining the selection made at the previous level by considering more detector information. The Level-1 (L1) trigger uses reduced-granularity information from the muon trigger chambers and the calorimeter to identify candidates for high p_{T} muons, electrons, photons, jets and τ-leptons and events with large $E_{\mathrm{T}}^{\mathrm{miss}}$. Since is does not use any information from the tracking system one cannot select b-jet candidates at this level. The L1 trigger is entirely hardware based and reaches a decision in $\lesssim 2.5\,\mu\mathrm{s}$. It also defines one or more regions-of-interest (RoI) - that is the η and ϕ coordinates where interesting features have been identified - which in turn seed the higher level trigger algorithms. The event rate after L1 trigger selection is $\sim 75\,\mathrm{kHz}$.

The Level-2 (L2) trigger and the event filter (EF) are based on regular computing hardware and perform a refined event selection by taking into account more event information. The L2 trigger has access to the full detector granularity, but considers only data from within the RoIs generated by L1. It reaches a decision in $\sim 40\,\mathrm{ms}$, reducing the event rate to $\sim 3.5\,\mathrm{kHz}$. The final decision on the event is taken by the event filter, considering the full detector information and an improved event reconstruction. Typical processing times are on the order of four seconds and the final event rate is $\sim 200\,\mathrm{Hz}$.

4.3 Event reconstruction in ATLAS

Physics analyses like the one presented in this thesis rely on mappings which allow to identify certain patterns in the detector's electronic response with physical objects like electrons, muons or particle jets. Such mappings are provided by the so-called reconstruction algorithms which transform energy depositions or pixel cell hits into higher-level physics objects. For the analysis at hand, a multitude of different objects has to be considered and I will briefly review the reconstruction algorithms associated with those objects in the following. It should however be noted that the ATLAS reconstruction software is not static, but undergoes constant evolution.

4.3.1 Track reconstruction

Charged tracks are reconstructed from inner detector measurements (see Section 4.2.1) - a task known as "tracking" which consists of three stages:

Pre-processing: Raw data from the pixel and SCT detectors are converted into clusters, TRT timing information is converted into drift circles and SCT clusters are in turn converted to space-points.

Track-finding: Tracks are seeded by clusters in the first three pixel layers and the first SCT layer. First track candidates are formed by extending the seeds through the SCT, these candidates are then fitted and ambiguities are removed. Fake tracks are rejected by quality cuts requiring a certain number of associated clusters. The remaining candidates are extrapolated to the TRT and associated with matching drift circles. The track is finally refitted using the information from all three sub-detectors.

Post-processing: A vertex-finder algorithm is used to identify primary vertices, followed by a dedicated algorithm for secondary vertex reconstruction.

A track is usually parametrized by its impact parameters in the transverse plane (d_o) and in the longitudinal direction (z_0), and by its momentum vector, defined by ϕ, θ and q/p, the charge associated with the track divided by its momentum. A detailed study of the ATLAS tracking performance and efficiency is presented in the ATLAS publication on proton-proton scattering [39].

4.3.2 Clustering

Calorimeter clusters form the input for both electron and jet reconstruction. Simply speaking, a calorimeter cluster is a collection of neighboring cells whose combined energy deposition reflects the energy loss of a particle in the calorimeter. Two different algorithms are relevant for the creation of calorimeter clusters, which is often called "clustering": sliding-window clustering and topological clustering. While the former algorithm creates clusters of fixed size, the latter one aims at the three-dimensional reconstruction of shower shapes which leads to clusters of variable size and form. These are called "TopoClusters". A detailed description of both algorithms can be found in [40].

4.3.3 Electron reconstruction

The reconstruction of electrons is based on sliding window clusters and tracks. The most general description of an electron candidate is a cluster in the electromagnetic calorimeter with an associated inner detector track pointing at it. Candidate clusters are required to have a transverse energy of at least 3 GeV, they have a size of 5×5 cells in the middle layer of the electromagnetic calorimeter. For track matching, a search window of 0.2×0.4 in $\Delta\eta \times \Delta\phi$ is constructed around the barycentre of the cluster. To be considered, tracks within that window are required to have at least three silicon hits, an $E/p < 10$ and must lie within a smaller window whose position depends on the track charge. The corners defining the windows for negative and positive tracks in terms of $\Delta\eta \times \Delta\phi$ with respect to the cluster center are placed at $(-0.05, -0.01), (0.05, 0.05)$ and $(-0.05, -0.05), (0.05, 0.1)$, respectively. The asymmetry accounts for the track curvature.

Starting from the cluster-track candidates the identification of electrons defines three quality levels with increasing jet rejection and decreasing efficiency:

Loose: Loose electron identification delivers a high efficiency but a low background rejection. Selection criteria are based on the hadronic leakage (ratio of E_T of the first sampling in the hadronic calorimeter to electromagnetic cluster E_T) and on the shower shape in the middle layer of the EM calorimeter.

Medium: On top of the criteria for loose electrons, the medium identification employs information from the first layer of the EM calorimeter, using its longitudinal segmentation to reject photons from neutral pion decays. It also requires additional track quality cuts and a tighter cluster-track match.

Tight: In addition to the medium criteria, tight electron identification requires a vertexing-layer hit to further reject photon conversions and a large fraction of high-threshold hits in the TRT. The latter are due to transition radiation and thus help to reject pions. Compared to medium, the cluster-track matching requirements are tightened even further.

For all quality levels there also exist more recent variants labeled `loose++, medium++, tight++`. These include additional quality cuts to mitigate the influence of pile-up in the 2011 and 2012 data which would otherwise strongly degrade the electron efficiency.

4.3.4 Jet reconstruction

Constructing jets is a complicated task: a good jet algorithm should be theoretically well-founded in the sense that it is insensitive to the details of the hadronization model and provide answers that are both infrared and collinear safe. The anti-k_T algorithm [41] used in ATLAS meets these requirements. It is a sequential combination algorithm which merges objects based on a distance measure defined in terms of their relative p_T. The objects that are used as inputs to the jet-finding algorithm are the TopoClusters discussed in Section 4.3.2. For each pair of objects i, j one defines:

$$\begin{aligned} d_{ij} &= \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{(\Delta R)^2_{ij}}{R^2} \\ d_{iB} &= k_{ti}^{2p} \\ (\Delta R)^2_{ij} &= (\phi_i - \phi_j)^2 + (y_i - y_j)^2 \end{aligned} \tag{4.5}$$

where k_{ti} is the transverse momentum of the object with respect to the beam direction and y is the rapidity, which can be approximated by the pseudo-rapidity η in the massless limit. R is a fixed distance measure which determines the size of the jets - in the analysis presented here, a value of $R = 0.4$ is used. The parameter p switches between the k_t ($p = 1$) and the anti-k_T ($p = -1$) algorithm. Both algorithms form jets by searching for the minimal entry in the list of all d_{ij} and d_{iB} and then adding the four-vectors of the objects i, j with minimal d_{ij}. If $d_{iB} < d_{ij}$ for all objects j, the object i is considered a complete jet.

The k_t algorithm is somewhat easier to interpret: for $p = 1$, d_{iB} is just the squared transverse momentum of an object with respect to the beam and d_{ij} is the (scaled) squared relative transverse momentum of the objects i, j. The drawback of the k_t algorithm is that it produces jets with irregular shapes, making the calibration of the jet energy difficult. The anti-k_T algorithm on the other hand produces well-defined jets which evolve to a conical form if they are isolated. It is now the default algorithm for jet-finding in ATLAS.

4.3.5 Muon reconstruction

Muons are probably the most important tagging objects to identify interesting event signatures. In ATLAS, several different methods for the identification of muons are used, relying on different subsystems of the detector. On top of that, there exist two families of algorithms in ATLAS which both implement the aforementioned methods: MuID and Staco. Usually only one algorithm is used for a given analysis, the other one is possibly employed for cross checks. The analysis presented in this thesis uses muons reconstructed with algorithms from the MuID family and additional "CaloMuons" (see below). I will briefly discuss the different approaches to muon identification in the following.

Moore and MuID standalone

The algorithms "Moore" and "MuID standalone" are used to reconstruct so-called standalone muons, which means that only information from the muon spectrometer is used. Moore creates muon spectrometer segments (two connected hits) and tracks (which consist of at least two segments) which are then extrapolated to the primary vertex by MuID standalone. Standalone muons can be constructed up to $|\eta| < 2.7$ but holes in the acceptance exist around $\eta = 0$ and $|\eta| = 1.2$.

MuID combined

So-called combined muons result from the combination of muon spectrometer and inner detector tracks. The MuID algorithm performs a global refit of the two tracks, combining vectors and covariance matrices of both measurements. Combined muons not only have a superior impact parameter resolution with respect to standalone muons, they also help to reduce the acceptance for secondary muons from pion or kaon decays in the calorimeter.

MuTagIMO and MuGirl

Using an (extrapolated) inner detector track as a seed, MuGirl searches for muon spectrometer segments and tracks. If a combination is found that yields a successful refit, a combined muon is created. If the refit is not successful, a tagged muon, where the track parameters are only taken from the inner detector track, is created. MuTagIMO aims to find muons with low transverse momentum which do not create hits in all muon spectrometer stations. A matching inner detector track is combined with a Moore segment to form a tagged muon.

CaloMuons

Although muons are minimal ionizing particles, they deposit some energy in the calorimeter along their flight trajectory. This is exploited to create so-called CaloMuons from an inner detector track and a matching extrapolated trajectory in the calorimeter. CaloMuons are usually used in the region $\eta \approx 0$

where no measurements from the muon spectrometer are available.

As for electrons, there exist different quality levels for muons (`loose, medium, tight`) which provide different efficiencies and background suppression factors. A recent measurement of the ATLAS muon reconstruction efficiency is presented in [42].

4.3.6 Missing transverse energy

Neutrinos provide an important fingerprint for many interesting physics processes. Unfortunately they are not directly detectable since they do hardly interact with the material in the detector. However, the conservation of the total transverse momentum in the proton-proton collisions at ATLAS allows for an indirect measurement of the neutrino momenta from the negative sum of all other momenta. The resulting quantity is called "missing transverse energy" and denoted as $E_{\mathrm{T}}^{\mathrm{miss}}$. Since a reliable measurement of $E_{\mathrm{T}}^{\mathrm{miss}}$ requires almost full detector coverage with regard to the solid angle, it is rather based on energy depositions in the calorimeter than on tracking information from the inner detector. The actual calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$ is somewhat involved due to many corrections (calibration for different objects, muons escaping the detector, dead material,...) that have to be taken into account. See [36] and [43] for details.

"Reality leaves a lot to the imagination."

John Lennon

CHAPTER 5

Selected aspects of event simulation and QCD

Simulation plays a crucial role on several levels in the analysis of LHC data. First and foremost, simulated collision events - often referred to as "Monte Carlo" for reasons that will become clear later - serve as theoretical references for most measurements and are thus vital for the understanding and interpretation of the data. In this chapter, I will introduce some key concepts of Monte Carlo event generation in the context of hadron collider physics and describe in short the samples that are used in the analysis at hand. A comprehensive and up-to-date review of Monte Carlo event generation is provided by the MCnet collaboration [44], a somewhat shorter and pedagogically motivated introduction is given in [45], from where much of the material presented here has been adapted.

5.1 Numerical integration with random numbers

The evaluation of phase-space integrals plays an important role in the calculation of (differential) cross-sections or detector count rates. Since the dimensionality of the phase-space increases rapidly with the number of final state particles, standard (non-random) numerical integration methods quickly break down in such situations due to convergence rates that depend on the dimensionality of the problem. Integration techniques which make use of random sampling - called Monte Carlo methods - on the other hand exhibit a stochastic convergence rate of $1/\sqrt{n}$ with n being the number of evaluations independent of the dimensionality of the problem. Random numbers and sampling tasks are thus an important ingredient in simulation problems: random phase-space points generated according to specific distributions are not only vital to perform integrations and make predictions for (differential) cross-sections, they also drive the simulation of individual events and observables. A discussion of specific sampling techniques and their usage in event generators is beyond the scope of this thesis. A general introduction to the topic is given in [46], a more detailed description of the algorithms employed in current Monte Carlo generators can be found in [44].

5.2 Essential QCD for the simulation of hadron-hadron collisions

To understand the simulation of events at hadron colliders, a discussion of specific aspects of QCD calculations performed in Monte Carlo generation is inevitable. I will thus sketch the key ideas of factorization, parton showers, matching and hadronization. The material is largely drawn from the review in [45].

5.2.1 Factorization

The difficulty in simulating events at hadron colliders like the LHC is that the simulation must address processes at various energy scales, including soft-QCD phenomena like the interactions of the partons inside the proton as well as perturbative QCD phenomena like hard scattering in parton collisions. Fortunately the two regimes can be separated by virtue of the factorization theorem [47], which allows to tackle the task by calculating the perturbative part on its own and then combining it with an almost universal non-perturbative part which is obtained from auxiliary measurements and fits to data. The idea of factorization can be understood from the following reasoning [45]: interactions among the partons inside the proton must be limited to momentum transfers below the confinement scale Λ, which means that they typically happen on timescales $\sim \frac{1}{\Lambda}$. Hard interactions between partons from different protons on the other hand typically involve momentum transfers of the order $Q \gg \Lambda$ and happen on timescales $\sim \frac{1}{Q} \ll \frac{1}{\Lambda}$. The structure inside the colliding hadrons thus appears to be static over the time of the hard interaction.

Adopting a particular factorization scheme called collinear factorization [48], the cross-section for a hadron-hadron interaction can be expressed as:

$$d\sigma_{h_1,h_2} = \sum_{i,j} \int_0^1 dx_i dx_j \sum_f \int d\Phi_f f_i^{h_1}(x_i, \mu_F^2) f_j^{h_2}(x_j, \mu_F^2) \frac{d\hat{\sigma}_{i,j\to f}}{dx_i dx_j d\Phi_f} \tag{5.1}$$

where the first sum runs over the partons i, j inside the colliding hadrons h_1, h_2, the second sum runs over the possible final states f and $d\Phi_f$ denotes the usual differential phase-space element. Note that the partonic cross-section $\hat{\sigma}$ can now be calculated in fixed-order perturbation theory. The parton distribution functions (PDFs) f_i, f_j on the other hand cannot be computed from first principles. They are typically constrained by measurements at lower energies - mainly deep inelastic scattering, see for example [49] - and then extrapolated by means of the appropriate renormalization group equation [50]. However, a useful event simulation must not only be able to predict cross-sections for partonic final states but provide estimations for observables that can be measured. This means that the simple factorization approach seems to fail in situations where the final state contains quarks and gluons as it is the case in typical LHC processes. The reason is that the experimental observables in such processes are based on hadrons in jets rather than partons. Perturbative QCD simply breaks down in the presence of scale hierarchies as they are encountered in jets. One inevitably runs into situations where a further resolution of substructure leads to divergent terms in the perturbative series. One must therefore introduce a fragmentation function (FF) D_f which describes the transition from a partonic final state to a hadronic observable. For a particular hadronic observable O, the differential cross-section then reads:

$$\frac{d\sigma}{dO} = \sum_{i,j} \int_0^1 dx_i dx_j \sum_f \int d\Phi_f f_i^{h_1}(x_i, \mu_F^2) f_j^{h_2}(x_j, \mu_F^2) \frac{d\hat{\sigma}_{i,j\to f}}{d\hat{O}} D_f(\hat{O} \to O, \mu_F^2) \tag{5.2}$$

where $\hat{O}$ denotes the same observable in the partonic final state. Just like the PDF, the fragmentation function D_f includes both resummations of perturbative corrections to all orders and non-perturbative corrections. Again this is a parametrized function that is independent of the hard-scattering process and can therefore be regarded as an ingredient to factorization. The distinction between the hard process on one hand and the PDF and FF on the other hand is made by the choice of an arbitrary scale called the factorization scale μ_F. A discussion of various choices for μ_F can be found in [51].

5.2.2 Fixed-order calculations

The cross-sections of hard partonic interactions can be calculated with a perturbative approach as an expansion in the strong coupling constant α_S. Calculations at the lowest or leading order (LO) can be carried out according to specific algorithms and have been automated in so-called matrix element generators like MadGraph [52]. Going beyond LO is typically difficult and has not been accomplished for every process. Such calculations at NLO, NNLO (next-to-leading order and so forth) are referred to as fixed-order calculations. If a NLO cross-section calculation for a given process is available the so-called K-factor may be defined as:

$$K_{\mathrm{NLO}} = \frac{\sigma_{\mathrm{NLO}}}{\sigma_{\mathrm{LO}}} \tag{5.3}$$

Higher order K-factors may be defined in analogy. These K-factors are typically used to calculate approximations of the differential cross-sections/distributions according to

$$d\sigma_{\mathrm{NLO}} = K_{\mathrm{NLO}} \cdot d\sigma_{\mathrm{LO}} \tag{5.4}$$

Kinematical corrections introduced at NLO are of course neglected in this approach, but since their calculation is difficult and the approximation is valid in the absence of additional real emissions it is a common compromise.

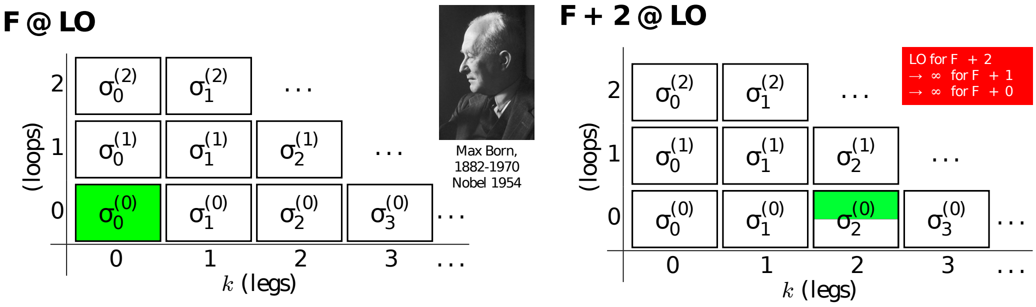


Figure 5.1: An illustration of the perturbative series calculated to fixed order. Left: Complete LO calculation for F. Right: LO calculation of $F + 2$ in the non-divergent region of the phase-space [45].

Let us introduce the notation convention from [45] for a short discussion of fixed-order calculations: a given final state is denoted as F, and k and l are the additional numbers of legs and loops with respect to the LO diagram for which $k = l = 0$. This is illustrated by the green square in the left pane in Figure 5.1. Very often, the LO diagram will not contain any loops and is thus called a "tree-level" diagram. With this convention, a LO estimate for the process $F + n$ jets can be obtained from F by setting $k = n$ and $l = 0$. A complete N^{n}LO result for F on the other hand must contain all contributions from combinations of k and l which fulfill $k + l \leq n$. The former (LO for $F + n$) can only be calculated in specific regions of the phase-space: when at least one of the n additional partons becomes soft or collinear, infrared singularities yield infinite answers in the phase-space integration. The half-shaded box in the right pane of Figure 5.1 illustrates this restriction. A proper N^{n}LO calculation for F (which by construction includes LO terms for $F+1, \ldots, F+n$) on the contrary is guaranteed to give a finite answer as stated by the KLN[1] theorem [53], [54]: the divergent terms which arise from additional legs are exactly canceled by divergent terms appearing in additional loops, rendering the complete answer finite

---

[1] Kinoshita-Lee-Nauenberg theorem: a complete perturbative expansion of the Standard Model must be infrared finite

up to order $k+l = n$. Unfortunately it is not straightforward to obtain a finite answer for such higher-order calculations in practice: correct canceling of the infinite contributions calls for a delicate arrangement of the divergent terms. Since we cannot restrict ourselves to the well-behaved (non-infrared) region of the phase-space when making QCD calculations for event simulation, efficient approximations of multi-leg/multi-jet scenarios - typically using resummation strategies - are indispensable. We will thus discuss the parton shower approach to leading-logarithmic resummation in Section 5.2.3.

5.2.3 Parton showers

A successful approach to approximating the infinite-order resummations enclosed in the FFs mentioned in Section 5.2.1 is called "parton shower". It allows for the description of the evolution of the partonic final state beyond the regime which can be covered with a fixed-order matrix element (ME) calculation. For the latter to give a reasonable answer for the process $F+n$ legs/jets, one must assert that all additional legs are sufficiently hard and well-separated (see Section 5.2.2) - a condition that is certainly not fulfilled in jet formation, where numerous emissions of soft and collinear partons take place. The matrix element may thus be interfaced with a parton shower algorithm which facilitates an approximate description of "infinite legs" and "infinite loops" [45].

The theoretical foundation is as follows: Given the squared amplitude $|\mathcal{M}_F|^2$ for the final state F it can be shown [45] that the squared amplitude for $F + 1$ leg factorizes according to

$$|\mathcal{M}_{F+1}|^2 = \mathrm{AF} \cdot |\mathcal{M}_F|^2 \tag{5.5}$$

in the phase-space region where additional radiation is enhanced. The power of this ansatz stems from the fact that it can be used recursively and thus can be used to obtain an approximation for the leading-order cross-section $\sigma^{(0)}_{F+n}$ for processes with an arbitrary number of legs n. The poles that would cause the tree-level matrix element for $F + n$ to diverge in the soft/collinear regime are now separated off and included in the so-called antenna-function AF (5.5). The expression that follows from repeated application of this procedure can in turn be rearranged in a way that allows for the cancellation of the infrared divergent terms against the divergent terms from loop corrections at so-called leading-logarithmic precision. See [45] for details.

The parton shower formalism realizes this idea in terms of an evolution operator which generates the all-order leading logarithmic corrections to the fixed-order matrix element by iteratively resolving more structure in the event. The evolution of successive $1 \to 2$ splittings is driven by an evolution parameter Q_E which may represent the parton virtuality, the relative transverse momentum of the parton or the angle between the two partons, depending on the algorithm. See [44] for a detailed review of the implementations in current MC generators. The probability to observe *no* split between two given scales $Q_{E1} > Q_{E2}$ is then defined by a so-called Sudakov factor. The evolution stops at the hadronization scale μ_{had} below which the current set of partons is passed to a hadronization algorithm (see Section 5.2.5).

A more technical discussion of parton shower algorithms, including an explanation of how random numbers are used to generate the evolution parameters employed for the determination of the splitting probabilities can be found in [55].

5.2.4 Matching matrix elements and parton showers

The parton shower approach to resummation is only well-defined in the regime of collinear and soft emissions. If we want to promote a LO×LL description of the process F (that is leading order ME for F plus leading logarithmic shower) to a description of $F + 1$ with an additional "hard leg", we need to include the LO matrix element for $F + 1$ as well and use $F + 1$ as a starting point for a new shower.

However, a problem arises: double-counting occurs for the LL terms in the inclusive $F + 1$ cross-section, since they were already included in the shower that started at F. The problem is illustrated in Figure 5.2: the left pane represents the LO cross-section calculation for F (green) with an additional LL shower (yellow). In order to improve the prediction for $F + 1$ from LL to LO, we try to add the LO ME for $F + 1$ and use this as a starting point for a new shower (middle pane). Since the LO ME for $F + 1$ is infrared divergent, we cannot cover the full phase-space in the calculation, as indicated by the half-shaded boxes. The problem of double-counting occurs if we try to combine both MEs which have a different number of legs (right pane).

Adding more legs (=more MEs) and showers leads to even more severe double-counting since one effectively sums up *inclusive* cross-sections where one should sum *exclusive* ones. Strategies to avoid the aforementioned problems are known as "matching".

Different approaches to matching exist, here I will just briefly mention the schemes used by several popular Monte Carlo generators. More details can be found in [55] or at a less technical level in [45].

Slicing: A popular approach used amongst others in the generator HERWIG [56]. The phase-space is divided into regions which are exclusively described by the ME or the shower. This is done by introducing a "matching scale" above which the shower is truncated and the higher-multiplicity MEs come into operation. Technical details on how to avoid discontinuities at the matching scale are important in practical implementations [57].

Subtraction: The difference of a LO calculation plus shower (LO×LL) and a NLO calculation plus shower (NLO×LL) is taken and used as a correction to the LO calculation. This approach provides NLO precision corrections to the Born level. An undesirable property of the subtraction approach is that negative event-weights occur which poses a problem in certain analysis tasks. Subtraction is used in MC@NLO [58].

Unitarity This approach aims at a point-wise correction of the shower prediction in phase-space by multiplicative factors which represent the ratio of the ME calculation to the shower at this point. The corrected shower associated to the process F will thus automatically reproduce the results for the $F + 1$ matrix element. PYTHIA [59] uses the original version of the algorithm, which only allows for one additional hard emission. An extension which includes NLO corrections to the Born level is used in POWHEG [60].

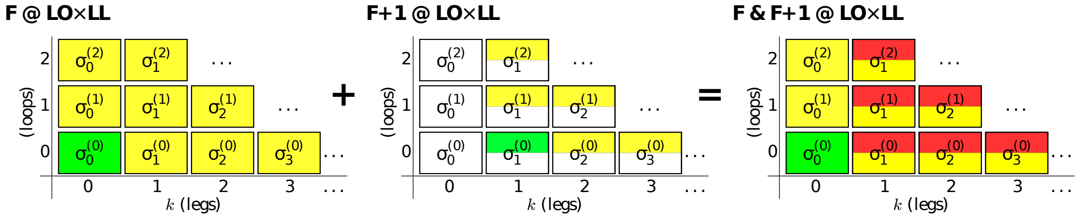


Figure 5.2: Illustration of the double-counting problems that occur when combining MEs and showers for different numbers of legs [45].

5.2.5 Hadronization

When the parton shower terminates at the hadronization scale μ_{had}, the current set of partons must be converted into a set of colorless hadrons. This process is called hadronization and is strictly non-perturbative which means one has to resort to phenomenological approaches. Two popular models for the parton-hadron transition exist: the cluster model [44] used by the generators HERWIG [56] and SHERPA [61] and the string model [62], which is built on “linear confinement” - the observation that the potential of the color-field between two color-charges grows linearly with the separation of the charges. The string model is used in PYTHIA [59] and will briefly be discussed below.

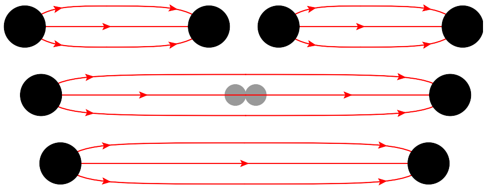


Figure 5.3: The string model of hadronization: a quark pair is created from fluctuations in the color field [45].

The string model of hadronization

Consider a color-connected quark-antiquark pair $q\bar{q}$ moving apart. According to the model of linear confinement, the potential between the quarks can be described by a one-dimensional string. As the quarks move apart and their kinetic energy decreases, the tension of the string increases as it is stretched. The string finally breaks up and its potential energy powers the creation of a real $q\bar{q}$-pair from fluctuations in the color field (see Figure 5.3). The process repeats until a collection of color-neutral hadrons is obtained. A more detailed discussion of the string model including aspects like the influence of gluons, baryon production or assignment of the quarks to hadron multiplets can be found in [45]. An in-depth explanation of the Lund model implemented in PYTHIA is given in [63].

5.3 Other aspects of hadron collisions that are relevant to simulation

Pile-up

Very often, the collision of two proton bunches at the LHC gives rise to more than one proton-proton interaction. Since hard scattering processes with high p_{T} are very unlikely in general, most of these interactions only yield low p_{T} particles/jets which are independent of a possible hard event within the same bunch collision. These processes are called “pile-up events”. The correct description of these additional interactions is important in order to understand their influence on experimental observables that are relevant to a given analysis task. Another common name for pile-up events is “minimum-bias events”. Strictly speaking, the latter term refers to *all* events that are selected with minimum trigger requirements (thus introducing minimal bias) to be as inclusive as possible. By definition this would also include hard processes, but since their share in minimum bias events is negligible, minimum-bias samples can be used to study the properties of pile-up. An account of the models used in Monte Carlo generators to describe pile-up/minimum-bias events can be found in [44].

Underlying event

The underlying event (UE) refers to additional interactions that involve partons from the same hadrons (“beam remnants”) which give rise to the hard process. Although these interactions mostly result in low-p_{T} scattering processes, multiple parton interactions (MPI) which yield observable energy depositions in the detector are common and need to be quantified. It has been observed that the activity in the UE is higher and less evenly distributed than in the average minimum-bias event. This is known as the “jet pedestal effect”. A detailed study of UE characteristics at the ATLAS experiment can be found in [64], a more general review of UE properties at the LHC is given in [65]. Finally, a description of the MPI model implemented in Pythia is found in [66].

5.4 Detector simulation

A comparison of simulated and measured events cannot take place without a proper description of the measurement process, in particular the response of the detector to all particles produced in a proton-proton collision. The description of the physical system after parton shower and hadronization must thus be fed into a realistic detector simulation. For this purpose, a detailed geometrical model of the ATLAS detector is constructed using Geant4 [67]. This model provides a simulation of all interactions with the detector, including energy loss and the generation of secondary particles. It is based on a macroscopic parametrization of particle-matter interactions which facilitates the simulation of particles traversing the detector without having to calculate all fundamental interactions on particle level. The simulation of a given event comes to a halt when all particles have left the detector volume or have been stopped in the material. After a subsequent digitization of the detector response in which the energy depositions are assigned to the fundamental detector components, the output format of the simulation is identical to that delivered by the ATLAS detector in response to a real proton-proton collision. The simulated output can thus be treated in the same way as real data and undergoes the same reconstruction process which then defines the objects that serve as an input to the analysis.

The complete simulation of the particle-detector interactions in Geant4 is very time-consuming, in particular for the calorimeter where most of the detector material is present. Since the analysis presented here relies on high statistics Monte Carlo samples, several processes were simulated with the ATLFAST-II [68] detector simulation, reducing the computation time by a factor of ~ 20. ATLFAST-II replaces the full calorimeter simulation of Geant4 with a parametrized version of the calorimeter response (FastCaloSim [68]) which was tuned by matching the results to the Geant4 response.

5.5 Monte Carlo samples used in the $VH, H \to b\bar{b}$ analysis

The signal and all relevant background processes for this analysis - except multijet - are modeled by Monte Carlo simulation, generated at the same center-of-mass energy and with the same amount of pile-up as the corresponding data samples. Whenever possible, a NLO generator was chosen, deviating choices are motivated below. All Monte Carlo samples undergo the same event reconstruction algorithms as the data, see Section 4.3 for details.

5.5.1 $VH \to Vb\bar{b}$ signal samples

The signal samples used in this analysis are generated with Pythia 8.165 [59], which is interfaced to Photos [69] to model QCD final state radiation and to Tauola [70] to model τ-lepton decays. The $V \to \tau\nu$ decays are considered to account for the small fraction of signal events which arises from $\tau \to \nu_\tau l \nu_l$ decays with $l = \mu, e$ in the WH channel. To facilitate a mass-sensitive analysis, separate samples are generated for different Higgs boson masses between 100 GeV and 150 GeV in intervals of 5 GeV. The use of the NLO generator Powheg was impeded due to a problem with low b-tagging efficiencies. As described in Section 3.4.1, the cross-section for the signal processes is calculated with

NNLO QCD corrections and NLO electroweak corrections. Uncertainties on the signal cross-section are treated as systematic errors, see Section 7.10.

5.5.2 Background samples

For the simulation of the background samples, several different Monte Carlo generators have been used. Nevertheless, the choice is consistent within different groups of background, i.e. for V + jets, di-boson and $t\bar{t}$/single-top.

The generation of the $W/Z+ \geq 1b/c/$light-jet samples calls for so-called multi-leg generators like Sherpa or Alpgen, with Sherpa being the preferred choice due to its superior matching algorithm. It is used in combination with the CT10 parton distribution functions. Specific heavy-flavor filters on generator level allow for increased statistics of the V + heavy flavor contribution which is particularly important for the analysis. Additional filters are also used to enrich the sample in high p_{T}^{V} events which tend to look more signal like than events with low p_{T}^{V}.

The WW, WZ and ZZ background samples are simulated with the Herwig generator, interfaced to the CTEQ6L1 PDFs and using the AUET2 tune [71] for the parton shower and hadronization model. The statistics of the ZZ sample is increased using an $E_{\mathrm{T}}^{\mathrm{miss}}$ filter. Attempts to use other generators were not successful.

The top-related backgrounds ($t\bar{t}$ and single-top) are generated with Powheg, which is preferred over the alternative choice MC@ANLO since it can be combined with the more recent Pythia8 as opposed to MC@NLO which can only be used with Herwig. The single-top t-channel forms an exception, since it can only be correctly described by AcerMC. Both generators rely on the “Perugia2011C” tune [71] and the CTEQ6L1 PDFs.

To avoid large uncertainties due to insufficient background statistics, the number of simulated background events was increased compared to earlier $VH \to Vb\bar{b}$ searches [72]. This guarantees that the statistical uncertainty is $< 5\%$ in the sensitive region of the phase-space. A summary of the Monte Carlo samples used in this analysis is presented in Table 5.1.

Process	Generator	$\sigma \times$ BR	N_{evt}
WH	Pythia 8.160		300000 (for each m_H)
ZH	Pythia 8.160		300000 (for each m_H)
$W \to l\nu$	Sherpa 1.4.1	10.97 nb	168M
$Z/\gamma^* \to ll, m_{ll} > 40\,\mathrm{GeV}$	Sherpa 1.4.1	1.24 nb	42M
WW	Herwig 6.510	55.43 pb	10M
$WZ, 66\,\mathrm{GeV} < m_{ll} < 116\,\mathrm{GeV}$	Herwig 6.510	22.69 pb	20M
$ZZ, 66\,\mathrm{GeV} < m_{ll} < 116\,\mathrm{GeV}$	Herwig 6.510	7.697 pb	7.5M
$t\bar{t}$	Powheg	238.06 pb	75M
single-top Wt-channel	Powheg	22.37 pb	20M
single-top s-channel	Powheg	5.61 pb	6M
single-top t-channel	AcerMC	87.76 pb	9M

Table 5.1: Overview of the Monte Carlo samples used for the analysis

“The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems ever conceived. It will be the stock analysts and petrochemical engineers who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.”

STEVEN PINKER

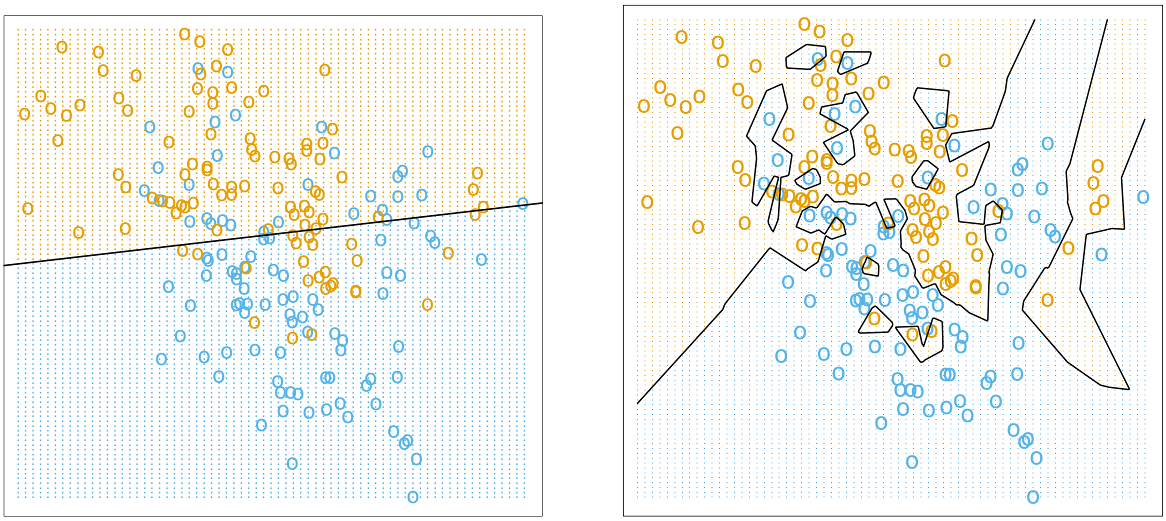
CHAPTER 6

Machine learning and boosted decision trees

An exceptionally demanding task in many high energy physics analyses is the identification of interesting events in a vastly background-dominated regime. The problem is most often tackled using a sequence of filters with increasing sophistication. At the LHC, where the event rates are particularly high, a first very coarse selection suitable for many subsequent tasks is performed online by dedicated trigger algorithms (see Section 4.2.4). The next level of filtering, which is analysis-specific, usually rejects any event whose signature is not in agreement with what is expected from the signal process under consideration. But even at that point, many rare processes are hardly detectable due to the overwhelming number of background events which either share the same signature as the signal process or have a non-vanishing probability to fake it if one or more final-state particles go undetected or are misidentified. The search for the Higgs boson in the $VH \rightarrow Vb\bar{b}$ channel is particularly challenging in this respect: a large number of different background processes (see Section 3.5) with cross-sections much higher than that of the signal process can produce almost indistinguishable final states. Two approaches can be followed in a situation like this - both of them rely on accurately simulated event samples for signal and background which underlines the importance of Monte Carlo simulation in the context of this analysis (see Section 5). In the first approach, one seeks to develop a very sophisticated set of selection criteria by hand. This requires the investigation and comparison of a large number of final state observables for signal and background both on theoretical level and on reconstruction level. Since it is very hard to visualize correlations between different observables beyond the two-dimensional case, one cannot guarantee that an optimal solution is found. The second approach relies on dedicated algorithms for statistical learning: a set of final state observables which convey as much information as possible is selected and their joint distribution is fed into a so-called “machine learning” algorithm which optimizes the selection with mathematical rigor.

In this thesis, the second approach - based on statistical learning - is adopted. However, before the learning algorithm is applied, a great deal of time is spent to make sure that the distributions presented to it correctly describe the data and that the observables chosen for the optimization indeed convey sufficient information to discriminate the signal candidates against the background. This is detailed in Sections 7.5 and 7.7.

In this chapter, I will discuss the foundations of statistical learning and describe a particularly appealing algorithm which is used in the analysis presented here: boosted decision trees (BDT).



(a)

(b)

Figure 6.1: The bias-variance trade-off in classification illustrated by two different algorithms. Left: a linear model with low variance and strong model bias. Right: $k-$nearest neighbors with $k = 1$ shows a high variance but no model bias [73].

6.1 Inference, classification and the bias-variance trade-off

Optimally separating signal and background events amounts to solving an inference task with respect to the event class, based on the underlying probability distributions for the final state observables. If the joint distribution of all final state observables was analytically known for both signal and background, one could easily construct an optimal hypothesis test from the likelihood-ratio of the two hypotheses (signal vs. background) as described by the Neyman-Pearson-lemma [74].

The problem that arises in high energy physics analyses however, is that the probability distributions of the final state observables are not known analytically - at least not after reconstruction. One thus has to extract the necessary information from the simulated Monte Carlo samples. This can be easily done in the limit of infinite statistics: a multi-dimensional histogram with sufficiently fine binning is just as good as the underlying probability density itself. Unfortunately, this approach will hardly work in practice since the available number of simulated events will usually not suffice to populate a histogram of the final state observables with more than two dimensions: the volume of the so-called feature space increases exponentially with the number of observables considered. This is known as the "curse of dimensionality": as the dimensionality of the problem increases, the available data become sparse.

Machine learning based classification seeks to mitigate this problem by imposing additional structure in the feature-space, often making assumptions about the underlying model and its parametrization: so-called training data, for which the correct class label is known, is inspected by the algorithm and the parameters of the assumed model are adjusted until optimal separation for the chosen model is achieved on the training data. This approach is known as "supervised learning" or "learning by example". After the training step is finished, the performance of the classifier is measured on a second, independent data set - the test data - whose class labels are also known. In high energy physics, both the training and the test data usually consist of Monte Carlo samples.

There exists a multitude of different algorithms for supervised learning problems, the main difference between them being how they compromise on bias versus variance. A very rigid approach, like a linear model, is of course robust, i.e. has small variance, but will introduce a strong bias if the true model is not linear. A very flexible classifier on the other hand, like k-nearest neighbors, will hardly introduce any bias but may adopt "too well" to the training data - a behavior known as "overtraining", similar in type to over-fitting. This is illustrated in Figure 6.1: a linear classifier and a $k-$nearest neighbor classifier with $k = 1$ are applied to a two-dimensional problem which is defined by the differently colored training data points shown in the plot. The black line indicates the so-called decision boundary, the contour which provides optimal separation between the two classes for a given classification model. For a more general discussion of machine learning and statistical inference, I refer the reader to one of the excellent introductory textbooks [73], [75], [76] available.

Two learning algorithms are particularly popular in high energy physics: artificial neural networks and, more recently, boosted decision trees. Both algorithms feature so-called "unparametrized" models, although there are of course parameters, albeit a very large number of them. These algorithms are very flexible and hardly impose any restrictions on the underlying model which makes them suitable for a large number of problems.

Artificial neural networks were among the first supervised learning algorithms employed in high energy physics and are still used for a wide range of different tasks, e.g in the b-tagging algorithm used for this analysis. Their strength lies in the flexibility to describe very complex classification scenarios, but in many cases they demand either a fair amount of manual tuning or require a lot of computing time to achieve optimal performance. A thorough discussion of neural networks is beyond the scope of this thesis, but a pedagogical introduction which is motivated from a high energy physics point of view can be found in [77].

Boosted decision trees (BDT) on the other hand are a more recent technique, making its first notable appearance in high energy physics at the MiniBooNE experiment in 2004 [78]. I will discuss the algorithm in the following section and motivate why it was chosen for the analysis at hand.

6.2 Boosted decision trees

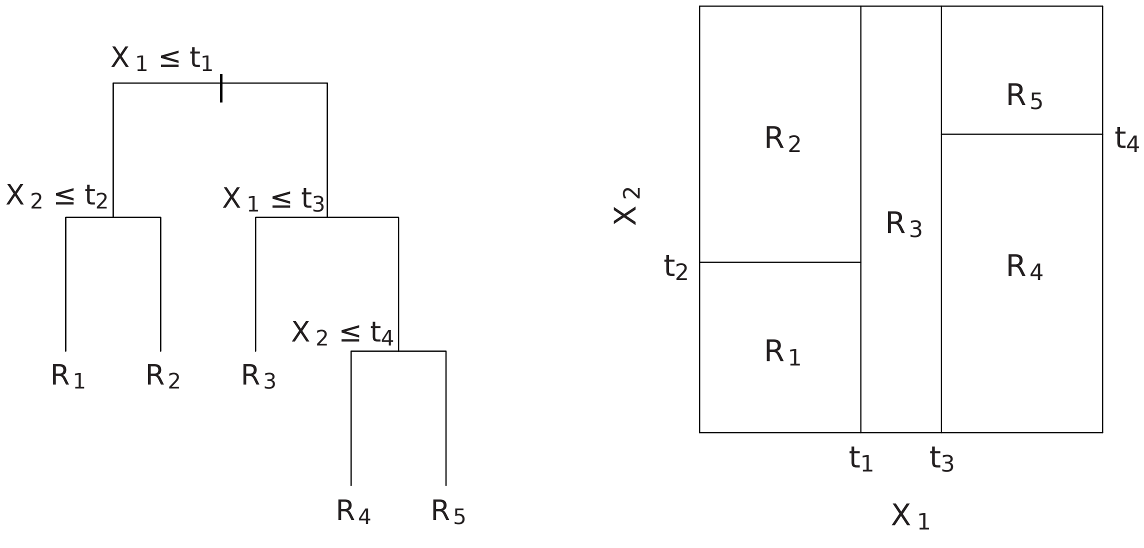
The BDT algorithm is an amalgam of two simple concepts which yield an exceptionally powerful tool for statistical learning when combined - decision trees and boosting. A decision tree comes as a natural extension of classification by binary cuts: consider a list of criteria (cuts) that may be checked to decide whether an event is signal- or background-like. A large percentage of the events in a typical analysis will neither be exactly background-like nor exactly signal-like but share properties of both classes. A decision tree allows to inspect combinations of several criteria instead of discarding an event which fails one of the cuts right away. This facilitates a more fine-grained selection process. I will discuss the construction of decision trees below.

Boosting is one of the most powerful concepts for statistical learning available. It relies on the principle of "the wisdom of crowds", forming a powerful "committee" from so-called "weak-learners". This means that a very simple algorithm - like a decision tree - can be used as a basis to construct a very sophisticated classifier.

6.2.1 Growing decision trees

A decision tree partitions the feature-space recursively into rectangular regions as illustrated in Figure 6.2. It is constructed by consecutively inspecting the available observables and gradually splitting the training data sample according to a given figure of merit. Assuming that we have two training samples, one with simulated signal events and one with simulated background events, which are characterized by a set of observables $\vec{x} = (x_1, \ldots, x_n)$, the algorithm works as follows:

1. Start from the root-node of the tree, which represents the entire training data.
2. If the node currently under consideration satisfies any stopping criterion (see below), make it a terminal node ("leaf") and exit.
3. Sort all training events according to each variable in $\vec{x}$.
4. Iterate through all variables $\vec{x}$ and for each variable x_i find the cut value t_i which would minimize the impurity in the two sub-samples that would be created if the cut was applied (see details below).
5. Select the pair (x_i, t_i) which minimizes the impurity and split the node - and thus the training sample - according to (x_i, t_i), producing two new nodes: one containing all events that pass the cut (x_i, t_i) and the other one containing all events that fail it.
6. Apply recursively on each node starting from step (2).



(a) (b)

Figure 6.2: Illustration of the recursive partition of the feature space induced by a decision tree [73].

Typical stopping conditions in step (2) include:

Leaf size: A node is declared as final if further splitting would reduce the number of events in the new nodes below a fixed minimum N_{min}. The number N_{min} can be chosen as an absolute value or as a percentage of the full training sample, ensuring that the impurity measure is statistically significant for each leaf.

Perfect separation: A node is declared final if it contains only events from one class.

Fixed tree depth: The depth of the tree may be fixed by the user to avoid complex decision boundaries.

Insufficient improvement: A node is declared final if no split yields significant improvement. This criterion is rarely used because even a weak split may allow for a very powerful split of one of the child nodes later.

Each leaf in the tree is finally assigned to one of the classes (signal or background) according to the majority of training events associated with it.

The splitting criterion which is used to find the optimal cut in step (3) is based on a measure of sample impurity. Such an impurity function must fulfill the following requirements:

- The impurity should be minimal for nodes which either contain only signal or only background events.
- The impurity should be maximal for an equal mix of signal and background.
- The impurity function should be symmetric in signal and background.
- The impurity function should be strictly concave, i.e. purer nodes are preferred.

If we denote the - possibly weighted - number of signal (background) events in a given node by s (b), we can define the signal purity p_s:

$$p_s = \frac{s}{s+b} \tag{6.1}$$

The background purity is then simply given by $p_b = \frac{b}{s+b} = 1 - p_s$. From this, one can easily construct several impurity functions which satisfy the requirements listed above:

Misclassification error: $1 - \max(p_s, 1 - p_s)$

Cross entropy [79]: $-\sum_{i=s,b} p_i \log p_i$

Gini index [80]: $2p_s(1 - p_s)$

The Gini index is the most popular impurity function in decision tree implementations since it is differentiable and easy to calculate. All three functions are shown in Figure 6.3.

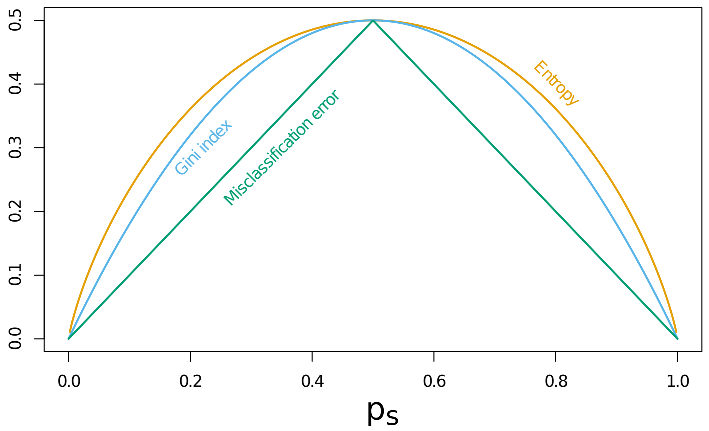


Figure 6.3: Common impurity functions used in decision tree implementations [73].

The application of a fully trained decision tree to data is straightforward: each event will gradually be passed down the tree until it reaches a leaf:

1. Apply the selection criterion t that was associated with the root node during training.
2. Continue to the pass- or fail-node according to the outcome of the test.
3. Apply the criterion associated with the new node and move to the next pass- or fail-node. Repeat until a leaf is reached.
4. The event is classified according to the class associated with the leaf in which it ends up. One can either assign a binary output (signal/background) at this stage or return the training purity of the leaf as a measure of the reliability of the classification.

Decision trees have a lot of desirable properties concerning robustness and the curse of dimensionality:

- Adding more variables, i.e. adding dimensions to the feature-space, will not degrade the performance of a decision tree since for each split all variables are considered in succession, not at the same time.
- Duplicate variables will simply lead to the same ordering and thus do not affect the tree at all.
- Non-discriminating variables will never be chosen to induce a split and thus do not add any noise.
- The training will result in the same decision tree regardless of the event order or any monotonous variable transformation applied.

These features facilitate the usage of $\gg 10$ variables in decision trees and even make them suitable for the identification of powerful variables which will be used for node-splitting more often than weak ones. A major drawback of decision trees on the other hand is that the recursive splits of the feature-space quickly reduce the statistics in the resulting regions, increasing the variance of the class estimate. A small change in the training sample can thus lead to a different split and thereby change the entire structure of all subsequent nodes. These instabilities prohibit the use of very deep trees, which are prone to overtraining. Shallow trees on the other hand do not provide very sophisticated rules for classification and may thus not produce the optimal decision boundary for a given problem - they are “weak learners”. Their robustness and simplicity however makes shallow decision trees the ideal candidates for boosting which will be discussed in the next section.

6.2.2 Boosting

Boosting is a committee-based learning approach which aims to form a powerful meta-classifier from a collection of so-called “weak learners”, i.e. individual classifiers which only slightly exceed the random-guessing error-rate of 0.5. Rather than selecting the base classifiers for reasons of performance, one can thus select them for reasons like robustness and immunity to the curse of dimensionality. As discussed previously, this makes shallow decision trees an ideal candidate for boosting. The following discussion of different boosting algorithms follows the presentation in [73].

Adaptive boosting

The most popular boosting algorithm is called “AdaBoost” (short for adaptive boosting), proposed by Freund and Schapire in 1995 [81]. AdaBoost produces a sequence of modified (reweighted) versions of the original training data and repeatedly applies the base classifier to it, thus creating a collection of weak classifiers (trees) $T_1, \ldots, T_M$, see Figure 6.4.

Let us assume that the class label in our training data is encoded in a binary variable $y \in \{-1, 1\}$, where $y = 1$ ($y = -1$) denotes a signal (background) event. Let us further assume that each single tree T_m returns a discrete response $T_m(x) \in \{-1, 1\}$ when applied to an event with observables x. The combined class prediction for that event is then given by

$$T(x) = \mathrm{sign}\left(\sum_{m=1}^{M} \alpha_m T_m(x)\right) \tag{6.2}$$

AdaBoost thus takes a weighted majority vote among the trees T_m to predict the class for an unknown event. The weights α_m are calculated by the algorithm and increase the influence of the more accurate classifiers.

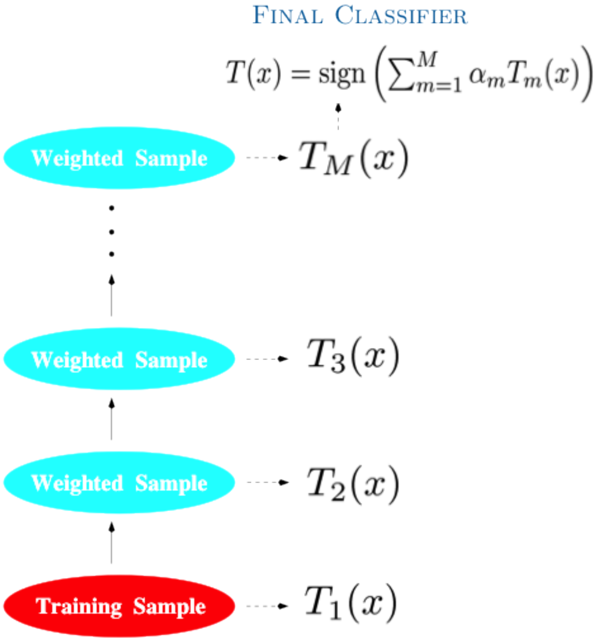


Figure 6.4: An illustration of the AdaBoost algorithm: a sequence of weak classifiers is trained on reweighted versions of the original training data and finally combined into a meta-classifier [73].

The reweighting prescription for the training data at iteration m involves a modification of individual event weights which depends on the class estimate of the tree T_{m-1}: events correctly classified by T_{m-1} have their weights decreased, whereas misclassified events have their weights increased. The next tree T_m will thus be forced to pay more attention to previously misclassified events, leading to a different tree structure. The complete AdaBoost algorithms reads as follows:

1. Initialize the weights of all training events to $w_i = 1/N$, where N is the total number of events.
2. For each tree T_m
    a) train the tree on the training sample with the current weights.
    b) compute the error rate $\epsilon_m = \frac{\sum_{i=1}^{N} w_i \cdot I(y_i \neq T_m(x_i))}{\sum_{i=1}^{N} w_i}$, where $I(y_i \neq T_m(x_i))$ evaluates to 1 only if the argument is true and to 0 otherwise
    c) compute the tree weight $\alpha_m = \log \frac{1-\epsilon_m}{\epsilon_m}$

   d) update the event weights according to $w_i \leftarrow w_i \cdot \exp(\alpha_m \cdot I(y_i \neq T_m(x_i)))$

3. Combine all trees: $T(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m T_m(x)\right)$

A great advantage of AdaBoost compared to other classification algorithms is that it needs very little tuning to reach optimal performance: choosing small trees as the base learners, it benefits from all the desirable features of decision trees mentioned previously. Basically one just needs to choose an appropriate number of iterations, which is typically between 100 and 1000 - another argument why decision trees, which can be constructed very fast, are the optimal choice for boosting.

Taking a more technical point of view, AdaBoost can be understood as a basis function expansion where the basis functions are given by the individual classifiers. To understand this in more detail, let us rewrite the final estimator $T(x)$ by making the dependence on the tree parameters γ_m (which denotes the collection of all cuts (x_i, t_i) chosen for node-splitting in tree T_m) explicit:

$$T(x) = \sum_{m=1}^{M} \alpha_m T_m(x, \gamma_m) \tag{6.3}$$

The parameters $(\alpha_m, \gamma_m)_{m=1,\ldots,M}$ of additive models like this are typically determined by fitting the model to the training data $(x_1, \ldots, x_N)$ using a so-called loss function L which measures the fit-quality:

$$(\alpha_m, \gamma_m)^{optimal}_{m=1,\ldots,M} = \arg \min_{\{\alpha_m, \gamma_m\}_1^M} \sum_{i=1}^{N} L\left(y_i, \sum_{m=1}^{M} \alpha_m T_m(x_i, \gamma_m)\right) \tag{6.4}$$

For large M, this is a formidably complicated optimization problem - how does AdaBoost efficiently find a solution?

The answer is that AdaBoost in fact solves a much simpler problem where only one tree is fitted at a time - trees which have already been fitted are not modified anymore. This approach is known as *forward stage-wise additive modeling* and often provides a reasonable approximation to the solution of Eqn. 6.4. Formulated as an algorithm:

1. Start with $T^{(0)}(x) = 0$.
2. For $m = 1, \ldots, M$:
   a) Determine the optimal parameters for T_m:

$$(\alpha_m, \gamma_m) = \arg \min_{\alpha, \gamma} \sum_{i=1}^{N} L(y_i, T^{(m-1)}(x_i) + \alpha T_m(x_i, \gamma)) \tag{6.5}$$

   b) Set $T^{(m)}(x) = T^{(m-1)}(x) + \alpha_m T_m(x, \gamma_m)$.

An approach like this will typically need many terms (trees) T_m to reach an approximate solution, but it is computationally much less expensive than trying to determine all parameters at once by global optimization.

It can be shown that the AdaBoost algorithm is fully equivalent to forward stage-wise additive modeling with an exponential loss function:

$$L(y, f(x)) = \exp\left(-y \cdot f(x)\right) \tag{6.6}$$

This is a convenient choice since it allows the formulation of the algorithm in terms of the reweighting procedure introduced earlier without the need for explicit minimization of L. However, there are situations when exponential loss will not provide the optimal solution because its performance may be degraded by outliers or false class labels in the training data. To understand this, let us introduce the so-called "margin" $y \cdot f(x)$ which measures the amount of misclassification on the training data: for training data with binary class labels $y \in \{-1, 1\}$ and a classifier function $f(x)$ with classification rule $T(x) = \mathrm{sign}(f(x))$ like the tree committee obtained from AdaBoost, a positive (negative) margin $y_i \cdot f(x_i) > 0$ ($y_i \cdot f(x_i) < 0$) implies that the event has been classified correctly (incorrectly). Figure 6.5 shows various loss measures as a function of the margin. For a discussion of classification, it is illustrative to inspect exponential loss and binomial deviance which both can be understood as approximations to the non-differentiable misclassification loss $L(y, f(x)) = I(y \cdot f(x) < 0)$ which does not penalize correct classification at all and gives constant penalty for misclassification.

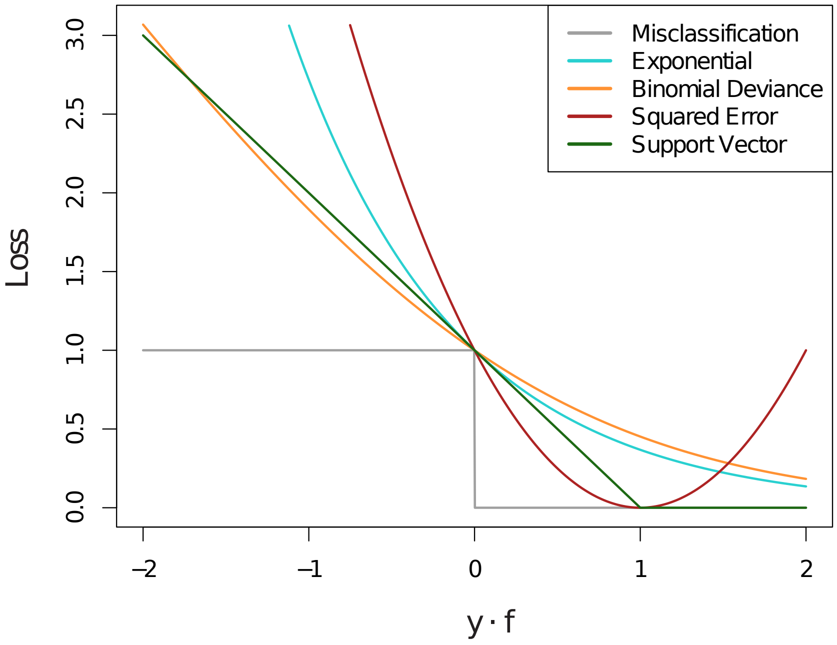


Figure 6.5: Different loss measures for classification as a function of the margin $y \cdot f(x)$ [73].

The binomial deviance loss function is given by

$$L(y, f(x)) = \log(1 + \exp(-2y \cdot f(x))) \tag{6.7}$$

Like exponential loss it yields an increasing penalty for increasingly negative margins, but while exponential loss grows exponentially, binomial deviance is asymptotically linear for $y \cdot f(x) < 0$. This means that an algorithm which uses exponential loss, e.g. AdaBoost, will be strongly influenced by training events with large negative margin, leading to a possible degradation of the performance in noisy settings. To be fair, one must state that situations like this are rare in high energy physics analyses and that AdaBoost is thus an appropriate choice for most tasks. However, since one can do better from a theoretical point of view, it makes sense to investigate alternative approaches to provide optimal performance for the analysis presented here. The major disadvantage of loss functions other than exponential loss is that they do not allow to express boosting as a simple reweighting prescription like AdaBoost. One thus has to resort to an approach inspired by numerical optimization: gradient boosting [82], [83].

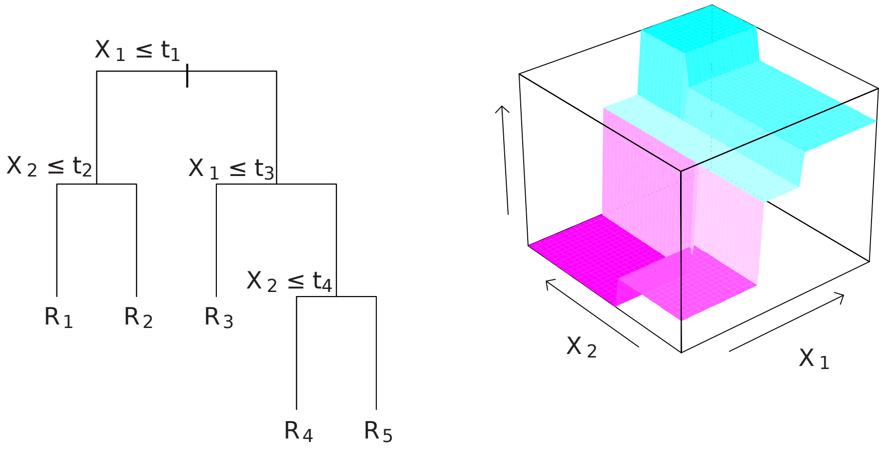


Figure 6.6: Illustration of a step function that a regression tree can represent [73].

Gradient boosting - regression trees for classification

Although the motivation for introducing a second boosting algorithm is to use an optimal loss function for classification, a discussion of gradient boosting requires a short introduction to regression trees. Regression trees are very similar to decision trees, but instead of event classification they aim at the approximation of an unknown target function by a step function which is represented by the tree: starting with a training set composed of observation/target-pairs (x_i, y_i) the feature space is recursively partitioned and each terminal region R_j is assigned a response value which is usually just the mean of the target values for all observations that fall into that region $c_j = \text{average}(y_i | x_i \in R_j)$. This is illustrated in Figure 6.6. The splitting criterion used during tree construction is that the sum of squares $\sum (y_i - f(x_i))^2$ in the resulting regions should be minimal. This is a computationally difficult task, but efficient approximations exist.

Returning to our problem of suitable loss functions for classification, let us drop the requirement that the predictor function $f(x)$ must be a sum of trees for a moment. We can then write the total loss on the training data as

$$L(f) = \sum_{i=1}^{N} L(y_i, f(x_i)) \tag{6.8}$$

and treat the problem as an exercise in numerical optimization: defining $\mathbf{f} = \{f(x_1), \ldots, f(x_N)\} \in \mathbb{R}^N$ as the parameter vector that contains the values of f evaluated at all training data points, one now seeks for

$$\hat{f} = \arg\min_{\mathbf{f}} (L(\mathbf{f})) \tag{6.9}$$

A minimization problem like this can be solved by the method of steepest decent. Starting with an initial guess $\mathbf{f}_0$ one successively constructs $\mathbf{f}_m$ from $\mathbf{f}_{m-1}$ by calculating the gradient $\mathbf{g}_m$ with components

$$g_{mi} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x_i) = f_{m-1}(x_i)} \tag{6.10}$$

and setting

$$\mathbf{f}_m = \mathbf{f}_{m-1} - \rho_m \mathbf{g}_m \tag{6.11}$$

The step length ρ_m can be determined from

$$\rho_m = \arg\min_{\rho}(L(\mathbf{f}_{m-1} - \rho \mathbf{g}_m)) \tag{6.12}$$

At first sight, this approach seems useless: perfect classification of the training data could easily be obtained by setting $\mathbf{f} = \{f(x_1), \ldots, f(x_N)\} = \{y_1, \ldots, y_N\}$, but this does of course not help in the construction of a function $f(x)$ which also delivers useful prediction for new data not presented in training. However, the idea of learning by gradient-descent is still useful when combined with the constraint that $f(x)$ should be a sum of trees. For a sum of trees, the components $\{f(x_1), \ldots, f(x_N)\}$ are not longer independent and the following idea emerges: the complicated step of tree parameter determination (Eqn. 6.5) in the forward stage-wise additive modeling algorithm discussed earlier is replaced by a gradient descent step. First, the gradient itself can be calculated easily for any differentiable loss function, but second it must be approximated by a tree. This is where the regression trees come into play - instead of constructing a classification tree at each iteration, a regression tree is fit to the N components of the gradient which are now the regression targets. Its parameters are simply given by:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{N} L^2(-g_{mi} - T(x_i, \gamma)) \tag{6.13}$$

The complete algorithm for an arbitrary differentiable loss function works as follows:

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$. This is just a regression tree which consists of a single root node, returning a constant value for the entire sample.
2. For $m = 1, \ldots, M$, where M is the total number of trees:
    a) Compute the components of the gradient $g_{mi} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x_i) = f_{m-1}(x_i)}$
    b) Fit a regression tree using $(x_i, g_{mi})_{i=1,\ldots,N}$ as a training sample. This will result in J_m terminal regions R_{mj}.
    c) Compute the optimal output for each region of the new tree according to the chosen loss function $c_{mj} = \arg\min_c \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c)$
    d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} c_{mj} I(x \in R_{mj})$

Various modifications of the original algorithm exist, introducing for example additional sub-sampling to increase the statistical robustness of the tree-growing procedure. See [83] for more details.

Both boosting algorithms discussed in this chapter are used for the $VH \to Vb\bar{b}$ search presented in this thesis, providing a valuable cross-check for the optimization of the classification performance. The implementations are provided within the framework of TMVA [84].

"You can know the name of a bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird... So let's look at the bird and see what it's doing – that's what counts."

RICHARD P. FEYNMAN

CHAPTER 7

A BDT-driven search for $H \to b\bar{b}$ in associated production with a vector boson

7.1 Outline of the analysis strategy

The Higgs boson search described in this thesis is designed to reach maximum sensitivity while at the same time providing a solid understanding of all features observed in the data. Several key ideas which will later be discussed in more detail shall be outlined here:

- The definition and selection of all physics objects in the analysis is in agreement with the convention adopted for earlier VH searches in ATLAS, in particular [85], to facilitate an easy comparison of results.
- A physically motivated description of the VH system by a minimal yet exhaustive set of Lorentz invariant quantities and angles is adopted.
- A highly optimized machine learning algorithm based on the statistical paradigm of boosting is used to reach optimal discrimination between signal and background.
- The pre-selection of candidate events is as wide-meshed as possible to exploit the full statistical power of the data. At the same time the selection must be able to reduce the phase-space for the analysis to a region where the selected data is understood to a very high level of detail and where MC simulation and data match as closely as possible. A strong suppression of the multijet background for which no appropriate simulation exists should be achieved.
- Whenever possible, data-driven estimates of background rates and shapes are used. A consistent choice of Monte Carlo generators is made across similar backgrounds, see Section 5.5 for details.

Bearing the aforementioned ideas in mind, the analysis proceeds as follows:

1. Physical objects (leptons, jets) which are of interest to the analysis are identified according to the criteria described in Section 7.3.
2. A selection of candidate events which complies with the requirements outlined above is performed. This is described in Section 7.7. Further categorization allows for the definition of several control regions which are used to constrain the most important backgrounds (see Section 7.8).

3. A multivariate classification algorithm based on a Lorentz invariant description of the VH system is applied to the events in the signal region. For details see Section 7.5 and Section 7.6.

4. The output distribution of the multivariate classification algorithm in the signal region and additional kinematical distributions from the control regions are used to determine an upper limit on the Higgs production rate in the $VH \to Vb\bar{b}$ channel. A detailed description of the fit procedure can be found in Chapter 8.

7.2 Datasets

The analysis presented in this thesis is based on data recorded by the ATLAS experiment during the 2012 run of the LHC which featured a center-of-mass energy of $\sqrt{s} = 8\,\text{TeV}$. Only data which passed the criteria defined by the official ATLAS Good Run List were used. This ensures that the detector systems which are crucial for a correct measurement of the objects defined for the analysis (electrons, muons, jets, $E_{\mathrm{T}}^{\mathrm{miss}}$) were working as expected and that the associated triggers were operational during data taking. The amount of data used for the analysis corresponds to an integrated luminosity of $20.3\,\text{fb}^{-1}$ [86].

7.3 Object definition, selection and calibration

To facilitate a fair assessment of the performance of the analysis presented here, the object definitions and quality requirements are chosen to be in agreement with those used in the latest VH search [85] published by the ATLAS collaboration. This includes the choice of triggers as well as the cuts imposed on jets and leptons that are taken into account in the analysis. In addition to the standard kinematical quantities, additional variables used in the object selection are defined:

- $\mathtt{iso}^{track}$: sum of the transverse momenta of all tracks in a cone of radius 0.2 around the lepton divided by the p_{T} of the lepton
- $\mathtt{iso}^{calo}$: sum of energy deposits in the calorimeter in a cone of radius 0.3 around the lepton divided by the lepton E_{T}
- d_0: transverse impact parameter with respect to the primary vertex
- z_0: longitudinal impact parameter with respect to the primary vertex
- JVF: sum of the transverse momenta of all tracks that are associated to a jet and that originate from the primary vertex divided by the sum of transverse momenta of all tracks associated to that jet

7.3.1 Trigger selection for the 2012 analysis

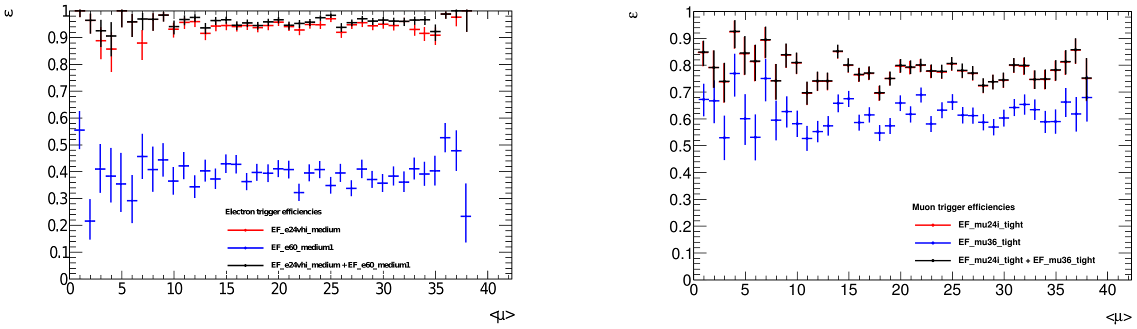
The analysis is driven by lepton triggers and for both electrons and muons the lowest unprescaled triggers are used. These are `e24vhi_medium1` (electrons) and `mu24i_tight` (muons) which both have a p_{T}-threshold of 24 GeV and include a track isolation requirement of $\mathtt{iso}^{track} < 0.1$. Additional triggers with higher p_{T}-thresholds (`e60_medium1` at 60 GeV for electrons and `mu36_tight` at 36 GeV for muons) are used to compensate for the inefficiency introduced by this isolation requirement. In the ZH channel, where two charged leptons are required, the aforementioned triggers are supplemented with the di-lepton triggers `2mu13` and `2e12Tvh_loose1`, respectively. The trigger efficiencies obtained with this trigger combination are shown in Figure 7.1 as a function of the mean number of interactions per bunch crossing $\langle \mu \rangle$. A fit using a constant function yields the efficiencies listed in Table 7.1.

	Trigger efficiencies	
	electron	muon
WH	0.967 ± 0.003	0.780 ± 0.006
ZH	0.998 ± 0.001	0.872 ± 0.005

Table 7.1: Trigger efficiencies for the trigger combination used in this analysis.

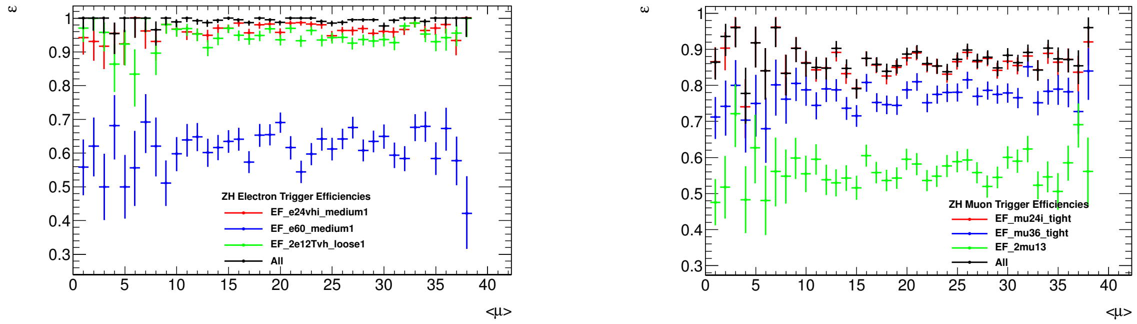
Earlier studies [87] have demonstrated that a part of the efficiency which is lost due to limited coverage of the L1 trigger chambers (see Section 4.2.3) in the muon channel can be recovered by using an additional $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger. While employed for the latest *VH* search [85] published by the ATLAS collaboration, this approach is not used in the analysis presented here due to missing trigger information in the available data. The overall loss of sensitivity should be less than 2%, see [88] for details.

Trigger matching is used for both lepton types to ensure that the offline lepton used in later analysis steps is indeed the lepton that fired the trigger and to avoid any bias in the estimation of the trigger efficiency.



(a) *WH* channel, electron

(b) *WH* channel, muon



(c) *ZH* channel, electrons

(d) *ZH* channel, muons

Figure 7.1: Trigger efficiencies estimated from Monte Carlo for different values of $\langle \mu \rangle$.

7.3.2 Lepton selection

Electrons

Two different quality levels for electrons are defined for the analysis: loose electrons, which are primarily used to reject unwanted events, and tight (or signal) electrons, which are used to identify the possible decay of a vector-boson which is part of the VH system. The labels “loose” and “tight” used in the analysis do *not* coincide with the quality definitions for electron identification discussed in Section 4.3.3: loose electrons are required to fulfill:

- $E_{\mathrm{T}} > 10\,\mathrm{GeV}$
- $|\eta| < 2.47$
- $\mathtt{iso}^{track} < 0.1$
- $d_0 < 0.1\,\mathrm{mm}$
- pass ATLAS `LoosePP` electron quality requirement (see Section 4.3.3 for a definition of the electron quality labels)

Tight electrons have to pass all requirements that apply to loose electrons and in addition fulfill:

- $E_{\mathrm{T}} > 25\,\mathrm{GeV}$
- pass ATLAS `MediumPP` electron quality requirement
- $\mathtt{iso}^{calo} < 0.07$ (only WH)
- $\mathtt{iso}^{track} < 0.04$ (only WH)
- pass ATLAS `TightPP` electron quality requirement (only WH)

Muons

As for electrons, two different quality levels (loose/tight) for muons are defined for veto and signal selection purposes respectively. The selection criteria for muons are similar to those for electrons with some subtle differences to improve the acceptance in regions with limited muon chamber coverage. Loose muons are defined by:

- $E_{\mathrm{T}} > 10\,\mathrm{GeV}$
- $|\eta| < 2.7$
- for $|\eta| < 0.1$: the muon is identified in the calorimeter (CaloMuon) and has an associated inner detector track
- for $|\eta| < 2.5$: the muon is reconstructed in both the muon spectrometer and the inner detector (combined or tagged muon)
- for $2.5 < |\eta| < 2.7$: the muon is reconstructed in the muon spectrometer (standalone muon)
- if muon has an associated ID track: $\mathtt{iso}^{track} < 0.1$

- if muon has an associated ID track: $d_0 < 0.1\,\mathrm{mm}$
- if muon has an associated ID track: $z_0 < 10\,\mathrm{mm}$
- if not a CaloMuon: pass ATLAS tight `MuID` criteria

Tight muons fulfill all of the above and in addition:

- $E_\mathrm{T} > 25\,\mathrm{GeV}$
- $|\eta| < 2.5$
- must be a combined/tagged muon
- $\mathtt{iso}^{calo} < 0.07$ (only WH)
- $\mathtt{iso}^{track} < 0.04$ (only WH)

The enumeration of lepton quality criteria shows that the definition of "tight" leptons is stricter for the WH than for the ZH channel. This is due to the increased level of multijet background in the one-lepton analysis which has to be suppressed by applying tight isolation cuts for the leptons.

Energy and resolution corrections for leptons

For both electrons and muons, energy scale and resolution corrections are applied on Monte Carlo. The correction algorithms are provided by the corresponding ATLAS combined performance groups and the parameters are obtained using tag-and-probe studies and fits to the Z line-shape. Details can be found in [89] and [90].

7.3.3 Jet selection

The jets used in this analysis are reconstructed with the anti-k_T algorithm with a radius parameter of 0.4 (see Section 4.3.4). Also for jets, two object quality classes are defined for the analysis: signal jets, which are used to reconstruct the Higgs boson candidate, and veto jets, which are used to reject unwanted events. All jets are required to have $p_\mathrm{T} > 20\,\mathrm{GeV}$. Jets with $p_\mathrm{T} < 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ are further required to fulfill JVF > 0.5 to reject pile-up. Signal jets must satisfy $|\eta| < 2.5$ to facilitate the application of b-tagging. In the forward region ($2.5 < |\eta| < 4.5$) only veto jets are considered. Since JVF cannot be calculated for these jets due to missing tracking information, the p_T cut is raised to $p_\mathrm{T} > 30\,\mathrm{GeV}$ to reduce pile-up. The main purpose of the veto jets is the reduction of the $t\bar{t}$-background.

7.3.4 Overlap removal

Some objects may pass the selection requirements for more than one category, for example an electron may also be identified as a jet, or a semi-leptonic heavy flavor decay within a jet may produce a lepton. To avoid double counting a procedure to remove this so-called overlap must be defined. In this analysis, the final object classification is done according to the following list of priority criteria:

- Jets which lie within a cone of radius 0.4 with respect to a loose electron with $p_\mathrm{T} > 15\,\mathrm{GeV}$ are discarded.
- Muons which lie within a cone of radius 0.4 with respect to a jet are discarded.

- Loose electrons with $p_\mathrm{T} < 15\,\mathrm{GeV}$ which lie within a cone of radius 0.4 with respect to a jet are discarded.

- Electrons which lie within a cone of radius 0.2 with respect to a muon are discarded.

7.3.5 b-tagging

An efficient identification of jets originating from b-quarks is crucial for the selection of Higgs candidates. The discrimination of so-called b-jets against jets without heavy-flavor content relies mainly on the large mass of b-hadrons and on the fact that their lifetime is considerably longer than that of hadrons which do not contain b-quarks. Both properties affect the impact parameter distribution of such jets and allows for the identification of secondary decay vertices which are a typical attribute of b-jets.

In this analysis, a more advanced algorithm is used: MV1 [91] combines several observables associated with b-jet identification (impact parameter, displacement of the secondary vertex, $b \to c$ decay chain fit) via an artificial neural network and returns a jet-based weight w_MV1 which takes values between zero and one, where high output values indicate a high probability for a jet to originate from a b-quark. The efficiency of a jet selection based on a given cut on the value of w_MV1 is measured in $t\bar{t}$ events as a function of the transverse jet momentum. Using this information, the efficiency in simulated events is calibrated by applying p_T^{jet}-dependent event weights, see [92] for details. Signal candidate events are required to have exactly two jets for which w_MV1 exceeds $w_\mathrm{MV1}^{70\%}$. This corresponds to a working point with a b-jet acceptance of 70%, a c-jet rejection factor of 5 and a light-jet rejection factor of 150. Jets passing this cut are refered to as "b-tagged jets". The b-tagging requirement for jets is not included at object selection level to allow for the construction of control regions later which contain mainly light- or c-jets (see Section 7.8).

7.3.6 p_T-dependent energy correction for b-jets

The default jet energy corrections provided by the corresponding performance group mitigate the effects of in-time pile-up, correct for inefficiencies which arise due the sampling structure of the calorimeter or holes in the acceptance and compensate scale differences between electromagnetic and hadronic interactions. However, several deficiencies which mainly affect b-jets are not taken into account:

1. The reference for the calorimeter response is derived from the average mixture of particles over all jets. The different composition of b-jets is not specifically addressed.

2. In case of semi-leptonic decays in a jet, the energy carried away by neutrinos is not taken into account.

3. The energy corrections are based on truth jets constructed with the anti-k_T-algorithm with a size parameter of 0.4. Gluons emitted at wider angles do therefore not contribute.

These issues are addressed by an additional, specific jet-energy correction for b-jets [93]: using a $t\bar{t}$ Monte Carlo sample, reconstructed b-jets are matched to the original partons from the hard process using a ΔR criterion. The ratio of jet to parton energy is calculated for all events in the sample and parametrized as a function of p_T^{jet}. In addition, scale factors which account for the difference between fully hadronic b-decays and semi-leptonic decays are derived. Based on this information, event-by-event jet energy correction factors for the analysis presented here are obtained, which are applied as follows: first, nearby muons or electrons are identified and a p_T correction factor is applied if a lepton is matched to the jet. In case of a nearby muon, the energy carried by the muon which eventually escapes the detector is also taken into account. This way, the p_{T} scale of the jet is adjusted to that of a jet with a fully hadronic b-decay. Second, out-of-cone and calorimeter response effects are covered by the aforementioned jet-to-parton correction function.

The b-jet energy correction improves the scale and the resolution of $m_{b\bar{b}}$, the invariant mass calculated from the two signal jets, significantly. Figure 7.2 shows the $m_{b\bar{b}}$ distribution in a Pythia WH Monte Carlo sample with $m_H = 125\,\mathrm{GeV}$ before and after the correction, normalized to unit area. Although the plot was originally created for an earlier version of the analysis aimed at the ATLAS data taken in 2011, it is sufficient to illustrate the influence of the correction for the analysis presented here: the fit-function shown in the plot is a so-called "Bukin-function", see [94] and [95] for details. Two parameters of this function are of interest to evaluate the performance of the correction: the position of the peak $\hat{x}$ and its width, which is given by $\sigma_{\hat{x}} = \frac{\mathrm{FWHM}}{2\sqrt{2\log 2}}$. Applying the correction described above, the peak position moves from $\hat{x}^{\mathrm{before}} = 115.4 \pm 0.3\,\mathrm{GeV}$ to $\hat{x}^{\mathrm{after}} = 125.4 \pm 0.2\,\mathrm{GeV}$ and the width reduces from $\sigma_{\hat{x}}^{\mathrm{before}} = 18.2 \pm 0.2\,\mathrm{GeV}$ to $\sigma_{\hat{x}}^{\mathrm{after}} = 14.40.1\,\mathrm{GeV}$, which corresponds to an improvement in the relative mass resolution of 27%. It should be noted that the correction is strictly jet-based and thus does not shape the $m_{b\bar{b}}$ distribution of the background samples in an undesirable way.

Since $m_{b\bar{b}}$ is the most important discriminating variable in the BDT analysis (see Section 7.5), the resolution improvement translates almost directly into an increased sensitivity for the entire $VH \rightarrow Vb\bar{b}$ search.

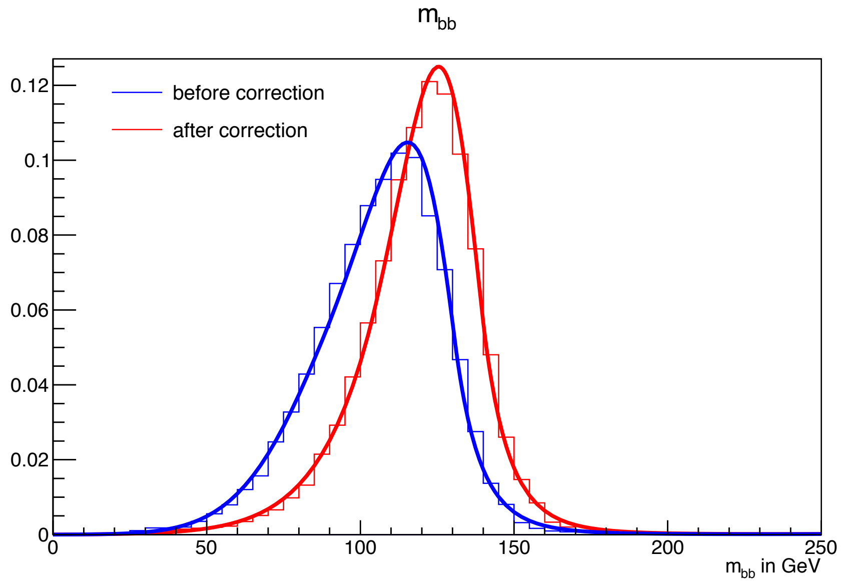


Figure 7.2: The effect of the jet energy correction on the m_{bbarb} distribution in the Pythia WH signal Monte Carlo sample with $m_H = 125\,\mathrm{GeV}$.

7.3.7 Missing transverse energy

Since the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ is calculated indirectly from the four-momenta of all other objects in the event (see also Section 4.3.6), it is affected by all corrections like lepton momentum smearing/scaling and jet recalibration. To account for this fact, $E_{\mathrm{T}}^{\mathrm{miss}}$ is reevaluated after all other object corrections have been applied and before any further event selection is performed.

7.4 Modeling of the multijet background

While the ZH channel is essentially free of multijet background after the pre-selection (see Section 7.7), the search in the WH channel suffers from a considerable amount of residual multijet events. The multijet background is a particular challenging one for two reasons: first, no appropriate Monte Carlo description of this background is available due to the overwhelming number of contributing processes, and second, the enormous cross-section and low selection efficiency would require a huge number of events to be simulated. For this reason, shape templates for the multijet background are derived from data in a control region where the track isolation requirement on the leptons is reversed. The assumption underlying this strategy is that the lepton isolation is orthogonal to other event properties, so that an event sample drawn from the multijet-enriched high $\mathtt{iso}^{track}$ region can serve as a template for the multijet background in the low $\mathtt{iso}^{track}$ region. One should note that the single lepton triggers used in the analysis of 2012 data (see Section 7.3) already include track isolation requirements to match the increasing pile-up conditions. At first sight this seems to render the inversion of track isolation impractical, but since the trigger-based isolation only considers tracks with $p_\mathrm{T} > 1\,\mathrm{GeV}$ compared to $p_\mathrm{T} > 0.4\,\mathrm{GeV}$ for the offline isolation, the approach is still valid. In fact, events passing the trigger but failing the offline isolation cut are more similar to those in the signal region than events that would be selected by a trigger without isolation requirement. The loss in statistics through the trigger is thus compensated by a better modeling. The dedicated multijet region is therefore defined by $0.04 \leq \mathtt{iso}^{track} \leq 0.1$ and a relaxed electron quality requirement of `med++` which enhances the statistics. Furthermore the cut on the calorimeter based isolation is tightened to $\mathtt{iso}^{calo} < 0.04$ since it was discovered that the m_T^W spectrum falls very rapidly for multijet events with $\mathtt{iso}^{calo} > 0.04$ and does not provide a good shape estimate for the multijet background in the signal region [96]. The distributions from the multijet region can directly be employed to obtain a multijet shape estimate for any variable that is of interest in the analysis. The normalization of these templates in the signal region and in all other control regions defined in the analysis (see Section 7.8 for details) is obtained in the following way: the multijet templates and the non-multijet backgrounds obtained from simulation are combined and the normalizations of the different components are fixed by a maximum-likelihood-fit to data using various variable distributions. See Section 7.9 for details. Both shape and normalization are determined independently for electrons and muons.

7.5 Choice of discriminating variables for the BDT analysis

The choice of variables which are used as inputs to the multivariate classification algorithm is guided by the following criteria:

1. Provide optimal discrimination between signal and background.
2. Provide an exhaustive kinematical description of the VH system.
3. Avoid redundancy and strong correlations.
4. Choose only variables which are correctly modeled in Monte Carlo.
5. Prefer variables which are robust with respect to systematic variations.

To determine the optimal performance achievable in terms of discrimination, a large set of possible variables encompassing variables employed in former multivariate analyses [97] and variables motivated by theoretical considerations [98] was taken as a starting point and subsequently reduced by removing poorly performing variables by repeatedly retraining and evaluating the BDT (see Section 6.2). Finally, an optimally performing subset comprised of the following fourteen variables was identified:

- $m_{b\bar{b}}$, $\Delta R(b\bar{b})$
- p_{T}^{V}, p_{T}^{jet0}, p_{T}^{jet1}, $E_{\mathrm{T}}^{\mathrm{miss}}$, HT (scalar sum of transverse energies of all objects in the event plus $E_{\mathrm{T}}^{\mathrm{miss}}$)
- $|\Delta\eta(b\bar{b})|$, $|\Delta\phi(V, b\bar{b})|$ where V denotes the four-vector of the vector boson
- $|\Delta\phi(lep, E_{\mathrm{T}}^{\mathrm{miss}})|$, $|\Delta\phi(lep, jet_{\mathrm{closest}})|$ (both only WH)
- m_{T}^{W}, p_{T}^{lep} (both only WH)
- m_{ll} (only ZH)

This variable set - shown in Figure 7.3 - was used to develop the event pre-selection (see Section 7.7) but *not* for the BDT which provides the input to the limit calculation. There are several reasons for this decision:

1. The variables related to the vector boson p_{T} and the angle between the signal jets are not perfectly modeled. Using them in the final discriminant would require dedicated correction and an additional systematic uncertainty.
2. Several variables in the set are very sensitive to systematic variations (see Section 7.10).
3. The variable set exhibits strong correlations and is probably redundant.

While these disadvantages are not of interest for the pre-selection, where one is only interested in a coarse identification of the signal-enriched part of the phase-space, they need to be taken seriously in the construction of the final discriminant. Focusing on robustness and to avoid redundancy, it seems promising to choose a physical description of the VH system which is based on Lorentz invariant quantities augmented with some angles which determine the orientation of the system. Using Lorentz invariant variables, the description of the VH system becomes independent of the transverse and longitudinal boost. This is highly desirable since the transverse boost is influenced by initial state gluon radiation which is not perfectly modeled and the longitudinal boost is related to the proton PDFs which also bear considerable uncertainties.

7.5.1 A Lorentz invariant description of the VH system

A complete parametrization of the final state of the system consists of four four-vectors, equalling 16 parameters. Ten of them can be chosen as Lorentz invariants: the pairwise products of the four-vectors (six) and four invariant masses of the final state particles. The other six parameters are defined by the Euler angles describing the orientation of the system in space (three) and - optionally - the boost vector of the system (three). Since the invariant masses of the charged lepton and the neutrino do not provide any information, they may be dropped. The jet masses may also be dropped since they are not well-modeled in the simulation. One of the three angles may be discarded since the process should be invariant with respect to rotation around the beam-line. Using the boost of the system would mean giving up on some of the desirable invariance properties of a Lorentz invariant parametrization - on the other hand it can be shown that the boost of the system carries useful discriminating information. Thus a compromise was adopted: the longitudinal component of the boost, which seems to be well-modeled, is kept and only the transverse boost is discarded. Following this line of thought, one obtains a description of the VH system which consists of nine variables:

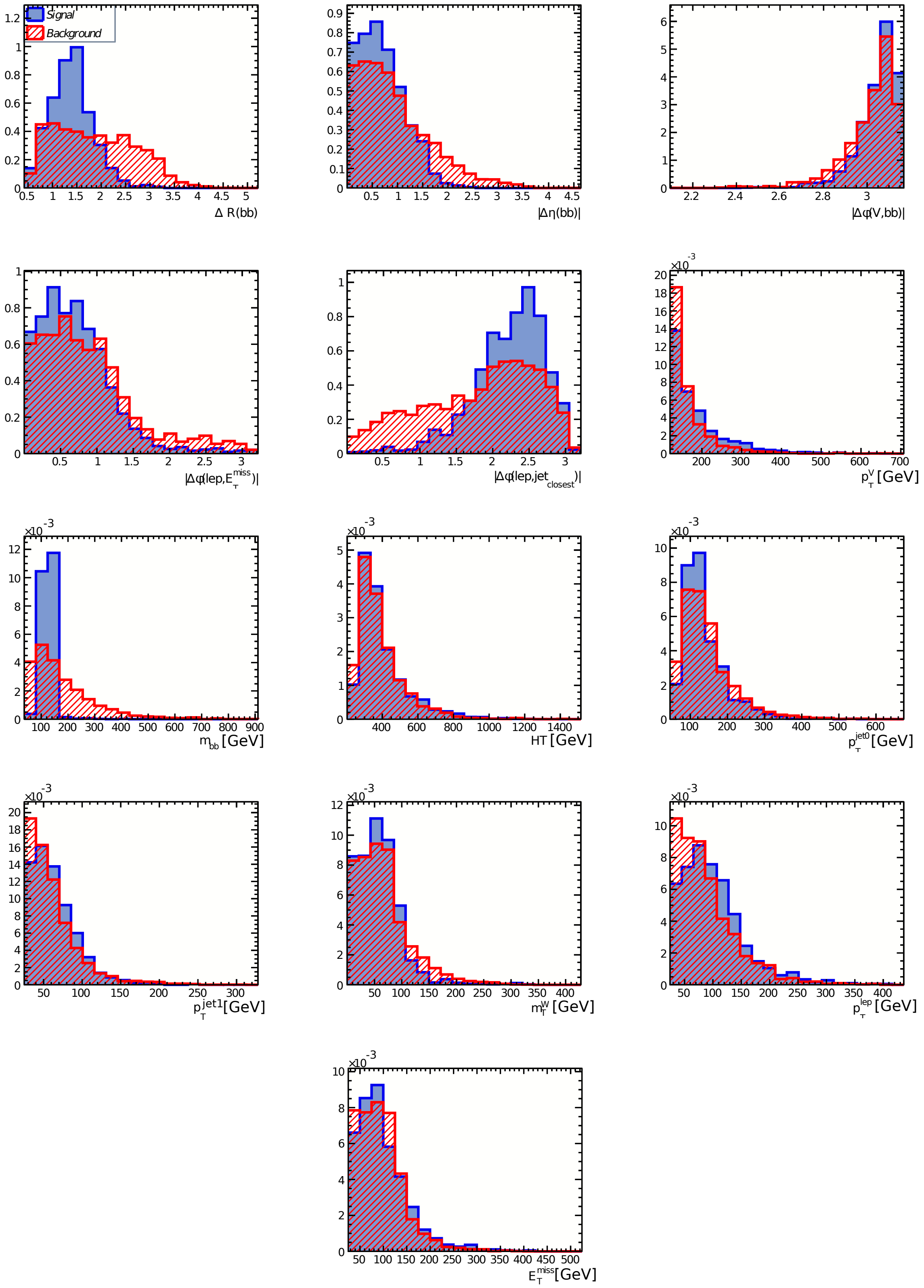


Figure 7.3: The set of optimally performing variables (definitions see text) in the WH channel. All distributions are normalized to unit area.

1. Pairwise four-vector products of all final state particles, denoted as
   - $p^{jet0}*p^{jet1}$
   - $p^{jet0}*p^{lep0}$
   - $p^{jet1}*p^{lep0}$
   - $p^{jet0}*p^{\nu}$(only WH)
   - $p^{jet1}*p^{\nu}$(only WH)
   - $p^{lep0}*p^{\nu}$(only WH)
   - $p^{jet0}*p^{lep1}$(only ZH)
   - $p^{jet1}*p^{lep1}$(only ZH)
   - $p^{lep0}*p^{lep1}$(only ZH)
2. The angle between the $b\bar{b}$-system and the beam-line, denoted as θ_1
3. The angle between the plane defined by the $b\bar{b}$-system and the beam-line on one hand and the $b\bar{b}$-system and the charged lepton (WH) or the lepton with the highest p_{T} (ZH) on the other hand, denoted as θ_2
4. The boost of the VH system in the direction of the beam, denoted as γ^z

While this parametrization can be calculated easily for the ZH channel, the z-component of the four-vector which corresponds to the neutrino in the WH channel is estimated in the following way [99]:

1. Asserting that both the lepton and the b-quark masses are sufficiently small with respect to the Higgs and W masses, one may assume that the energy E^{ν} carried by the neutrino corresponds on average to 25% of the WH mass m_{WH} in the rest frame of the WH system.
2. To arrive at the laboratory frame, one needs to boost along the beam axis:
   $\langle p_z^{\nu} \rangle = \beta\gamma \langle E^{\nu} \rangle = \frac{p_z^{WH}}{m_{WH}} \langle E^{\nu} \rangle = 0.25 p_z^{WH}$
3. Assuming a uniform distribution of the energy among the decay products, one arrives at
   $p_z^{\nu} \simeq \frac{0.25}{0.75}(p_z^{lep0} + p_z^{jet0} + p_z^{jet1})$

While this may seem like a very crude approach at first, it has the advantage of always providing a unique estimate for p_z^{ν}, in contrast to a constrained W mass fit which does not always find a solution and suffers from ambiguities.

Using the parametrization described above, not only is the number of variables for the final discriminant reduced from fourteen to nine, the new variables also exhibit considerably less correlations (see Figure 7.6) and are well modeled throughout all relevant control regions (see Section 7.8). Before they are fed to the BDT, all variables x are mapped to the range $[0, 1]$ by a transformation of the form $x' = x/(c + x)$, where the constant c is chosen such that $< x' >= 0.5$ and a dimensionless quantity is obtained. As discussed in Chapter 6, the performance of the BDT algorithm is not affected by this. The resulting shapes are shown in Figure 7.5, where both the distributions for the signal (WH for $m_H = 125\,\mathrm{GeV}$) and for the sum of the backgrounds have been scaled to unit area. Although it looks as if some variables in this set do not provide sufficiently different shapes for signal and background, it turns out that the performance of the classification algorithm is only very slightly diminished by exchanging the original set of fourteen-variables for the set composed of Lorentz invariants and angles

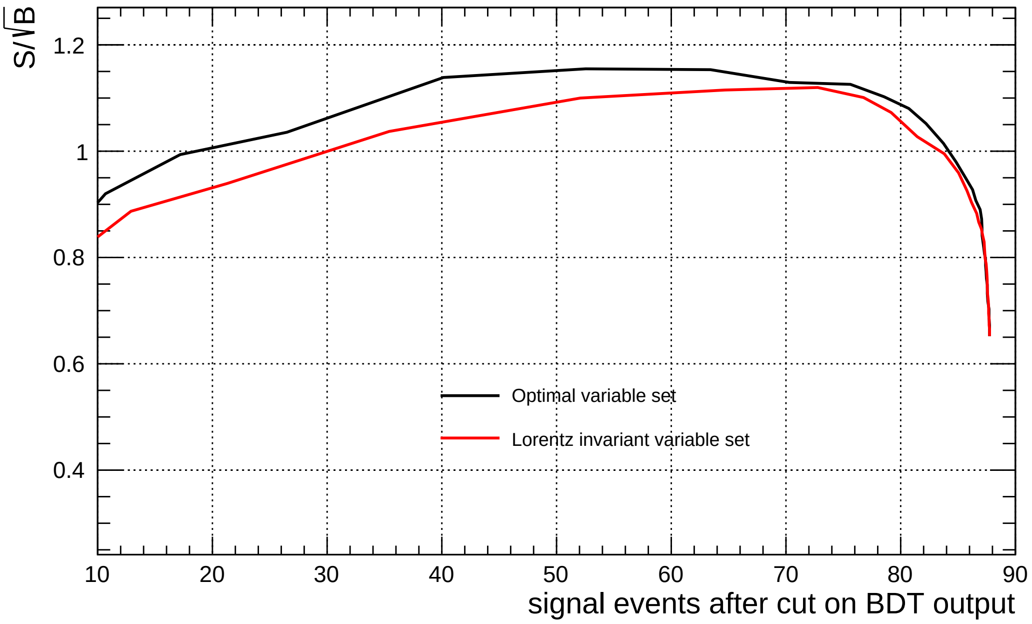


Figure 7.4: Comparison of the BDT performance for the two variable sets considered in the analysis.

(see Figure 7.4). The figure of merit used for this comparison is $S/\sqrt{B}$ after the best possible cut that could be performed on the BDT output w^{BDT}, which is represented by the maximum value attained by the curves in Figure 7.4. Note that this is a sound measure of the relative performance of different BDT configurations, but it does *not* correspond to the overall sensitivity that the analysis achieves. Additional studies have shown that the observed information loss can be ascribed to the neglect of the transverse boost in the Lorentz invariant variable set. However this loss in performance is compensated by a decreased sensitivity to systematic variations (see Section 7.10), which makes the Lorentz invariant variable set the superior overall choice.

7.6 Configuration of the final discriminant BDT

The concept of the BDT algorithm was introduced in Section 6.2. The implementation used in this thesis stems from the toolkit for multivariate data analysis in ROOT, TMVA [84]. The BDT algorithm is used for two different purposes in this analysis: first, to identify the signal-enriched phase-space to facilitate the construction of a coarse cut-based pre-selection. This is described in Section 7.7. Second, to combine the discriminating variables discussed in Section 7.5.1 into a single-valued estimator which is in turn used to calculate an upper limit on the $VH \to Vb\bar{b}$ cross section.

To make optimal use of the mass dependence of the chosen observables, separate BDT classifiers are trained for each simulated Higgs mass and for both the WH and the ZH channel. With signal Monte Carlo samples ranging from $m_H = 115\,\mathrm{GeV}$ to $m_H = 140\,\mathrm{GeV}$ in steps of $5\,\mathrm{GeV}$, this amounts to a total of 12 classifiers which are evaluated for each candidate event in the analysis.

Both boosting algorithms (adaptive boosting and gradient boosting) introduced in Section 6.2 were

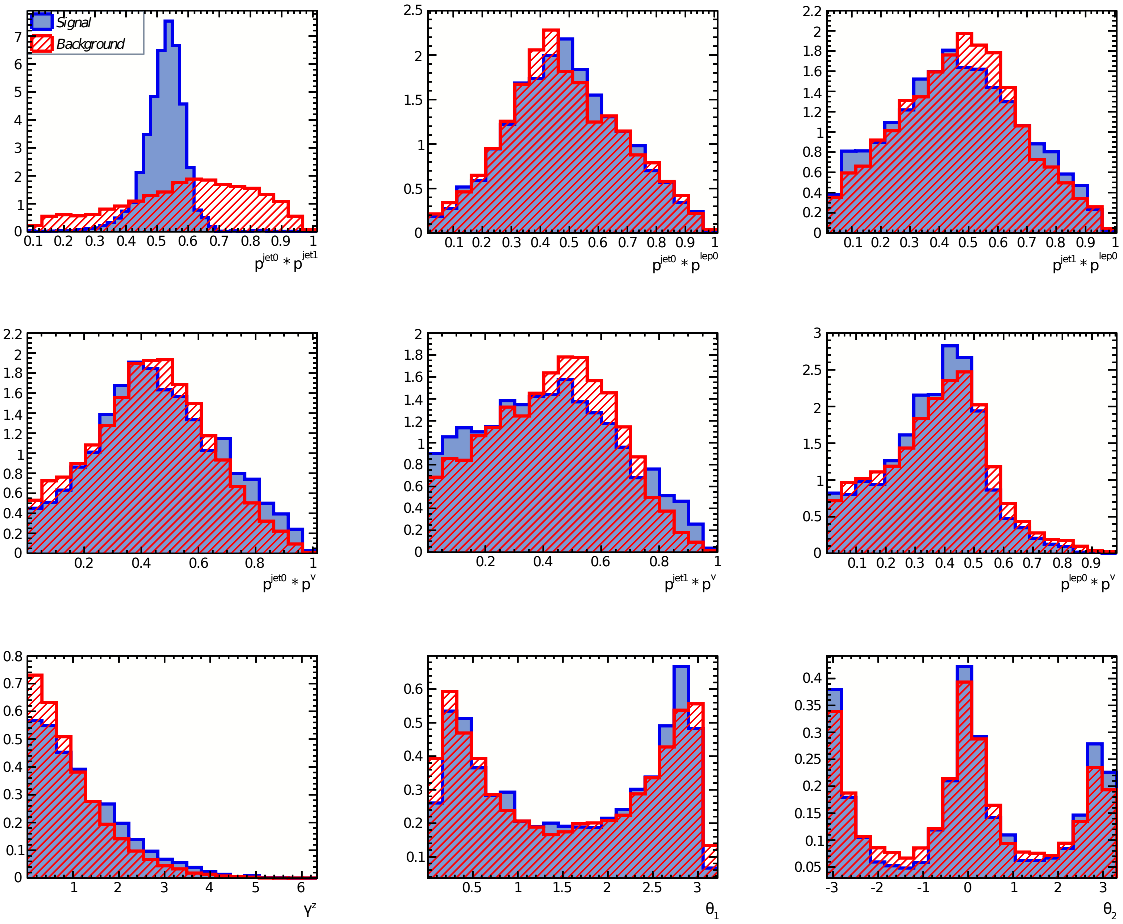
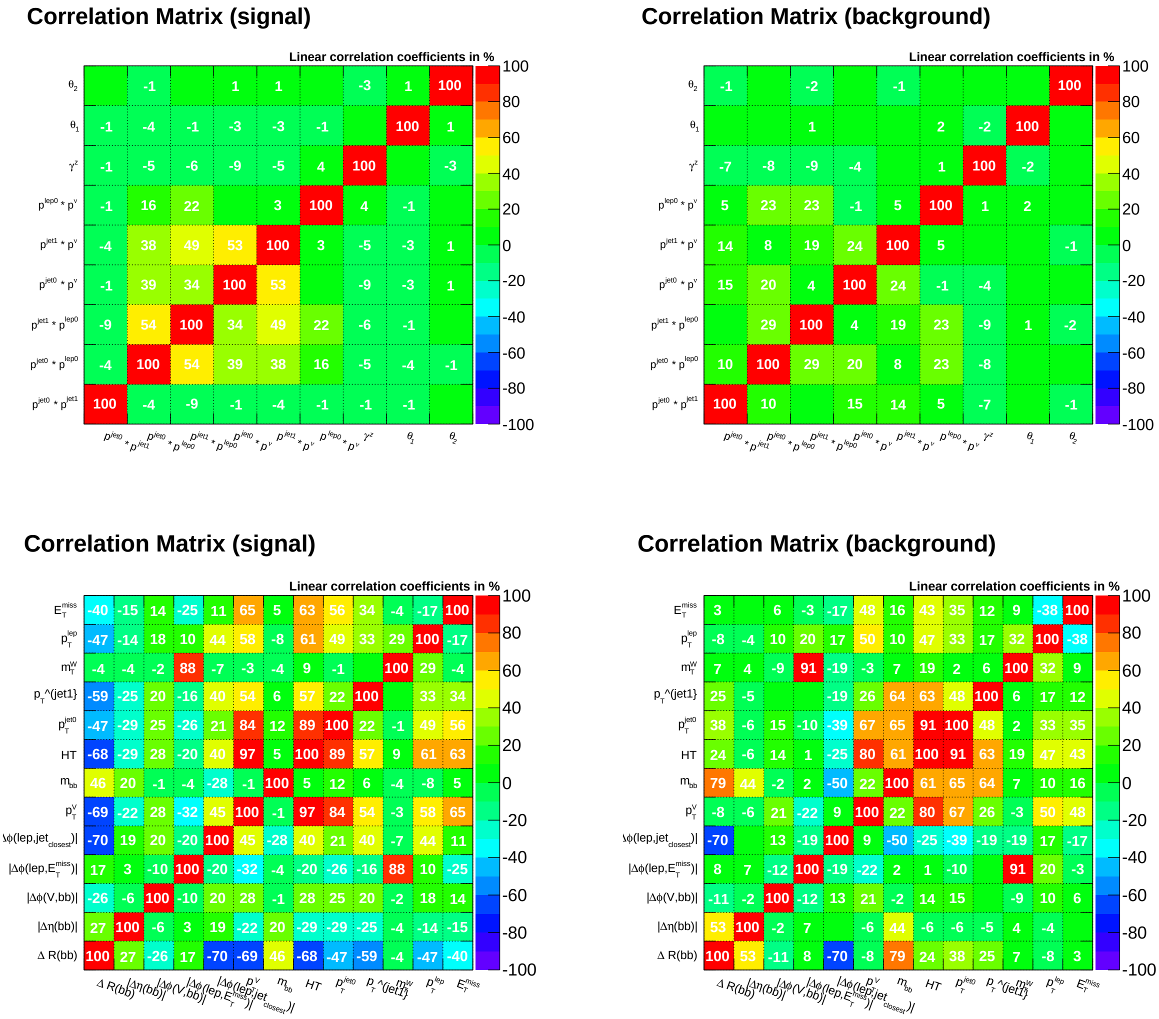


Figure 7.5: The set of discriminating variables composed of Lorentz invariant quantities and angles used in the WH channel.



Correlation Matrix (signal)

Correlation Matrix (background)

Correlation Matrix (signal)

Correlation Matrix (background)

Figure 7.6: First order correlations within the two variable sets studied in the WH analysis. Top row: Set based on Lorentz invariant quantities and angles. Bottom Row: Optimally performing set of variables used for the pre-selection.

considered for the analysis presented here. Although gradient boosting is the superior algorithm from a theoretical point of view, adaptive boosting was finally chosen to be the default one for this analysis since the current implementation of adaptive boosting in TMVA provides a consistent treatment of simulated events with negative weight in the training while the gradient boosting algorithm simply discards such events. However, the gradient boosting algorithms was independently optimized and used to cross check the performance obtained with adaptive boosting to ensure that the training reaches optimal convergence. Figure 7.7 shows the ROC (receiver-operator-characteristics, see also [100]) curves for both algorithms after the training for the discrimination of WH with $m_H = 125\,\text{GeV}$ against the sum of all backgrounds is completed: one immediately notes that both algorithms perform almost identical which suggests that the information provided for the training has been optimally exploited. Similar plots are obtained for all mass points and in both the WH and the ZH channel.

In agreement with the considerations presented in Section 6.2, where it was stated that small trees should exhibit superior performance, the following configuration was found to deliver the best discriminatory power independent of the Higgs mass:

- Adaptive boosting:
  - Number of trees: 400
  - Maximum tree depth: 4
  - Impurity measure: Gini index
- Gradient boosting:
  - Number of trees: 1000
  - Maximum number of nodes: 5
  - Fraction of events to be sampled at each iteration (see [83]): 0.6
  - Impurity measure: Gini index

7.6.1 Optimal usage of the available training statistics

In order to avoid a bias from possible overtraining, the simulated events which are used to train the BDT must not be reused for data/MC comparisons later in the analysis. Since this requirement would effectively reduce the Monte Carlo statistics for the limit fit and thus degrade the sensitivity of the analysis, the following strategy is adopted [101]:

1. The Monte Carlo samples are randomly bisected using the parity of the event number.
2. *Two* independent BDT classifiers are trained, one on each half of the simulated events.
3. In the application phase of the analysis, each event is evaluated by the classifier which has *not* seen this event during training. Two output distributions are obtained, which are then summed to yield the final discriminant output. This way, the complete sample can be used.

The drawback of this procedure is that the two classifiers do not yield the exact same shape, which has a similar effect as if the BDT output distribution w^{BDT} would be slightly smeared. However, it has been demonstrated [101] that the gain in statistics outweighs this effect. This also means that the entire analysis would benefit from an increase in Monte Carlo statistics which would make the described procedure obsolete.

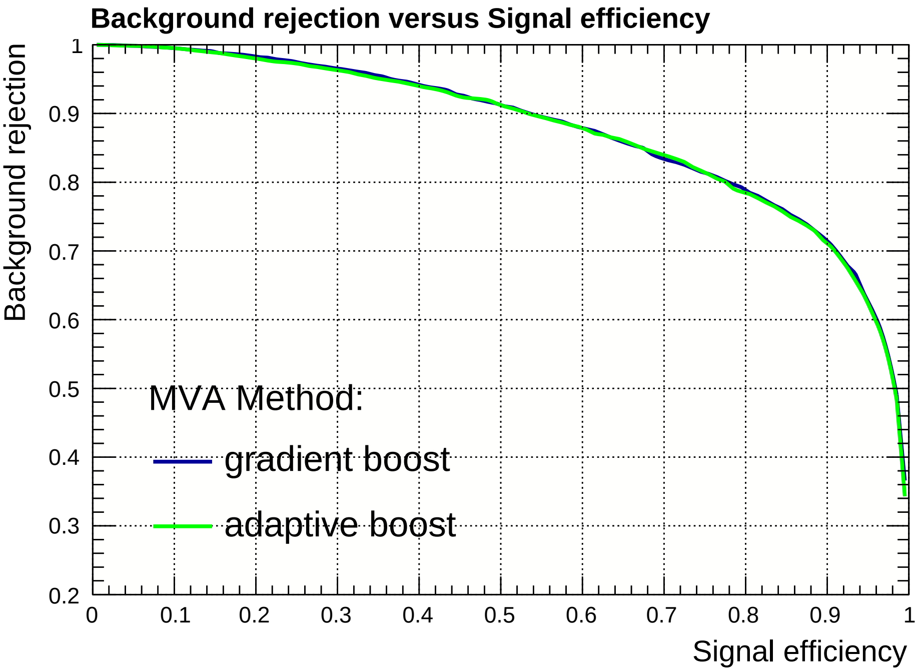


Figure 7.7: Comparison of signal efficiency / background rejection characteristics for the two boosting algorithms considered.

7.7 Selection of candidate events

Not all events in the data set considered for this analysis are admitted to the BDT discriminant whose output is used to calculate the upper limit on the $VH \to Vb\bar{b}$ cross section. Besides basic event quality requirements like a minimum of three tracks associated to the primary vertex, a two-level pre-selection is used to ensure that the application of the final BDT discriminant is performed only in a region of the phase-space which is well understood in terms of a reliable description through Monte Carlo simulation. The first part of the pre-selection is directly based on the multiplicities of the objects defined in Section 7.3 - this is called “signature-selection”. For an event to be considered in the WH branch of the analysis it must contain exactly one tight lepton, for the ZH analysis it must contain exactly one tight and one loose lepton. Two b-tagged signal jets are mandatory in both cases. Additional leptons or jets will trigger a veto, with the exception of a third non-b-tagged jet which is allowed in the ZH channel.

The second part of the pre-selection process is more sophisticated and follows two objectives: first, suppress the multijet background which cannot be described with simulated samples and is thus estimated from data by means of a separate control region which leads to an increased systematic uncertainty concerning the shape and normalization of this background. Second, identify the part of the phase-space which contains the signal events that are the most “typical” ones, i.e. that are most easily identified by the BDT and will thus yield the highest sensitivity in the final discriminant. A reduction of the analysis phase-space by additional pre-selection cuts should remove as few of these high-sensitivity events as possible. At the same time, the pre-selection cuts must be tight enough to ensure a reliable modeling of the remaining phase-space. To satisfy both requirements, the following strategy is adopted: using the maximally informative variable set introduced in Section 7.5, a BDT is trained on all (Monte Carlo) events that pass the signature-selection. The phase-space region which contains mostly events with a high BDT output w^{BDT}_{pres} will be responsible for the sensitivity of the analysis and must not be discarded prematurely. Even if the description of the background processes is not perfect after signature-selection, it is safe to use the BDT output at this stage as a guide to identify cuts which discard only events with low w^{BDT}_{pres} and will thus have little effect on the signal region. This can be done by plotting the distributions of the discriminating variables for different ranges of w^{BDT}_{pres} and deriving cuts on the variables which show a strong dependence on w^{BDT}_{pres}. Each time a promising cut is identified and applied, the multijet reduction and the data/MC agreement in the remaining phase-space is checked and the procedure is iterated until a satisfying agreement is obtained. The distributions that are considered to judge the data/MC agreement correspond to the variables that are used in the final discriminant. At the same time, a rough estimator of the significance based on the would-be optimal cut on w^{BDT}_{pres} is monitored to ensure that the sensitivity is actually not reduced more than necessary. This procedure yields the following pre-selection cuts:

1. $p_{\mathrm{T}}^{jet0} > 45\,\mathrm{GeV}$, $p_{\mathrm{T}}^{jet1} > 25\,\mathrm{GeV}$
2. $HT > 180\,\mathrm{GeV}$
3. $m_{b\bar{b}} > 40\,\mathrm{GeV}$
4. $m_{\mathrm{T}}^{W} < 100\,\mathrm{GeV}$ (WH only)
5. $71\,\mathrm{GeV} < m_{ll} < 121\,\mathrm{GeV}$ (ZH only)

It is interesting to note that no cut is applied on $E_{\mathrm{T}}^{\mathrm{miss}}$ (and no lower cut on m_{T}^{W} in the WH channel), contrary to what one would probably have expected for an analysis that aims to identify leptonic decays of vector bosons. A study of the corresponding distributions in the signal Monte Carlo has shown that a sizable fraction of the WH signal is in fact found at low values of m_{T}^{W} and $E_{\mathrm{T}}^{\mathrm{miss}}$ but can still be identified by the BDT: cutting on these variables would therefore unnecessarily reduce the sensitivity of the analysis.

The MC/data agreement for the variables defined in Section 7.5.1 in the remaining phase-space is very good as can be seen in Figure 7.9/ Figure 7.10 (WH channel) and Figure 7.11/ Figure 7.12 (ZH channel). Here, the cross sections of the Monte Carlo samples have already been scaled according to the result of the normalization fit described in Section 7.9. For visualization purposes, the signal is overlayed as a red line with ten times the standard model cross section in all plots. The grey band in the ratio plot indicates the statistical uncertainty of the Monte Carlo prediction. A large version of the common legend for all the distributions considered in the analysis is shown in Figure 7.8.

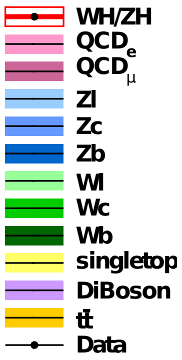


Figure 7.8: Common legend for all data/MC comparison plots.

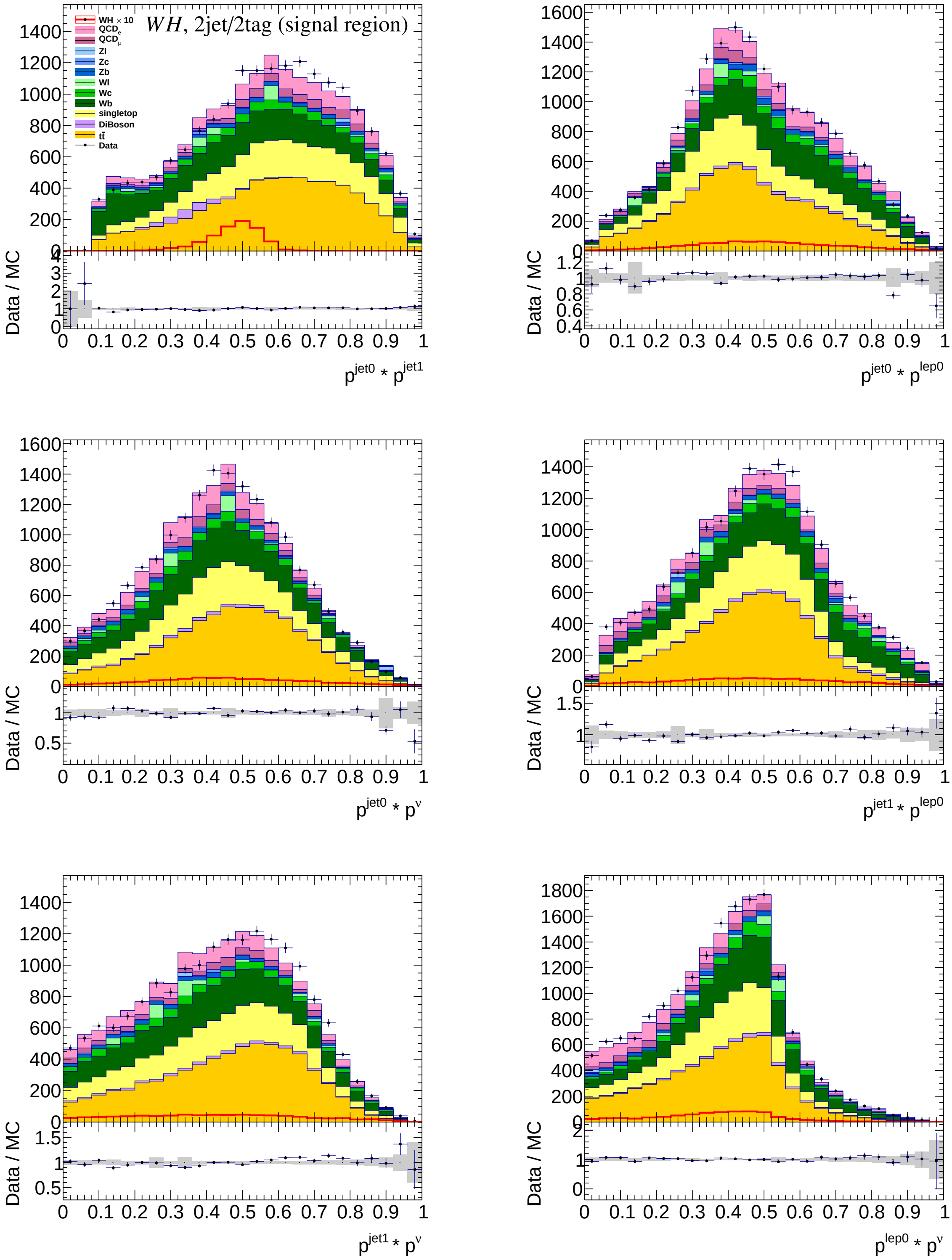


Figure 7.9: The Lorentz invariant variables given by the four-vector products of the final state objects in the signal region of the WH channel. These variables (and the ones on the next page) form the input to the final discriminant BDT in the WH channel.

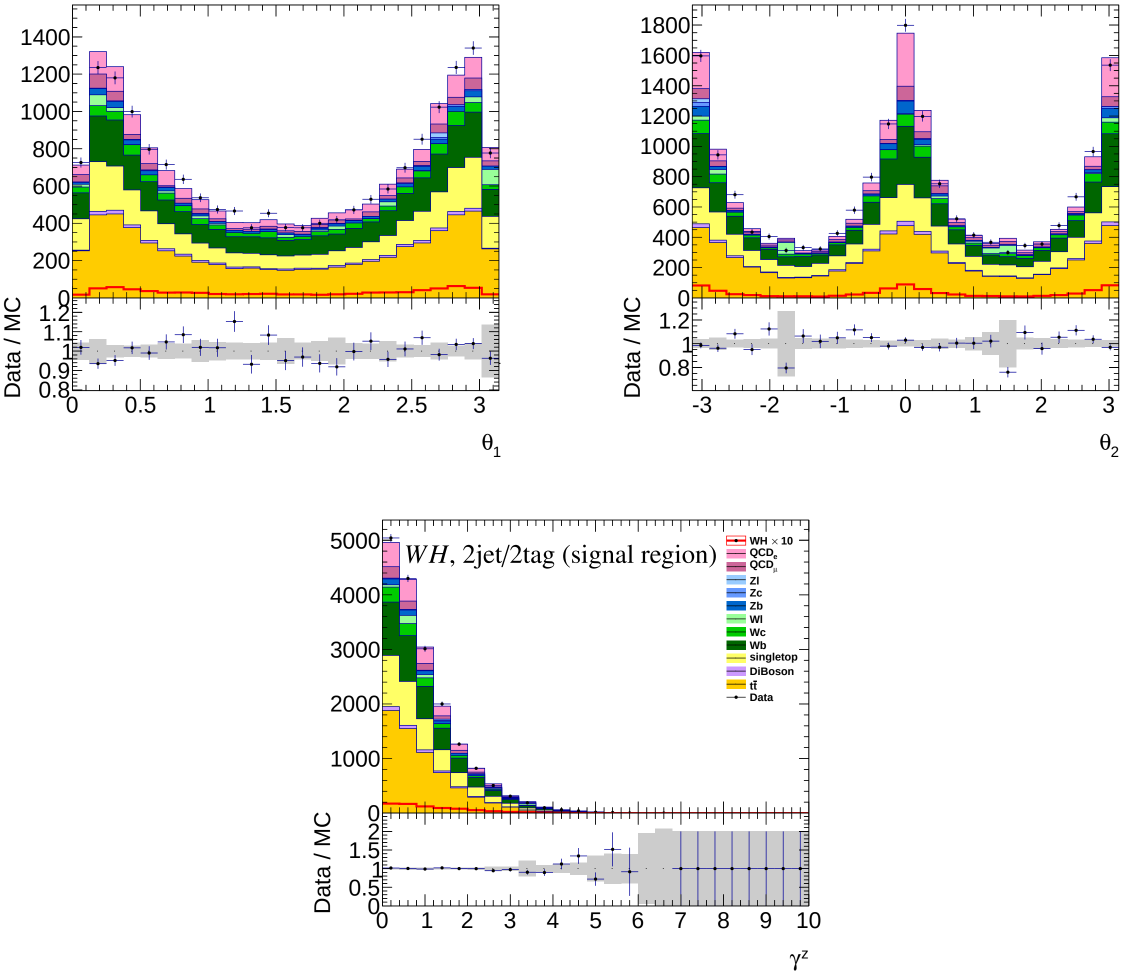


Figure 7.10: The angular variables and the longitudinal boost in the signal region of the WH channel.

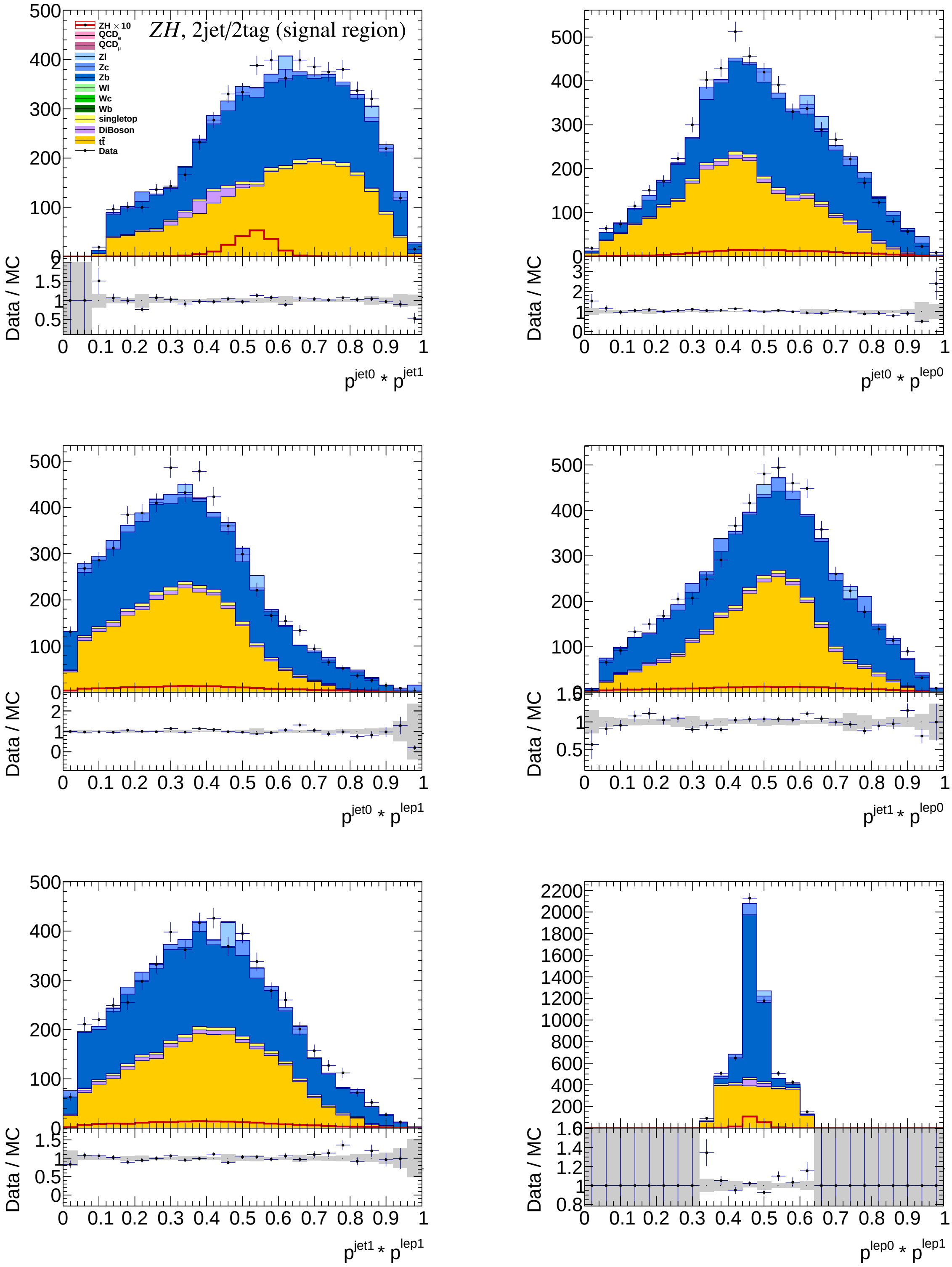


Figure 7.11: The Lorentz invariant variables given by the four-vector products of the final state objects in the signal region of the *ZH* channel. These variables (and the ones on the next page) form the input to the final discriminant BDT in the *ZH* channel.

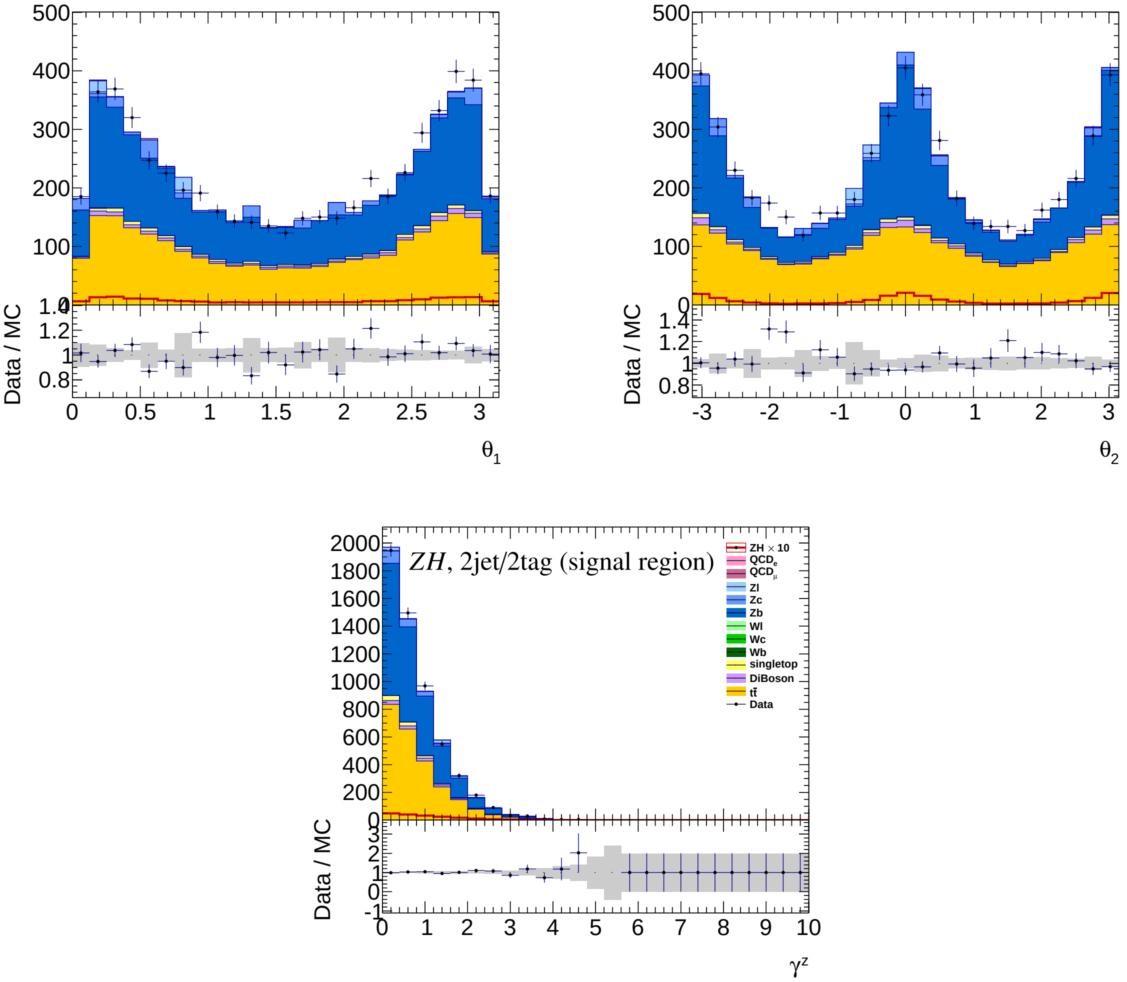


Figure 7.12: The angular variables and the longitudinal boost in the signal region of the ZH channel.

7.8 Definition of control regions

Processes resulting in the production of a vector boson and additional jets (referred to as V+jets) are one of the most important background sources for the $VH \to Vb\bar{b}$ search. A particular difficulty stems from the fact that the flavor composition of the jets originating from these processes is not exactly known. Determining this composition is important in order to obtain a reliable estimate for the expected background, notably for $V + b\bar{b}$ which yields an irreducible contribution to the signal region. To get a handle on the background composition, additional regions of the phase-space are considered on top of the signal region which is defined by the pre-selection discussed in Section 7.7. These are called "control regions" and serve to normalize the different background components as well as control the systematic uncertainties in the global fit (see Section 8.2). Each control region should be sensitive to a particular kind of background while at the same time being kinematically largely equivalent to the signal region in order to minimize the impact of systematic uncertainties related to extrapolation. Moreover the control regions should be statistically independent. An obvious choice which meets this requirement is to define control regions according to the number of b-tagged jets while keeping the other cuts of the aforementioned pre-selection: the region without any b-tag is used to constrain the normalization of the V + light background, while the region with exactly one b-tag mostly constrains $V + c$. Here V + light refers to all events with a vector boson and (at least) two jets within the acceptance which are neither bottom- nor charm-flavored. $V + c$ then refers to all events with at least one charm-flavored jet but without any bottom flavored jets within the acceptance.

Although the phase-space region with exactly two b-tagged jets (which contains also the signal) is in principle wide enough to allow for a normalization of $V + b$ using the signal-free part of the BDT output distribution with low w^{BDT}, ambiguities between the $V + b$ and the $t\bar{t}$ background lead to an increased uncertainty on the $V + b$ normalization compared to the other flavors. Due to the signal-like nature of the $V + b$ background, an increased uncertainty of its normalization immediately degrades the sensitivity of the analysis. Therefore, an additional control region with exactly three jets, two of them b-tagged, is introduced. This region is largely dominated by $t\bar{t}$ and single top production and thus helps to fix the normalizations for these background which in turn facilitates a more precise determination of the $V + b$ background normalization from the region with only two jets.

In the following, I will simply name the control regions after their jet and b-tag multiplicity:

2jet/0tag: This control region features events with exactly two jets, none of them b-tagged. It was originally intended to fix the V + *light* background but then discarded in favor of the 3jet/0tag region, see Section 7.8.1 for details.

3jet/0tag: This control region features events with exactly three jets, none of them b-tagged. It is used to fix the V + *light* background.

2jet/1tag: This control region features events with exactly two jets, with exactly one of them b-tagged. Is was originally intended to fix the $V + c$ background but then discarded in favor of the 2jet/1tag$_{slj}$ region, see Section 7.8.1 for details.

2jet/1tag$_{slj}$: This control region features events with exactly two jets but only the sub-leading one being b-tagged. It is used to fix the $V + c$ background.

2jet/2tag: The signal region contains events with exactly two jets, both of them b-tagged. The signal-free part of this region (low w^{BDT}) provides a constraint on the $V + b$ normalization.

3jet/2tag: This control region features events with exactly three jets, two of them b-tagged. Is is largely dominated by $t\bar{t}$ and single top.

The composition of all regions used in the analysis is illustrated in Figure 7.13. All control regions are intended for use in the final limit-fit (see Chapter 8) and also in the so-called "pre-fit" which is needed to initialize the multijet normalization for the limit-fit and to facilitate data/MC comparisons. See Section 7.9 for details.

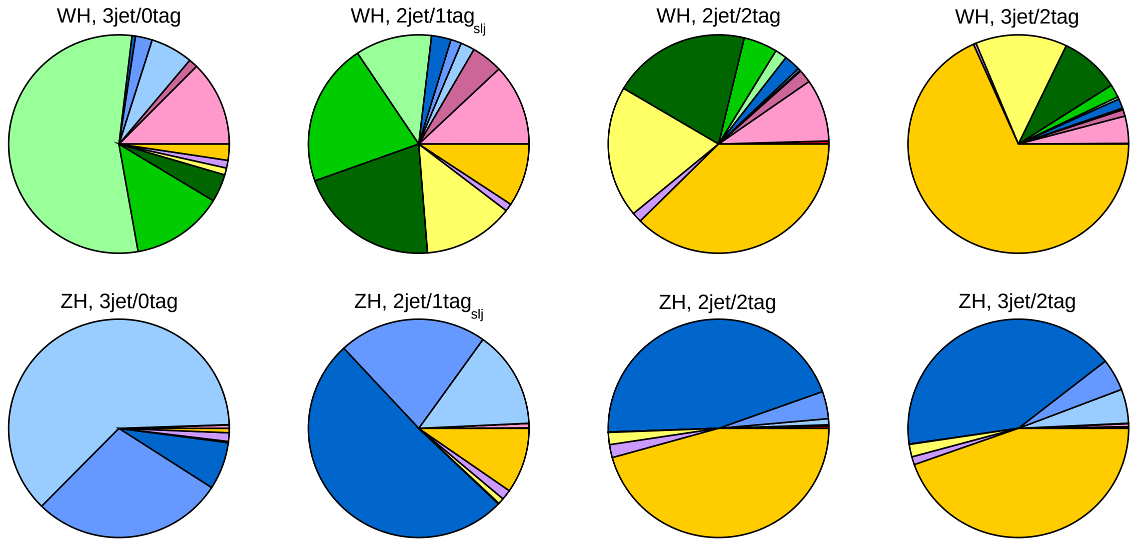


Figure 7.13: The background composition in all regions used in the analysis. The color code is explained in Figure 7.8.

7.8.1 Background modeling in the control regions

Although the sensitivity of the analysis is determined by the 2jet/2tag and - to a lesser extent - the 3jet/2tag region, it is instructive to inspect the distributions of the variables considered for the BDT also in the control regions which serve to constrain the background contributions. Studying the control regions helps to detect possible deficiencies in the Monte Carlo simulation and allows to validate the size of the systematic uncertainties attributed to modeling problems.

Starting with the 2jet/0tag region, it is evident that some of the variables - in particular those involving the four-vector of the sub-leading jet p^{jet1} - are systematically mismodeled by the simulation (see Figure 7.14), leading to a prominent slope in the data/MC ratio.

A similar, yet less pronounced effect is also seen in the 2jet/1tag region. The exact source of this mismodeling is still the subject of intensive discussion in the ATLAS $H \to b\bar{b}$ community at present, one of the hypotheses being that ISR/FSR is not correctly described by the current V+jets (Sherpa LO) Monte Carlo simulation. This is substantiated by generator studies which demonstrate that a NLO Monte Carlo simulation where ISR/FSR jets can also be generated from the matrix element improves on the description of many observables in V+jets [102], [103]. In the analysis presented here, several observations indeed suggest that a sizable fraction of the jets in the affected control regions stem from gluon radiation:

1. Dedicated studies of the 2jet/0tag region [104], [105] have shown that the data in this region is composed of two distinctive components, which will henceforth be referred to as the "soft" and the "hard" component. Surprisingly, the two components can be most effectively separated by

a cut on the invariant mass of the sub-leading jet, which suggests that the components may be characterized by the production mechanism of the sub-leading jet.

2. The jet pairs in the hard sub-sample commonly exhibit a back-to-back topology and are approximately balanced in p_{T}. In the soft sub-sample, the leading jet is rather back-to-back with the system formed by the lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$. No significant correlation between p_{T}^{jet0} and p_{T}^{jet1} is observed in the soft sub-sample.
3. The interpretation of these observations is as follows: the soft component of the data is dominated by W+1-jet production with an additional jet from gluon radiation. This hypothesis is substantiated by the $\Delta R(jet0, jet1)$-distribution of the soft sub-sample: it shows a distinctive peak at low ΔR and a plateau up to high values of ΔR. The former probably corresponds to events with gluon FSR emitted by the leading jet, while the latter correspond to events with gluon ISR emitted by one of the incoming partons. In this case, the emission is to first order independent of the W production and thus has no bearing with the direction of the leading jet.

Redefinition of the 0tag and 1tag control regions

Assuming that the data/MC agreement in the regions with 0/1 b-tag is mainly degraded by events where one of the jets stems from ISR/FSR, a strategy to reduce this contribution is developed based on two observations:

1. If gluon radiation is present, it most likely generates the least energetic jet in an event.
2. Jets from gluon radiation are rarely b-tagged (the exception being gluon-splitting to a $b\bar{b}$-pair which is not likely to happen).

As a consequence of this reasoning, the 2jet/0tag region is replaced with the 3jet/0tag region, assuming that a possible contamination with ISR/FSR will now mostly affect the third most energetic jet which is *not* used in the calculation of the relevant observables. Figure 7.16 and Figure 7.17 demonstrate that the data/MC agreement in this region is in fact superior to the description obtained in the 2jet/0tag region, although a residual slope is observed for the variables $p^{jet1} * p^{lep0}$ and $p^{jet1} * p^{\nu}$ which involve the sub-leading jet and the leptons. Therefore the 3jet/0tag region is only used to provide an overall normalization constraint for the global fit to avoid any influence that may be caused by residual shape mismodeling.

In the 2jet/1tag region, modeling can be improved by requiring that the sub-leading jet must be b-tagged which virtually rules out that it stems from ISR/FSR. To make this requirement explicit, the region is now labeled 2jet/1tag$_{slj}$. Figure 7.18 and Figure 7.19 demonstrate that this procedure really yields a decent data/MC agreement which even allows to extract shape information from the 2jet/1tag$_{slj}$ region in the limit fit (see Section 8.2).

The data/MC agreement in the 3jet/2tag region (see Figures 7.20 and 7.21) is sufficient to use this region without any correction.

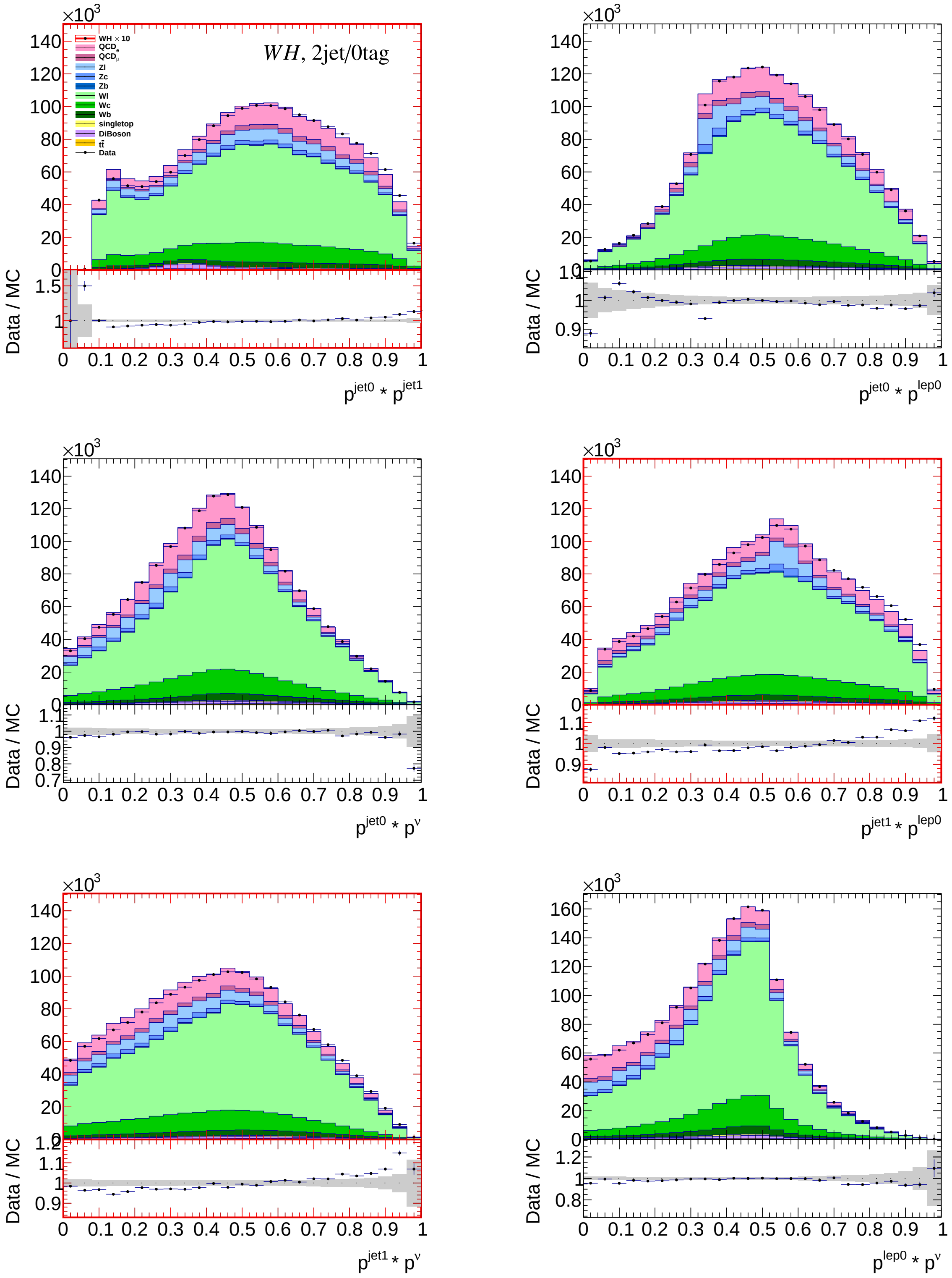


Figure 7.14: The Lorentz invariant variables given by the four-vector products of the final state objects in the 2jet/0tag control region of the *WH* channel. Systematic modeling deficiencies are clearly visible in the observables which involve the four-vector of the sub-leading jet p^{jet1} (red frames). The region is therefore not used in the final fit.

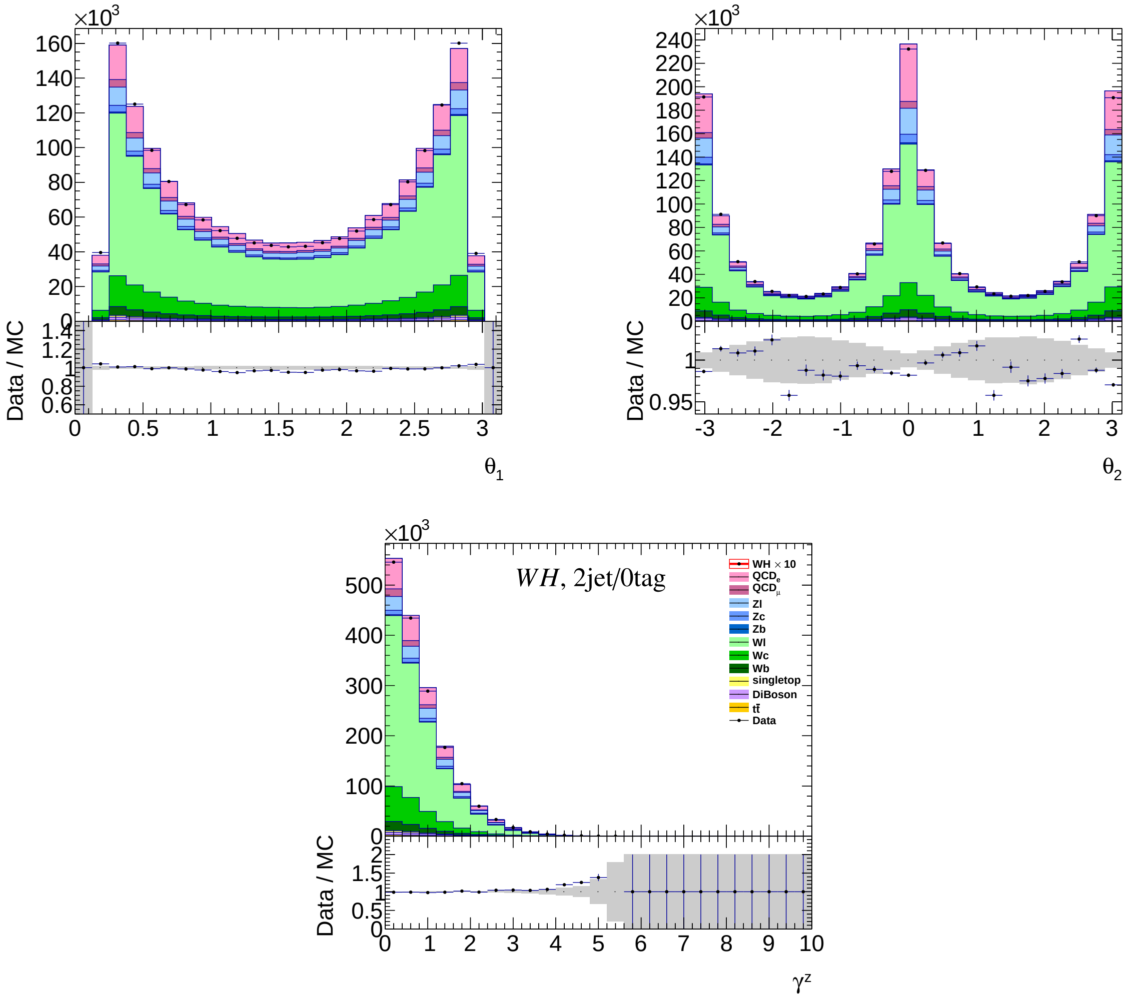


Figure 7.15: The angular variables and the longitudinal boost in the 2jet/0tag control region of the *WH* channel. This region is not used in the global fit due to modeling deficiencies.

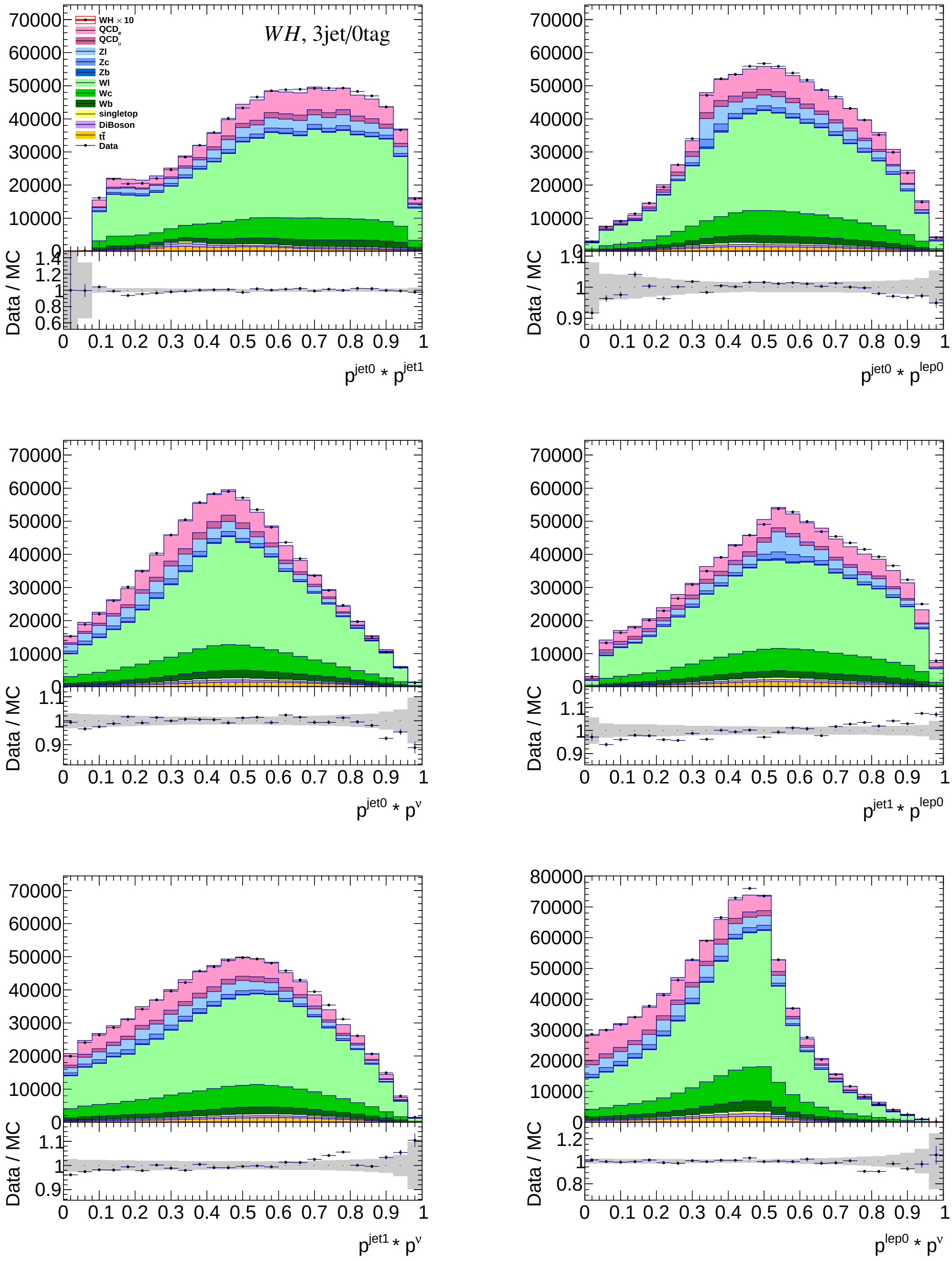


Figure 7.16: The Lorentz invariant variables given by the four-vector products of the final state objects in the 3jet/0tag control region of the WH channel.

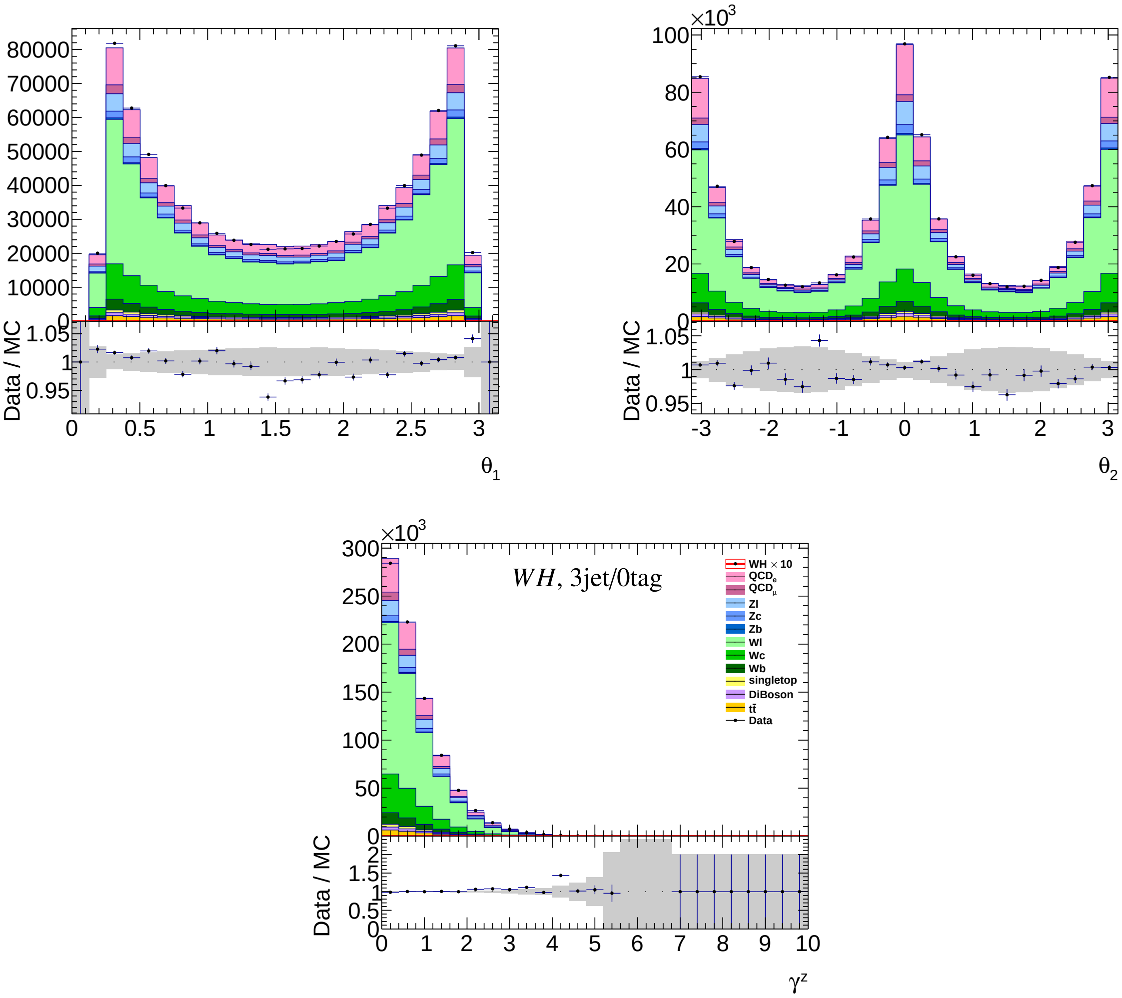


Figure 7.17: The angular variables and the longitudinal boost in the 3jet/0tag control region of the WH channel.

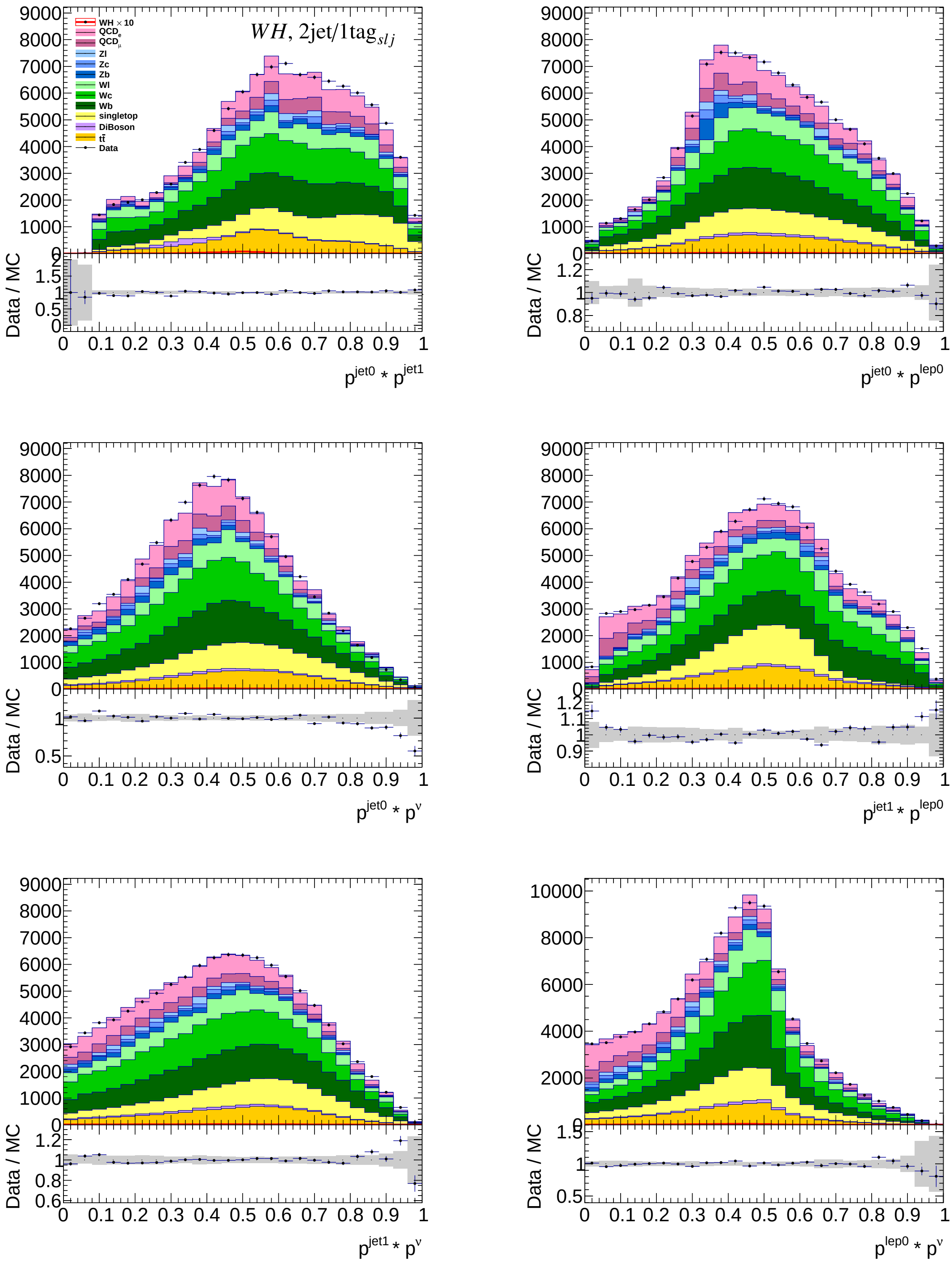


Figure 7.18: The Lorentz invariant variables given by the four-vector products of the final state objects in the 2jet/1tag$_{slj}$ control region of the WH channel.

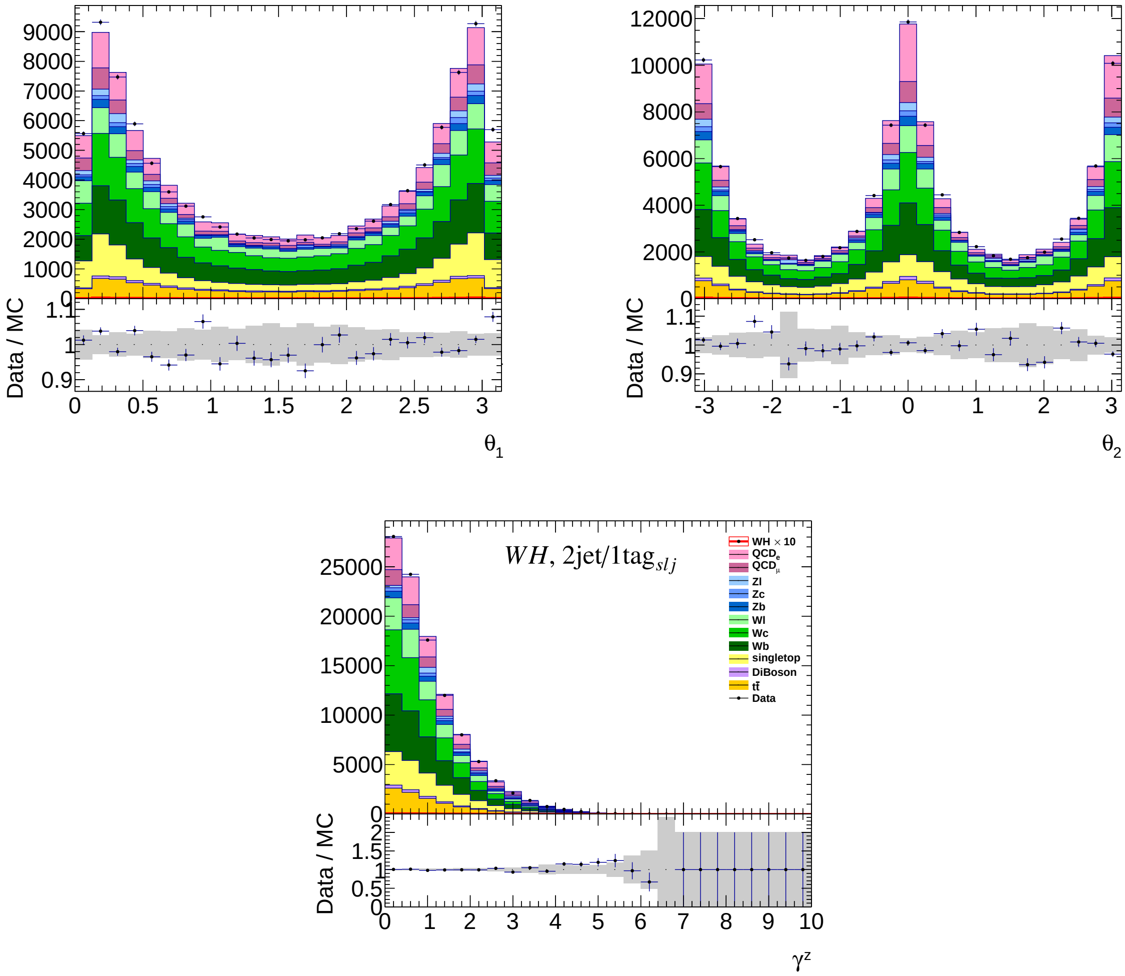


Figure 7.19: The angular variables and the longitudinal boost in the 2jet/1tag$_{slj}$ control region of the WH channel.

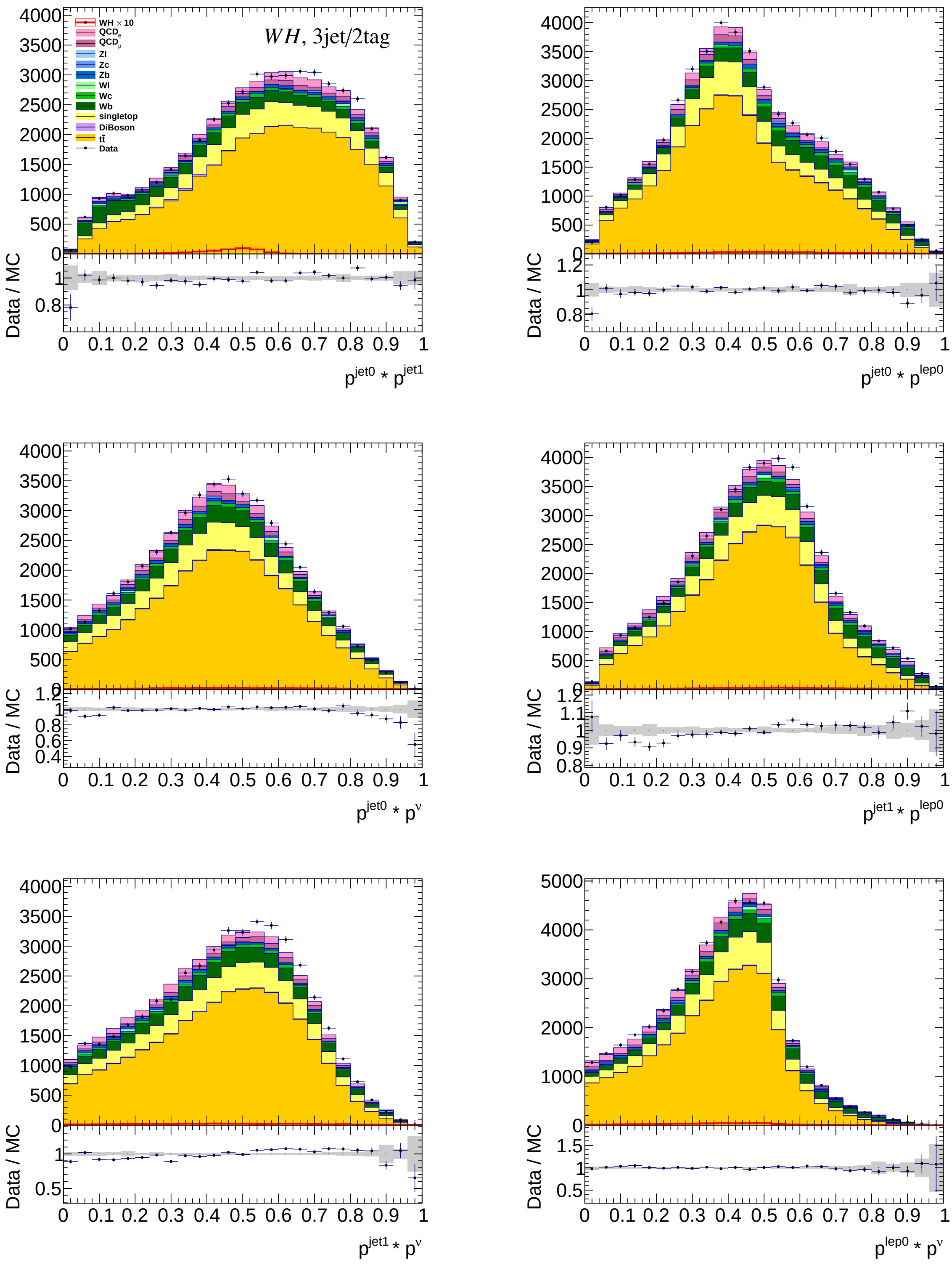


Figure 7.20: The Lorentz invariant variables given by the four-vector products of the final state objects in the 3jet/2tag control region of the WH channel.

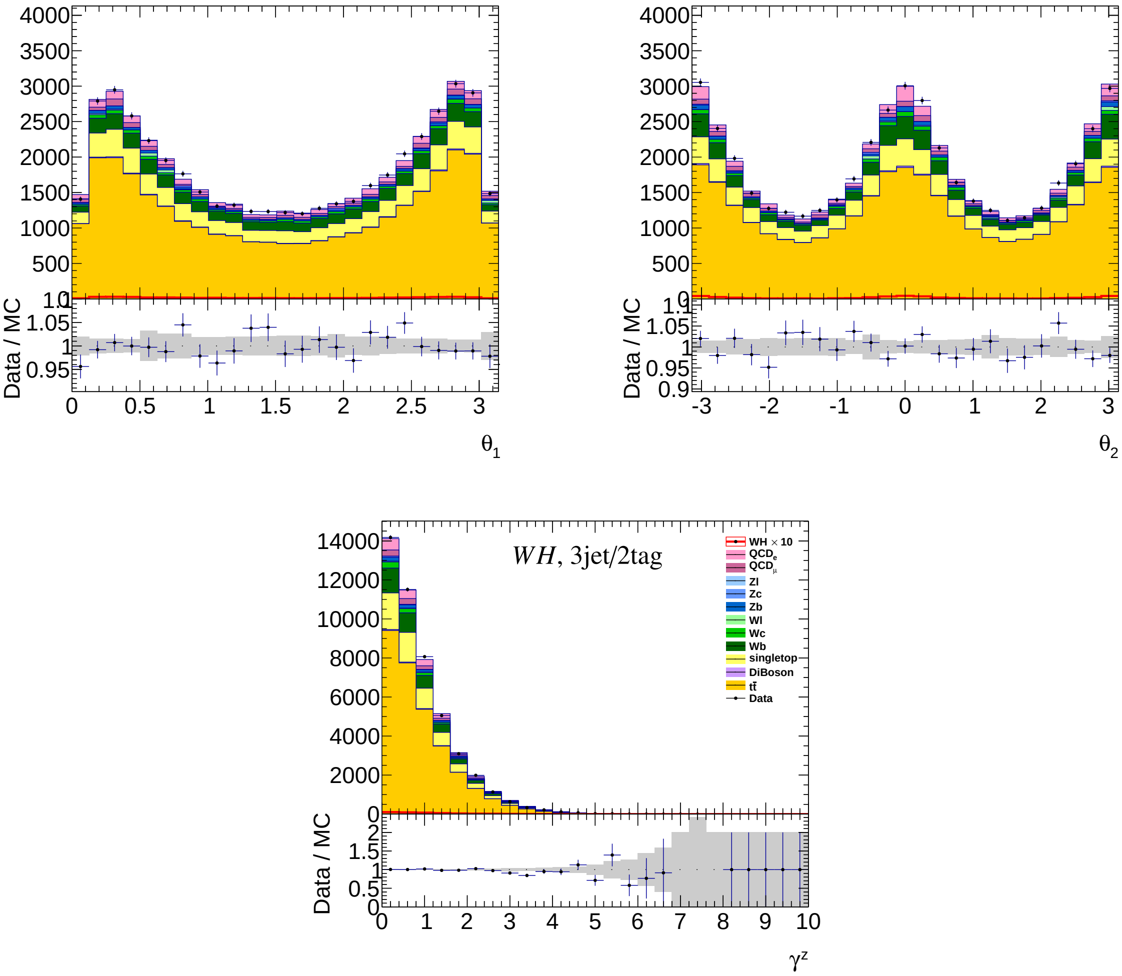


Figure 7.21: The angular variables and the longitudinal boost in the 3jet/2tag control region of the WH channel.

7.8.2 Monte Carlo modeling corrections

Correction of p_{T}^{V} in $t\bar{t}$ samples

As discussed in the ATLAS note on the differential $t\bar{t}$ cross section measurement [106], the p_{T} distribution of the top quarks is not accurately described by the available Monte Carlo samples. In the analysis presented here, this affects the modeling of the vector boson p_{T} in the regions with substantial $t\bar{t}$ contribution. Although the discriminating variables used in this analysis are deliberately chosen to avoid a dependency on the transverse boost of the vector boson, a residual correlation remains, most notably in the two angles. Therefore, a dedicated correction of p_{T}^{V} which was developed for the cut-based $VH \to Vb\bar{b}$ search presented in [85] is applied to the $t\bar{t}$ Monte Carlo. The correction reweights the Monte Carlo as a function of truth p_{T}^{top}, using the unfolded measurement from [106] as a reference. The systematic uncertainty associated with this correction is obtained by varying the nominal correction by ±50%.

Test of a $\Delta\phi(b\bar{b})$ correction for the V+jets samples

Studies carried out within the scope of the analysis presented in [85] suggest that the mismodeling observed in the 2jet/0tag and the 2jet/1tag region are due to an inaccurate description of the $\Delta\phi(b\bar{b})$ distribution in the Sherpa V+jets Monte Carlo samples. Carrying out a suitable correction could thus supersede the redefinition of the control regions. To check this, a corresponding study was carried out [107], where the recommended correction from the aforementioned analysis was applied. While the data/MC agreement is indeed improved for some distributions in the original 2jet/0tag and 2jet/1tag control regions, it is degraded for others [107], in particular in the aforementioned ISR/FSR depleted control regions. The proposed correction does thus *not* provide a complete remedy for the mismodeling observed in the relevant regions and is *not* used in the analysis presented here. Instead, the modified regions defined in Section 7.8.1 are kept without any further correction. However, to account for a possible residual mismodeling of $\Delta\phi(b\bar{b})$ which could also affect the variables used in this analysis, the systematics associated with the proposed correction are implemented according to the recommendations in [85]. This includes a shape variation which corresponds to 50% of the nominal $\Delta\phi$ correction, see also Section 7.10.2.

7.9 Pre-fit of background normalizations and the multijet template

Comparisons between data and Monte Carlo are essential to spot possible modeling inaccuracies which could degrade both the reliability and the sensitivity of the analysis. Such comparisons are only possible if the normalizations of all background contributions are accurately known. At least for the multijet background whose shape is described by templates obtained from a dedicated multijet control region (see Section 7.4), this requirement is not a priori met. Thus a dedicated fit is performed to determine its normalization, both to facilitate data/MC comparisons and to obtain a sensible initial value for its normalization in the limit fit. Since the limit fit allows the normalization of the V+jets and $t\bar{t}$ samples to float (see Chapter 8 for details), which in turn affects the multijet estimate, the pre-fit should be designed in a similar way, but *without* using any distribution which is sensitive to the possible presence of a Higgs signal. Therefore, a dedicated fit procedure for the background samples has been developed: the distribution which is most sensitive to the normalization of the multijet background is $p^{lep0} * p^{\nu}$, which conveys similar physical information as m_{T}^{W}. Besides this, several other distributions from the control regions defined in Section 7.8 are used in the fit to constrain the other background components, in particular the contributions of the different flavors in the V+jets background. All distributions are fitted simultaneously to data. A detailed record of the regions and distributions can be found in Table 7.2. Note that the signal-sensitive $p^{jet0}*p^{jet1}$ distribution is *not* used in the signal region. The scale factors resulting from this fit only serve to perform detailed data/Monte Carlo comparisons prior to the global limit-fit: with the exception of the multijet normalization, the scale factors are discarded when then input for the limit fit is constructed. The multijet normalization is used in the limit-fit, albeit with a large systematic uncertainty of 30%.

All distributions used in the fit are shown on the following pages (Figures 7.22, 7.23, 7.24, 7.25), ordered by the number of b-tagged jets: the left column always shows the unscaled (pre-fit) distributions and the right column the post-fit distributions. While the distributions from the ZH channel mostly help to constrain the normalization of the Z+jets and $t\bar{t}$ background, the distributions from the WH channel fix W+jets, single-top, $t\bar{t}$ and, most important, the multijet background. The ratio plots clearly indicate that the description of the data improves significantly after the fit and that the multijet templates described in Section 7.4 provide a reasonable shape estimate.

One should note that in some control regions more than one observable is used for the fit - this means that the statistical uncertainty of the scale factors determined in the fit is underestimated. However, since these scale factors and their uncertainties do *not* enter the limit calculation, this does not introduce a bias.

Distributions used in the background fit				
	2 jet, WH pres.	3 jet, WH pres.	2 jet, ZH pres.	3 jet, ZH pres.
0 b-tags	-	$p^{lep0}*p^{\nu}$	-	$p^{jet0}*p^{lep0}$
1 b-tag (sub-leading jet tagged)	$p^{jet0}*p^{jet1}$ $p^{jet1}*p^{lep0}$ $p^{lep0}*p^{\nu}$ (elec. only) $p^{lep0}*p^{\nu}$ (muon only)	-	$p^{jet0}*p^{lep0}$	-
2 b-tags	$p^{jet0}*p^{lep0}$ $p^{jet1}*p^{lep0}$	$p^{jet0}*p^{jet1}$ $p^{jet0}*p^{lep0}$	$p^{jet0}*p^{lep0}$	-

Table 7.2: Input distributions/regions used in the fit of the background normalizations

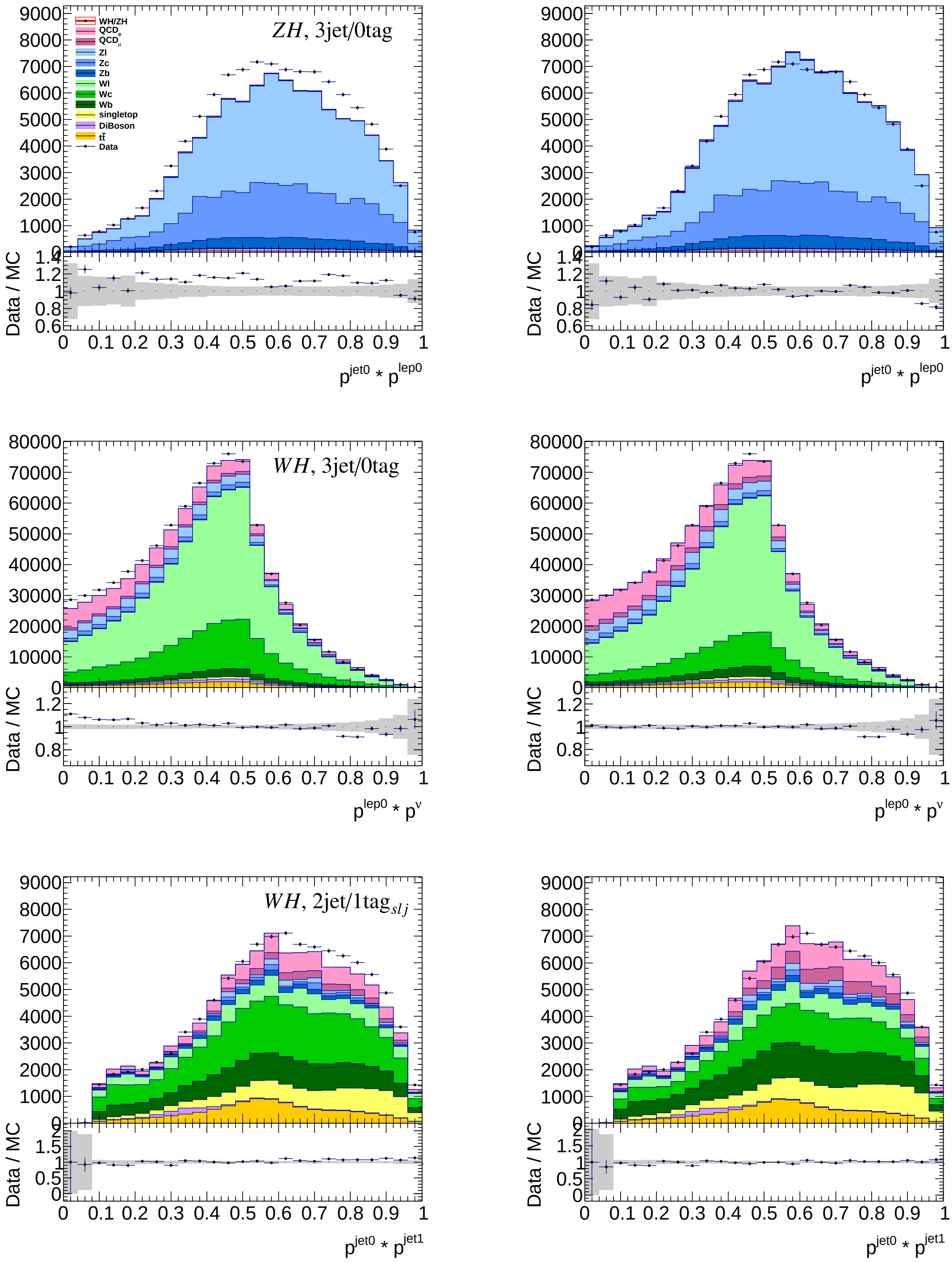


Figure 7.22: Distributions before (left) and after (right) the background normalization fit. Top row: $p^{jet0}{*}p^{lep0}$, 3jet/0tag, ZH pre-selection. Middle row: $p^{lep0}{*}p^{\nu}$, 3jet/0tag, WH pre-selection. Bottom row: $p^{jet0}{*}p^{jet1}$, 2jet/1tag$_{slj}$, WH pre-selection.

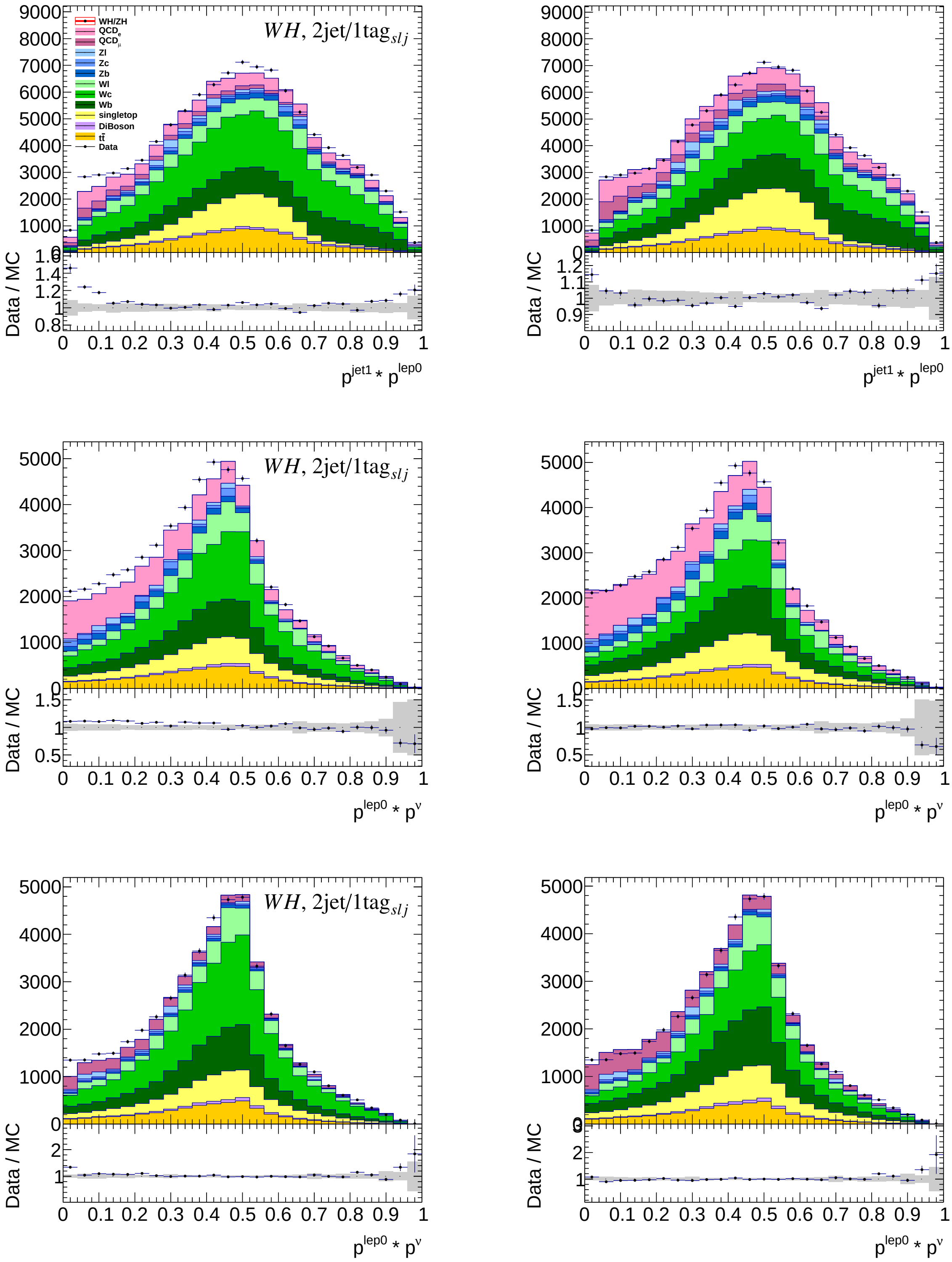


Figure 7.23: Distributions before (left) and after (right) the background normalization fit. Top row: $p^{jet1} {*} p^{lep0}$, 2jet/1tag$_{slj}$, WH pre-selection. Middle row: $p^{lep0} {*} p^{\nu}$ (electron only), 2jet/1tag$_{slj}$, WH pre-selection. Bottom row: $p^{lep0} {*} p^{\nu}$ (muon only), 2jet/1tag$_{slj}$, WH pre-selection.

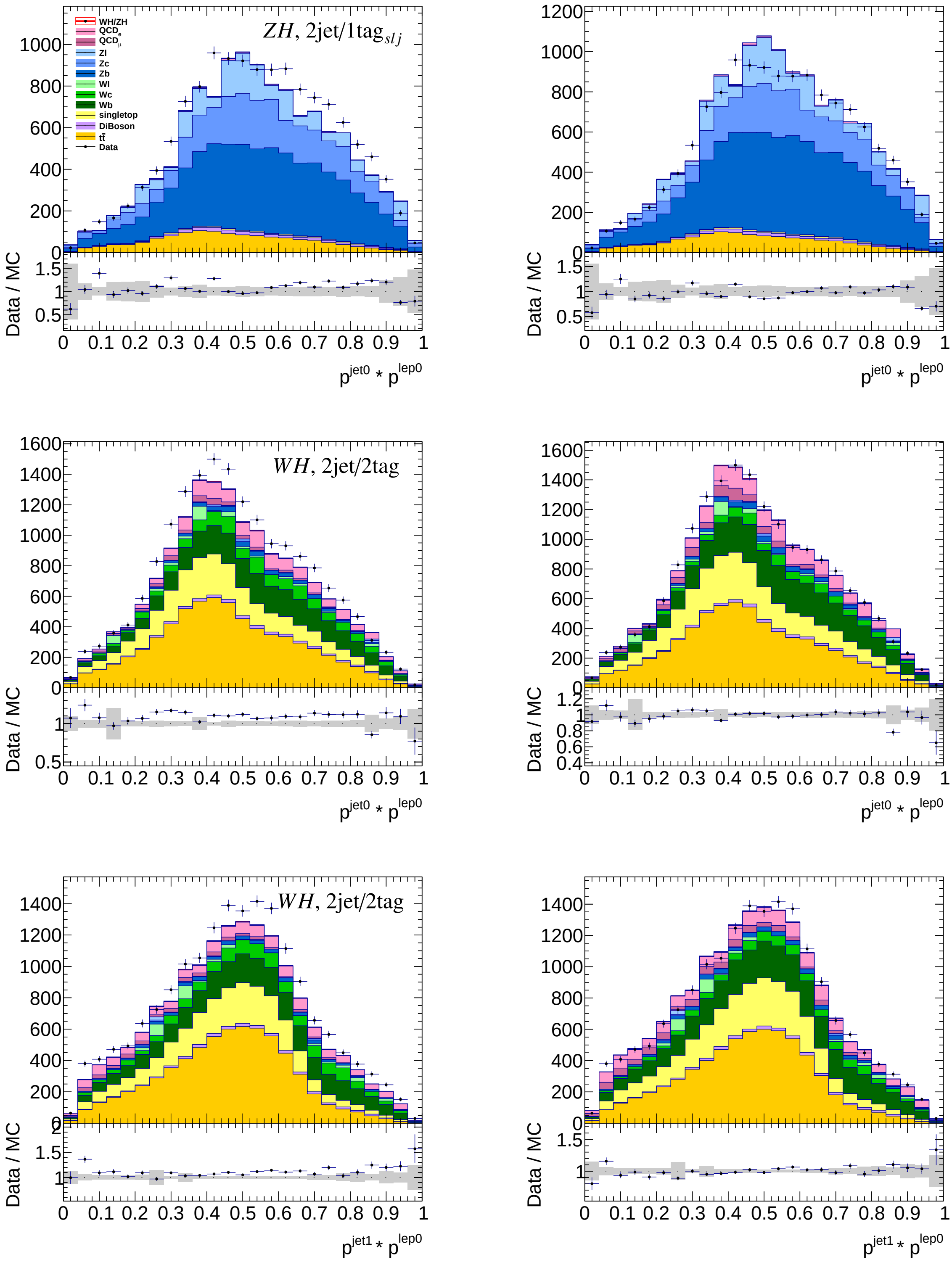


Figure 7.24: Distributions before (left) and after (right) the background normalization fit. Top row: $p^{jet0}{*}p^{lep0}$, 2jet/1tag$_{slj}$, ZH pre-selection. Middle row: $p^{jet0}{*}p^{lep0}$, 2jet/2tag, WH pre-selection. Bottom row: $p^{jet1}{*}p^{lep0}$, 2jet/2tag, WH pre-selection.

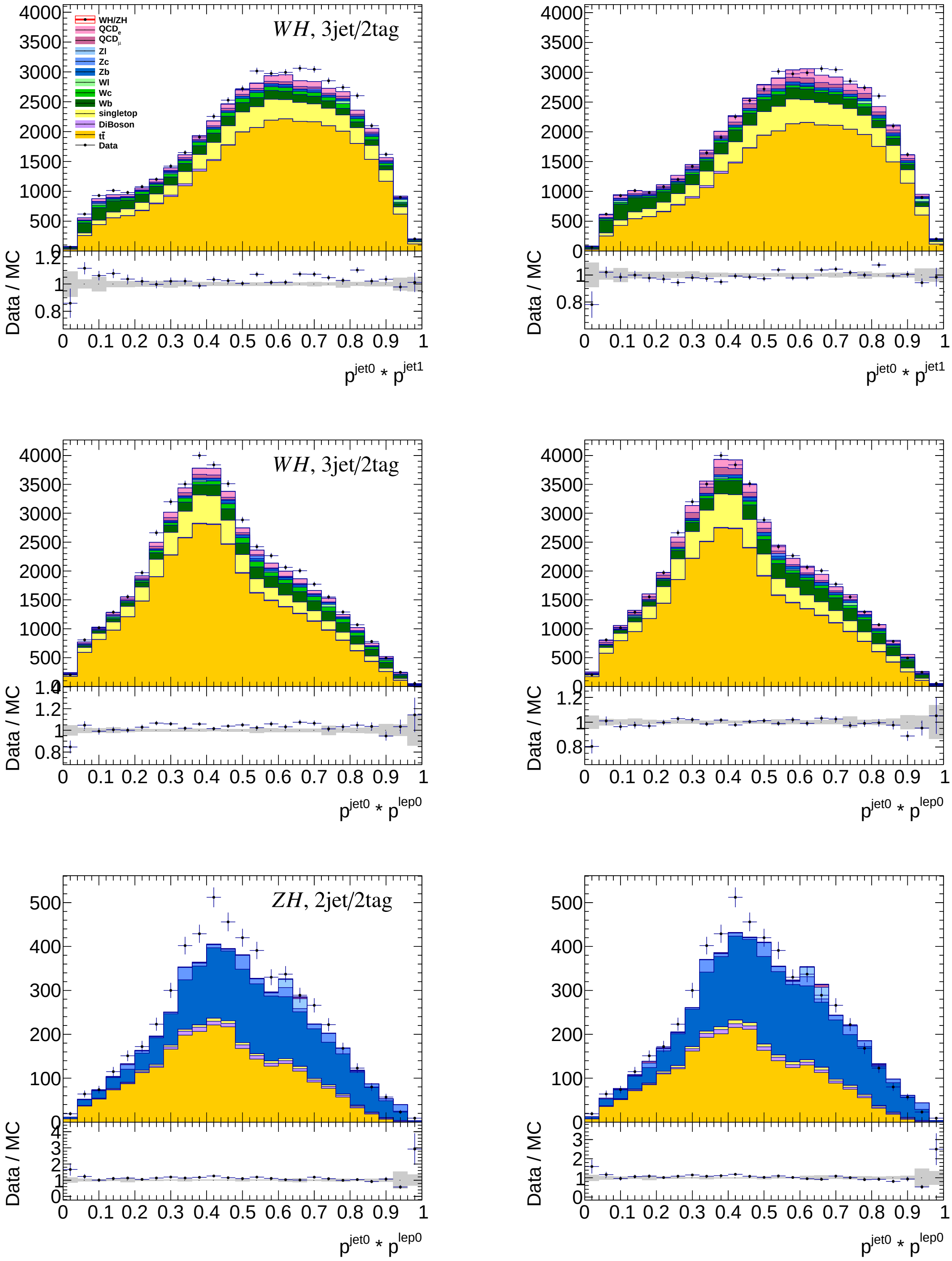


Figure 7.25: Distributions before (left) and after (right) the background normalization fit. Top row: $p^{jet0} {*} p^{jet1}$, 3jet/2tag, WH pre-selection. Middle row: $p^{jet0} {*} p^{lep0}$, 3jet/2tag, WH pre-selection. Bottom row: $p^{jet0} {*} p^{lep0}$, 2jet/2tag, ZH pre-selection.

7.10 Systematic uncertainties

A wide range of systematic uncertainties is considered in the analysis. Most of them are identical to the uncertainties considered in the most recently published ATLAS $VH \to Vb\bar{b}$ search [85] and are treated in exactly the same way to facilitate a fair performance comparison with the approach presented therein. Technically, all uncertainties enter the final limit fit in the form of additional nuisance parameters which are incorporated in the overall likelihood used to test the signal/background hypotheses. This procedure is described in Chapter 8, the names of the parameters are given below. In the following I will discuss the most important sources of systematic uncertainty and how they affect the shape and/or normalization of the BDT discriminant. The presentation is based on the latest ATLAS $VH \to Vb\bar{b}$ publication [85]. Three categories are differentiated: first, experimental uncertainties which affect the object properties and the accuracy with which these can be measured, second, background modeling uncertainties which account for possible data/MC disagreement, and third, theoretical uncertainties related to the signal process which mainly affect the signal cross-section and thus the expected number of signal events.

Plots illustrating the effect of all systematic uncertainties on the Monte Carlo prediction in the signal region can be found in Appendix D.

7.10.1 Experimental uncertainties

Measurement, event reconstruction and object identification all introduce various sources of systematic uncertainty which are subsumed in this category. As expected, the $VH \to Vb\bar{b}$ search depends mostly on the subset of uncertainties related to the jet energy scale and the b-tagging efficiency. Furthermore, different corrections applied to the physics objects (see Section 7.3) have associated uncertainties, too. Recommendations for the correct treatment of those are usually provided by the correspondent physics performance groups.

Pile-up and luminosity

To account for uncertainties associated to the modeling of additional proton-proton interactions (pile-up) in Monte Carlo, a variation of the expected number of pile-up interactions $< \mu >$ is performed, resulting in shape variations of the discriminating variables. The corresponding nuisance parameter is labeled **SysMuScale**.

The uncertainty on the integrated luminosity of the 2012 data affects the normalization of all background samples except multijet, which is directly taken from data. According to the calibration with beam-separation scans in November 2012 [108], the luminosity scale uncertainty is 2.8%.

Trigger

The trigger efficiencies for both muons and electrons are known to a precision of $< 1\%$, which is the result of dedicated tag-and-probe studies. The corresponding uncertainties are thus considered negligible.

Lepton identification and selection

Two different sources of uncertainty are considered for the lepton selection: the efficiency of the lepton identification and the efficiency of the isolation requirement. Both are obtained from tag-and-probe measurements using the Z mass peak and are on the order of 1%. Since the WH analysis uses a different isolation cut than the tag-and-probe study, an additional scale factor was introduced. The nuisance parameters associated with the lepton selection uncertainties are labeled **SysElecEffic**, **SysMuonEffic** and **SysLepIso**, respectively

Lepton corrections

The uncertainties associated with the lepton energy and resolution corrections discussed in Section 7.3.2 are determined from the Z mass line shape. Recommendations for their application are provided by performance groups in charge. The correspondent nuisance parameters are **SysMuonEResolMS**, **SysMuonEResolID** and **SysElecEResol**. All three of them have very little impact on the BDT discriminant.

Uncertainties related to the jet energy scale and resolution

The uncertainties related to the jet energy scale (JES) and resolution form the largest group of uncertainties in the analysis presented in this thesis. Most of them stem from the in-situ JES calibration analysis which includes studies of the Z+jet balance, the γ+jet balance and the multijet balance, see [109] for details. The following uncertainties are covered in these studies: differences between generator predictions from Alpgen, Herwig and Pythia, uncertainties associated with the selection, calibration and modeling of the physics objects involved and finally also uncertainties which arise from limited statistics in the calibration samples. The latter ones correspond to the nuisance parameters **SysJetStatX** (**X**= $1, 2, 3$) in this analysis, while the object-related uncertainties correspond to **SysJetModelX**, where **X** runs from one to four. In general, each uncertainty from the in-situ study falls in one of the following categories: detector description, physics modeling, detector/physics modeling mix, statistics and analysis method. Besides these, the following uncertainties are introduced:

SysJetEtaModel, SysJetEtaStat: Uncertainties associated with the modeling of additional radiation which affects the p_{T} and η distributions of the di-jet system. Referred to as "η-intercalibration" uncertainties.

SysJetNonClos: Uncertainty which accounts for the variations in the jet response when using JES calibration results based on different generators. Referred to as "MC non-closure term".

JetSysMixed1, JetSysMixed2: Uncertainties accounting for flavor and topology differences between the JES calibration samples. Evaluated by studying generators with different gluon/quark fragmentation functions.

JetSysHighPt: Specific uncertainty for high p_{T} jets ($p_{\mathrm{T}} > 1\,\mathrm{TeV}$) obtained from studies of single-particle energy depositions. Negligible in the analysis presented here.

SysJetPilePt, SysJetPileRho: Uncertainties related to pile-up effects.

SysJetFlavB: Uncertainty which accounts for the differences between the jet energy scales for b-jets and jets originating from other flavors. Obtained by comparing the b-jet response for different Monte Carlo generators.

JetEResol: The relative jet energy resolution (JER) ranges from $\sim 25\%$ at $20\,\mathrm{GeV}$ to $\sim 5\%$ at $1\,\mathrm{TeV}$. The associated uncertainty accounts for differences in this resolution between data and Monte Carlo. It is determined in in-situ studies using the di-jet balance and bisector methods. The effect on the analysis presented here is evaluated by smearing the jet p_{T} according to the uncertainty and then recalculating all observables.

SysJetFlavComp, SysJetFlavResp: Since the calorimeter response depends on whether a jet is gluon- or quark-initiated, the composition of a given sample will have an effect on the JES and JER that is obtained for it. This can be parametrized by the relative fraction of quark- and gluon-initiated jets in the sample and the associated uncertainty can be minimized if the sample composition is well known. As in [85], a 4% uncertainty is assigned to the composition of the relevant background samples, resulting in separate nuisance parameters for each physics process: **SysJetFlavComp_Diboson**, **SysJetFlavComp_Top**, **SysJetFlavComp_Wjets**, **SysJetFlavComp_Zjets**, **SysJetFlavResp_Diboson**, **SysJetFlavResp_Top**, **SysJetFlavResp_Wjets**, **SysJetFlavResp_Zjets**.

If not stated otherwise, all jet-related uncertainties are evaluated by means of a dedicated "JetUncertaintyTool" which is provided by the corresponding performance group. This tool allows for the modification of jet properties according to the different sources of uncertainty considered here. Each modification is propagated through the entire analysis, including object and event selection and the application of the BDT discriminant function. This way, the impact of each uncertainty on the distribution of w^{BDT} can be studied.

b-tagging uncertainties

The analysis presented here uses the same b-tagging calibration as presented in [85] which is justified by the conformity of the object definitions in both analyses. The calibration is obtained from $t\bar{t}$ events in 2012 data [92] and applied in the form of Monte Carlo scale factors which even out possible discrepancies between data and Monte Carlo. The systematic uncertainties associated with these scale factors correspond to the uncertainties of the calibration which are explained in [92]. Since the calibration itself is p_{T}^{jet} (and in some cases also η^{jet}) dependent, multiple nuisance parameters are introduced to account for the p_{T}-dependence of the corresponding uncertainties, following the eigenvector method introduced in [110]. For b-jets, the resulting parameters are labeled **SysBtagBXEffic** where **X** runs from 0 to 6, for c-jets they are labeled **SysBtagCXEffic**, **X**$= 0, \ldots, 5$. The parameter associated with the mistag efficiency is named **SysBTagLEffic**.

For the Monte Carlo samples where truth tagging is used, additional uncertainties are introduced which account for the fact that the efficiency maps which are used to calculate the truth tagging weight depend on the process and on the generator. These uncertainties are associated with the nuisance parameters **SysBTagSherpaBEffic** and **SysBTagSherpaCEffic** and correspond to an additional scaling uncertainty of 2% and 10% respectively.

Missing transverse energy

The $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement is mostly affected by the uncertainties associated with the other physics objects, since any changes on the vectorial sum of these will immediately result in a variation of $E_{\mathrm{T}}^{\mathrm{miss}}$. See also section 7.3.7 for a discussion of the re-evaluation of $E_{\mathrm{T}}^{\mathrm{miss}}$ in the presence of object corrections.

In addition to that, the scale and resolution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ contribution originating from unassociated calorimeter clusters ("soft terms") are varied within their uncertainties. This produces two additional nuisance parameters: **SysMETResoSoftTerms** and **SysMETScaleSoftTerms**.

7.10.2 Uncertainties related to the Monte Carlo background modeling

As discussed in Section 7.8.2, systematic discrepancies between data and Monte Carlo are observed in certain control regions and for certain Monte Carlo samples ($t\bar{t}$ and V+jets). To account for these modeling problems and for the uncertainties related to possible countermeasures like reweighting, a multitude of scale and shape variations is introduced and the effect on the final discriminant is carefully studied. Almost all modeling systematics that have been studied in [85] are adopted to the analysis presented here, possible modifications are detailed if necessary. The modeling systematics can be categorized as follows:

Normalization: Uncertainties related to the normalization of a given background sample. If the sample normalization is *not* floating in the final fit (see Chapter 8 for details), its uncertainty is taken from the corresponding cross-section calculation. For the diboson samples, this yields relative scaling uncertainties of $5 - 7\%$ and for the single-top samples $4 - 7\%$, depending on the production channel. As explained in Section 7.4, the multijet normalization is assigned an uncertainty of 30%. This is in agreement with the conservative approach taken in earlier studies [72], while the latest ATLAS publication [85] was done with a floating multijet normalization in the final fit.

3-jet/2-jet ratio: As discussed in Section 7.8, three-jet (and a four-jet) control regions are used in the analysis to constrain the normalization of the $t\bar{t}$ and single-top contributions. Since the predicted ratio of 2-jet events to 3-jet events in the $t\bar{t}$ sample differs between the generators Powheg, MC@NLO and Alpgen, an additional 5% uncertainty on the normalization in the 3-jet region (**Systtbar3JNorm**) is assigned. In the same way, the uncertainties labeled **SysStopWt3JNorm** (13%) and **SysStopst3JNorm** (9%) are introduced for the single-top samples. Finally, since the flavor composition of the V+jets contribution in the 3-jet control region differs between generators, one also has to introduce additional flavor-dependent scale factors for these samples to avoid undesirable tensions between the 3-jet region and the dedicated flavor-control regions. See [85] for details on the 2-jet/3-jet ratio studies.

$m_{b\bar{b}}$-modeling: Although $m_{b\bar{b}}$ is *not* the final discriminant used in the limit fit, it is still the most powerful variable in the BDT. Shape variations developed in [85], which are motivated by different generator predictions are thus adopted in the analysis presented here and propagated from $m_{b\bar{b}}$ to the BDT output distribution. For different background samples, different shape variations are used, details can be found in Table 7.3.

p_{T}^{V}-modeling: Analogous to $m_{b\bar{b}}$, the p_{T}^{V} distributions found in the $t\bar{t}$, single-top and diboson samples are compared among different generators and systematic uncertainties are assigned if required. As described in Section 7.8.2, a dedicated p_{T}^{V}-correction is applied to $t\bar{t}$. An extra uncertainty, defined as half the correction, is associated with this.

$\Delta\phi$-modeling: The $\Delta\phi$-correction proposed in [85] is *not* used in this analysis, as discussed in Section 7.8.2. However, the systematic variations associated with this correction are still applied, since they can be translated into a variation of the discriminating variables which accounts for the residual mismodeling even if the nominal distributions are not changed. To do so, $\Delta\phi$-dependent event weights are calculated and propagated through the entire analysis, resulting in a shape variation of the BDT output distribution.

7.10.3 Theoretical uncertainties on the signal cross section

The analysis presented here uses the latest signal cross-sections published by the LHC Higgs boson cross section working group [21], [29]. Associated with these numbers are several sources of uncertainty, which are taken into account via the respective nuisance parameters:

SysTheoryHbbBr: Relative uncertainty of the $H \rightarrow b\bar{b}$ branching ratio in the relevant mass range (3.3% for $m_H = 125\,\text{GeV}$).

SysTheoryWHScale, SysTheoryZHScale: Uncertainties related to the choice of the renormalization/factorization scale (~ 0.5%). More information about the calculation of these uncertainties is given in Section 3.4.1.

SysTheoryWHPDF, SysTheoryZHPDF: Uncertainty related to the parametrization of the parton distribution functions (~ 3.5%). Calculated according to the PDF4LHC recommendation [29].

SysTheoryWHEW, SysTheoryZHEW: Studies in [85] have shown that the size of the NLO electroweak correction to the signal cross section depends strongly on p_{T}^{V}. Therefore, a p_{T}^{V} dependent correction for the cross section provides a better description of the signal processes than the application of an inclusive scale-factor. Such a correction is derived in [85] and also used in the analysis presented here. The uncertainty associated with that correction is also p_{T}^{V} dependent and calculated on an event-wise basis. Details can be found in [85]

Background modeling uncertainties					
Process	Normalization	3jet/2jet ratio	p_{T}^{V}	$m_{b\bar{b}}$	$\Delta\phi$-correction
Wl	floating			**SysWMbb**	**SysWDphi** **SysWDphi3J**
Wc	floating	**SysWc3JNorm**			**SysWccDPhi** **SysWccDPhi3J**
Wb	floating	**SysWbb3JNorm**		**SysWbbMbb**	**SysWbDPhi** **SysWbDPhi3J**
Zl	floating			**SysZMbb**	**SysZDPhi** **SysZDPhi3J**
Zc	floating	**SysZc3JNorm**			**SysZcDPhi** **SysZcDPhi3J**
Zb	floating	**SysZbb3JNorm**		**SysZbbMbb**	**SysZbDPhi** **SysZbDPhi3J**
$t\bar{t}$	floating	**Systtbar3JNorm**	**SysTopPt** **SysTopPt3J**	**SysTopMbb** **SysTopMbb3J**	
singletop	**SysStopNorm**	**SysStop3JNorm**			
diBoson		**SysVV3JNorm**			
mutijet	**SysMJNorm**				

Table 7.3: Overview of the systematic uncertainties associated with background modeling and the corresponding nuisance parameter names. Normalization parameters which are floating in the global fit are marked with “floating”. Some nuisance parameters which are discarded in the final fit due to marginal effect have been omitted, see also Section 8.3.

"If your experiment needs a statistician, you need a better experiment."

ERNEST RUTHERFORD

CHAPTER 8

A fit model for limit setting

The goal of the statistical treatment presented in the following is to infer an upper limit on the Standard Model Higgs boson production cross-section and to measure the ratio μ of the observed signal strength to the Standard Model expectation. The corresponding hypothesis test is based on a binned maximum-likelihood fit which uses four different distributions: the w^{BDT} distribution in the signal region (2jet/2tag), which provides the sensitivity to the Higgs signal, and one distribution from each of the control regions, which serve to constrain the scale factors f_{bg} associated with the normalization of the background components that are floated in the fit (see Sections 8.2 and 8.3). In addition, the systematic uncertainties discussed in Section 7.10 are incorporated in the fit via a set of nuisance parameters θ, which means that the likelihood employed in the fit is a multi-parameter function of the form $\mathcal{L}(\mu, f_{bg}, \theta|\mathbf{x})$, where $\mathbf{x}$ denotes the data.

8.1 The profile likelihood method

Setting an upper limit on the Higgs boson production cross section is a particular variant of hypothesis test. The hypothesis H_0 is represented by the Standard Model without the Higgs boson, while the alternative hypothesis H_1 is represented by the Standard Model containing the Higgs Boson and a defined Higgs boson mass m_H. The specification of the mass in H_1 is necessary to facilitate a measurement of the signal strength parameter μ: while the expected signal strength depends strongly on the Higgs boson mass, the sensitivity of the analysis to the mass is poor due to limited resolution. The hypothesis test is based on the likelihood function $\mathcal{L}(\mu, \theta)$, where θ denotes additional nuisance parameters. Following the Neyman-Pearson lemma [74], an optimal test statistic is given by the likelihood ratio:

$$\lambda = \frac{\mathcal{L}(H_1)}{\mathcal{L}(H_0)} \tag{8.1}$$

Using toy Monte Carlo, one can determine the distributions of the test statistic λ under the two hypotheses and compare the observed value λ_{obs} to these distributions. Exclusion of the hypothesis H_1 is usually announced if the associated (one-sided) p-value is smaller than 5% - this corresponds to a rejection of H_1 at the 95% confidence level. The problem with this theoretically simple approach is that it is not straightforward to deal with additional nuisance parameters in the likelihood function: strictly speaking, a hypothesis can only be rejected if the corresponding test is carried out for all possible values of θ, regardless of the fact that the true value of θ is usually not of interest for the measurement. Obtaining a valid statement about H_1 irrespective of θ can be achieved by marginalizing with respect to θ in a quasi-Bayesian fashion. This approach is known as the hybrid likelihood approach which was proposed by Cousins and Highlands. However, carrying out the integration is computationally expensive.

A more effective treatment of the nuisance parameter problem is the *profile likelihood* method: a new test statistic $\lambda(\mu)$ is defined

$$\lambda(\mu) = \frac{\mathcal{L}(\mathbf{x}|\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\mathbf{x}|\hat{\mu}, \hat{\theta})} \tag{8.2}$$

where $\hat{\mu}$ and $\hat{\theta}$ denote the global best fit values for μ and θ given the observed data $\mathbf{x}$, and $\hat{\hat{\theta}}(\mu)$ denotes the best fit value for θ that can be found for a fixed value of μ. Note the difference to the first approach: while the likelihood ratio defined in Eqn. 8.1 tests H_1 against H_0, $\lambda(\mu)$ tests either of them against the hypothesis preferred by the data, for which $\mu = \hat{\mu}$. This construction has some very desirable properties:

1. As expected, the profile likelihood as a function of μ is broadened with respect to the original likelihood function with fixed θ. This reflects the loss of information introduced by nuisance parameters.

2. The profile likelihood construction fulfills the conditions of Wilk's theorem [111], which states that the distribution of the quantity

$$q_\mu = -2\ln\frac{\mathcal{L}(\mathbf{x}|\hat{\theta}_1, H_1)}{\mathcal{L}(\mathbf{x}|\hat{\theta}_0, H_0)} \tag{8.3}$$

   asymptotically, i.e. for a sufficiently large data sample, approaches a χ^2 distribution with a number of degrees of freedom that is equal to the difference between the number of free parameters in $\mathcal{L}(\mathbf{x}|H_1)$ and $\mathcal{L}(\mathbf{x}|H_0)$. Here, $\hat{\theta}_0$ and $\hat{\theta}_1$ denote the maximum likelihood estimates of θ for the respective hypothesis. Applied to the profile likelihood, this means that $q_1 = -2\ln(\lambda(\mu = 1))$ follows a χ^2 distribution with one degree of freedom under H_1.

The second property assures that the profile likelihood approach can handle the presence of nuisance parameters without the need for numerical integration. Instead, one can simply compute the desired exclusion significance for a given signal strength μ via

$$Z_\mu = \sqrt{q_\mu} = \sqrt{-2\ln(\lambda(\mu))} \tag{8.4}$$

Several variations of the basic profile likelihood concept which are tailor-made for the LHC experiments have been proposed and asymptotic approximations which extend Wilk's theorem accordingly have been computed. These are discussed in [112].

8.2 Distributions used in the fit

In addition to the w^{BDT} distribution in the signal region, which delivers the largest sensitivity to the possible presence of a Higgs signal, one distribution from each of the control regions is used in the fit. As discussed in Section 7.8, this allows for a measurement of the flavor composition in the V+jets background which is difficult to model. Moreover, it also helps to fix the normalization of the $t\bar{t}$ background in the specific phase-space considered for this analysis where one W boson escapes detection (WH channel). While the 2jet/1tag$_{slj}$ region and the 3jet/2tag region can be used to provide shape information, the 3jet/0tag region is only employed to provide an overall background normalization which mostly helps to fix the V+light background fraction. The input distributions used in the WH channel are shown in Figure 8.1, an overview is presented in Table 8.1. While the w^{BDT} distributions are obviously chosen for their sensitivity to the possible presence of a Higgs signal, the $p^{jet0} * p^{lep0}$ distribution in the 2jet/1tag$_{slj}$ region was found to provide the best constraint on the single-top normalization among the observables considered.

In principle, the fit model presented here was designed to allow for a combined fit of the WH and the ZH channel: all regions/distributions are present in both channels, which makes for a total of 8 distributions. However, for the limit-fit this thesis will concentrate on the WH channel only. The ZH channel has not been fully optimized and a robust combination demands a more detailed study.

Fit input distributions		
	2 jet	3 jet
0 b-tags	-	norm only
1 b-tag (sub-leading jet tagged)	$p^{jet0} * p^{lep0}$	-
2 b-tags	w^{BDT}	w^{BDT}

Table 8.1: Input distributions/regions used in the limit-fit.

8.3 Floating normalizations and nuisance parameters

The use of dedicated control regions allows to determine the final normalization of the most important background contributions for the WH channel from data - in total, four parameters associated with sample normalizations are floated in the fit: **norm_Wl**, **norm_Wc**, **norm_Wb**, **norm_ttbar**. The normalizations associated with the main backgrounds in the ZH channel (**norm_Zl**, **norm_Zc**, **norm_Zb**) are treated as additional nuisance parameters with priors reflecting the uncertainties of the respective cross-sections. A combined fit of both channels would allow to determine those from the data, too.

Contrary to the strategy followed in [85], where the normalization of the multijet template is floated independently for each b-tag and jet multiplicity, a consistent treatment of the multijet normalization across all regions is adopted for the analysis presented here: the normalization is initialized to the value obtained from the fit described in Section 7.9 and associated with a nuisance parameter with a log-normal prior distribution to avoid unphysical negative values.

The other nuisance parameters in the fit control the systematic uncertainties described in Section 7.10. While some of these uncertainties only affect the normalization of one or more samples, others also affect the shape of the distributions. For the distributions used in the fit, it is thus fair to say that the content of each bin is a function of the nuisance parameters θ, although the bins are of course correlated. In the fit, the parameters θ are allowed to vary according to their prior distributions: parameters which are associated with uncertainties that change the normalization are assigned a log-normal prior (which is only defined for positive values) and parameters associated with shape variations are assigned a Gaussian prior. The width of each prior is chosen according to the respective size of the associated uncertainty. The statistical uncertainty of the Monte Carlo prediction is also taken into account by assigning a nuisance parameter to each bin for which the relative statistical uncertainty is larger than 5%. Systematic uncertainties that affect both the shape and the normalization of a given distribution are split into two separate sources of uncertainty and two parameters are assigned. If either of the two causes a variation of less than 0.5% it is discarded. For any parameter θ_i the optimal value found in the fit

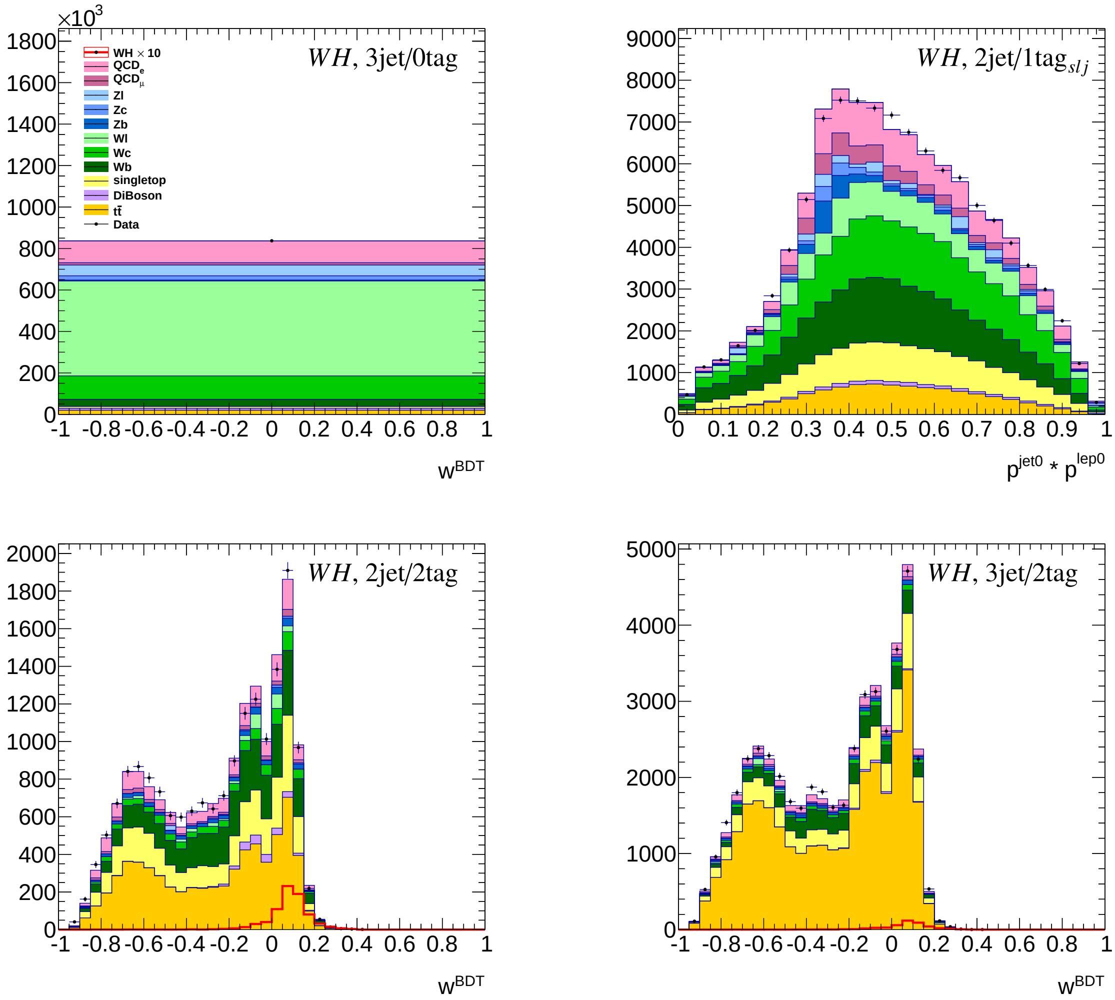


Figure 8.1: The distributions used in the limit fit for the WH channel.

may differ from the nominal value preferred by the prior - within the given uncertainties this is to be expected, but large deviations may indicate a mismodeling in the simulation which is not fully corrected for or which is not covered by the corresponding systematic uncertainty. To spot possible problems, so-called "pull-plots" are generated, where the post-fit deviation of the nuisance parameters from their initial values is depicted relative to the associated uncertainties. These pull distributions also allow to assess the validity of the size of the systematic uncertainties assigned by comparing the post-fit errors of the nuisance parameters to the pre-fit errors. See Section 9.1 for details.

8.3.1 Correlations among the nuisance parameters

While some sources of uncertainty are common to all processes in the analysis and may thus be parametrized consistently across all distributions and samples, others may for example depend on the flavor composition or the object multiplicity and should thus be treated differently depending on the input region. The scheme adopted here follows the reasoning presented in [85]:

Systematic uncertainties related to jet flavor composition and response are treated as uncorrelated across four different groups of processes since the flavor composition may depend on the process. The four groups are W+jets, Z+jets, $t\bar{t}$/single-top and diboson/VH. The systematic uncertainty associated with the top p_T modeling is split into two parameters: one for each jet multiplicity in the WH channel. The $\Delta\phi$ systematic is parametrized separately for each of the six V+jets samples and both jet multiplicities, resulting in twelve nuisance parameters. All other systematic uncertainties are treated as fully correlated across all regions and samples, for more details see [85].

“The most exciting phrase to hear in science, the one that heralds the most discoveries, is not "Eureka!" (I found it!) but "That’s funny.."”

ISAAC ASIMOV

CHAPTER 9

Results

9.1 Global fit quality

9.1.1 Behavior of nuisance parameters

Using the fit model described in Section 8.2, the floating Monte Carlo scale factors and all nuisance parameters are adjusted to give the best fit to the data. The resulting nuisance parameter pull distributions are shown in Figures 9.1- 9.7. While most of the nuisance parameters are well-behaved in the fit in the sense that they stay close to the value assumed in the corresponding prior, some display a sizable deviation from their pre-fit value and/or are strongly constrained after the fit:

SysBTagLEffic: This parameter is associated with the b-tagging efficiency for light jets. The fact that it is shifted downwards can be understood on closer examination of the W+light background component in the 2jet/2tag and 3jet/2tag regions and its behavior under variation of the relevant parameter (see Figure 9.8): the effective luminosity of the W+light samples is much lower than the data luminosity, which is compensated by assigning large weights to the events. As a consequence, the few events which penetrate into the signal region create "spikes" in the w^{BDT} distribution. Lowering the b-tagging efficiency for light jets, the spikes are attenuated which results in a smoothened distribution. Given that the W+light component is a minor background, the pull of this parameter is considered tolerable in this thesis, but a follow-up study with a larger W+light sample would be needed to settle this question completely.

SysWMbb: This parameter only affects the W+light background, introducing a shape variation in the m_{jj} distribution which in turn changes the shape of the BDT output. Since the effect of this variation on the Monte Carlo prediction for W+light is only visible in the regions with two b-tags (see Figure 9.9), it is very likely that the observed behavior is another artifact resulting from the low statistics of the W+light sample after b-tagging. As with **SysBTagLEffic**, the observed behavior is considered tolerable in the study presented here.

SysStopNorm: This parameter affects the normalization of the single-top background, which is enhanced after the fit. The region most affected by the variation of this parameter is the 3jet/2tag region, where a 1σ deviation of the single-top normalization corresponds to a 2% change in the overall background normalization. The observed pull is a result of a delicate interplay between the regions: the reduced b-tagging efficiency for light jets smoothens the distributions in the 2jet/2tag and 3jet/2tag regions. At the same time, it affects the background normalization in the 2jet/1tag$_{slj}$ region (but leaves the 3jet/0tag region intact). Since the $W + c$ normalization is constrained by the 3jet/0tag region, the $W + b$ background must account for the change in the 2jet/1tag$_{slj}$. This is indeed the case, the $W + b$ normalization is enhanced (see the discussion of the floating background normalizations in Section 9.1.2 for details). The increased amount of $W + b$ background is compensated by a reduction of the $t\bar{t}$ background in the 2jet/2tag region. Finally, to maintain a proper normalization in the 3jet/2tag region, the normalization of the single-top background must increase. The fact that certain background normalizations are correlated in such a delicate way is reflected in the large uncertainties of these parameters (Section 9.1.2).

SysTopMbbLo: This parameter varies the shape of the $m_{b\bar{b}}$ distribution in the $t\bar{t}$ background based on generator comparisons. A low value of this parameter as it is observed after the fit corresponds to a broadening of the $m_{b\bar{b}}$ distribution. This broadening effectively moves $t\bar{t}$ events away from the Higgs mass peak and thus also from the high w^{BDT} region to the low w^{BDT} region. This effect is most pronounced in the 2jet/2tag region where it affects the overall background normalization in the left tail of the BDT output by up to 5% (see Figure 9.11). Since the data in this region is slightly above the Monte Carlo prediction (pre-fit), a low value of this parameter is preferred by the fit.

SysZccNorm: This parameter controls the normalization of the $Z+c$ background, which only makes a sizable contribution in the 2jet/1tag$_{slj}$ region. In this region, the $Z+c$ background has a distinctive peak-like shape (see Figure 9.12). Reducing the height of this shape gives a better description of the data, as can be seen from a comparison with the top right plot in Figure 8.1.

For the other nuisance parameters which remain close to their nominal values but are strongly constrained by the fit (**SysJetFlavB**, **SysJetFlavComp_Wjets**, **SysJetFlavResp_Wjets**, **SysMETResoSoftTerm**), an inspection of the fit to the Asimov data set [113] created from the Standard Model background expectation reveals that their pre-fit uncertainties are overestimated for this particular application. This applies in particular to the normalization of the multijet background (**SysMultijetNorm**) which proves that the 30% uncertainty assigned to the normalization of the corresponding templates is very conservative.

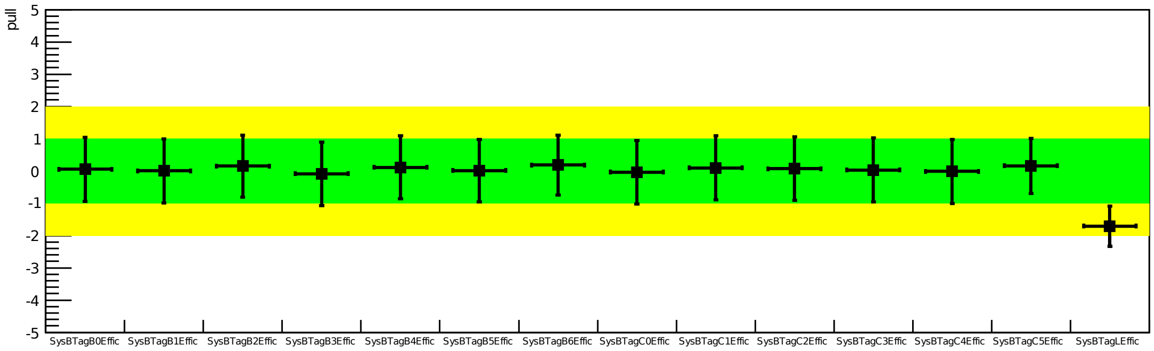


Figure 9.1: Overview of the post-fit nuisance parameter pulls associated with the b-tagging uncertainties obtained from the global fit in the WH channel.

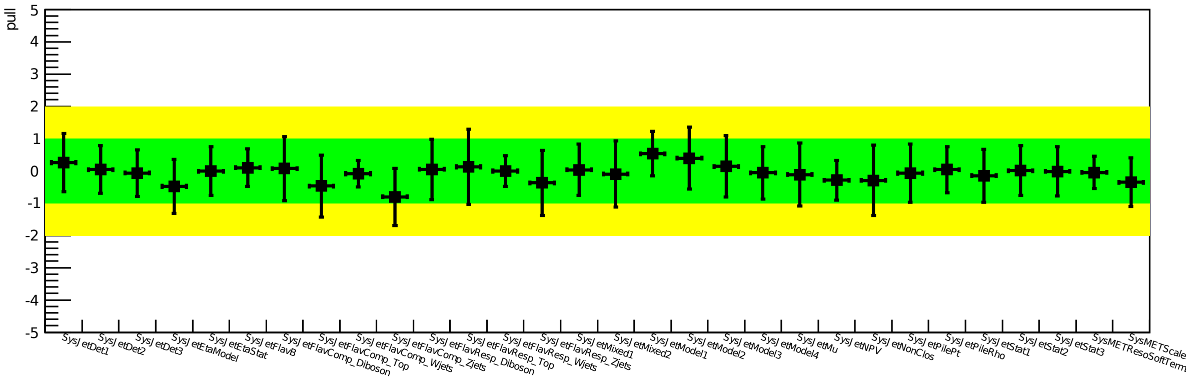


Figure 9.2: Overview of the post-fit nuisance parameter pulls associated with the jet energy scale uncertainties obtained from the global fit in the *WH* channel.

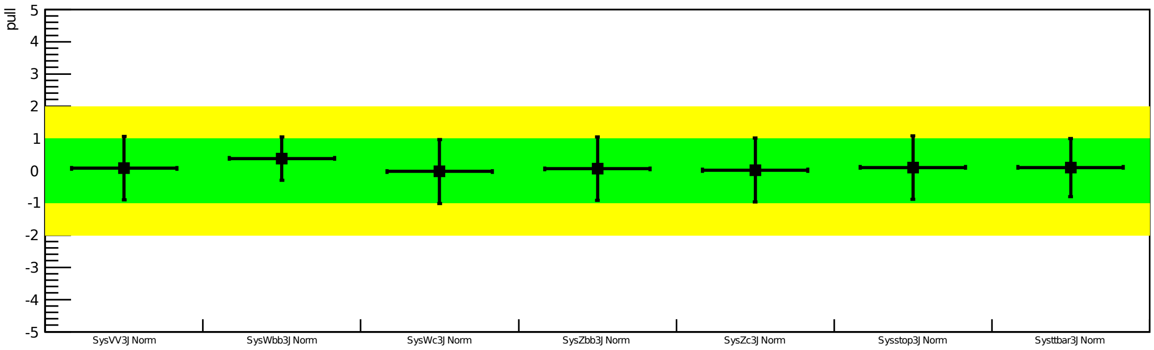


Figure 9.3: Overview of the post-fit nuisance parameter pulls associated with the extrapolation between 3jet and 2jet regions obtained from the global fit in the *WH* channel.

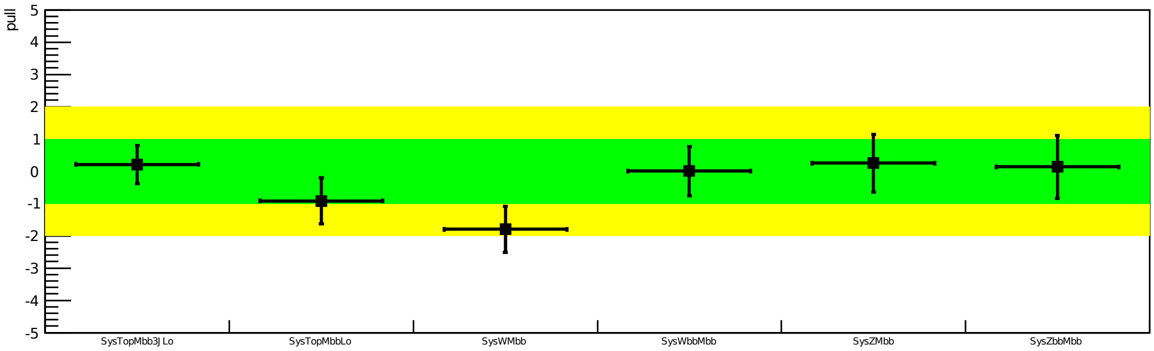


Figure 9.4: Overview of the post-fit nuisance parameter pulls associated with the modeling of the $m_{b\bar{b}}$ distribution obtained from the global fit in the *WH* channel.

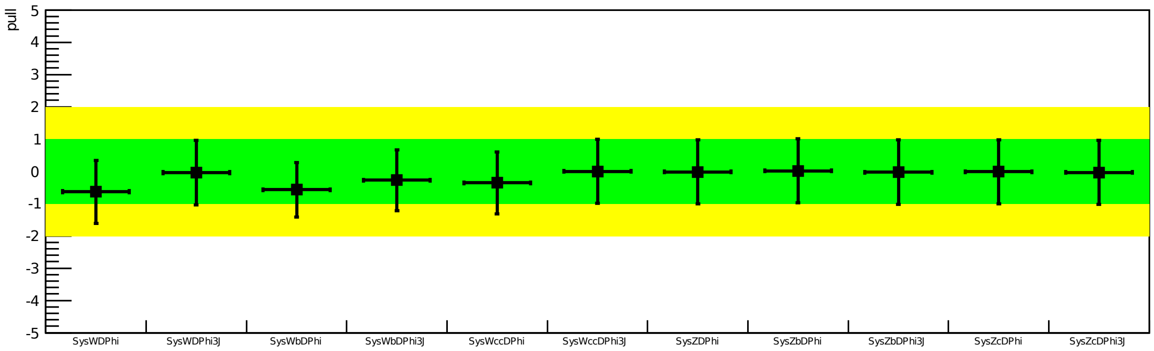


Figure 9.5: Overview of the post-fit nuisance parameter pulls associated with the modeling of the $\Delta\Phi(jet0, jet1)$ distribution obtained from the global fit in the WH channel.

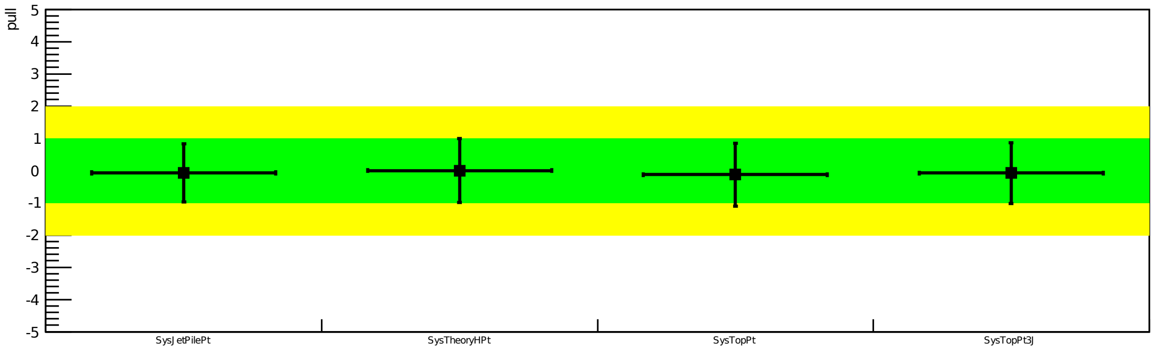


Figure 9.6: Overview of the post-fit nuisance parameter pulls associated with the modeling of the jet and vector boson p_{T} distributions obtained from the global fit in the WH channel.

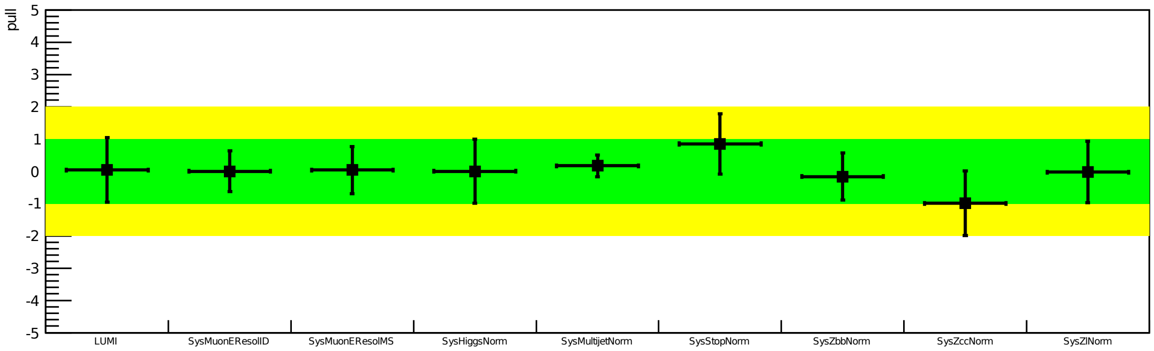


Figure 9.7: Overview of the post-fit nuisance parameter pulls not associated with any of the other categories obtained from the global fit in the WH channel.

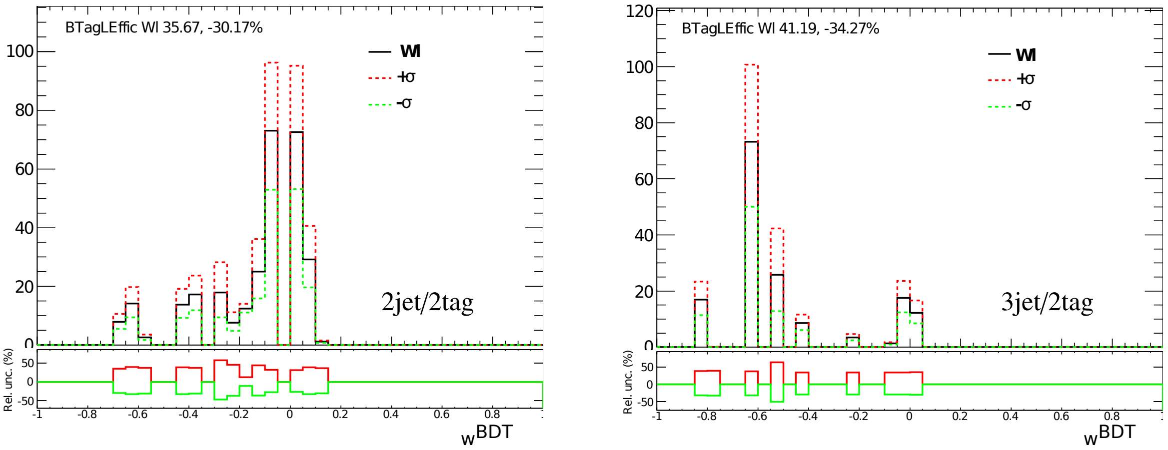


Figure 9.8: The remaining W+light Monte Carlo in the 2jet/2tag and 3jet/2tag regions and its behavior under variation of the **SysBTagLEffic** nuisance parameter.

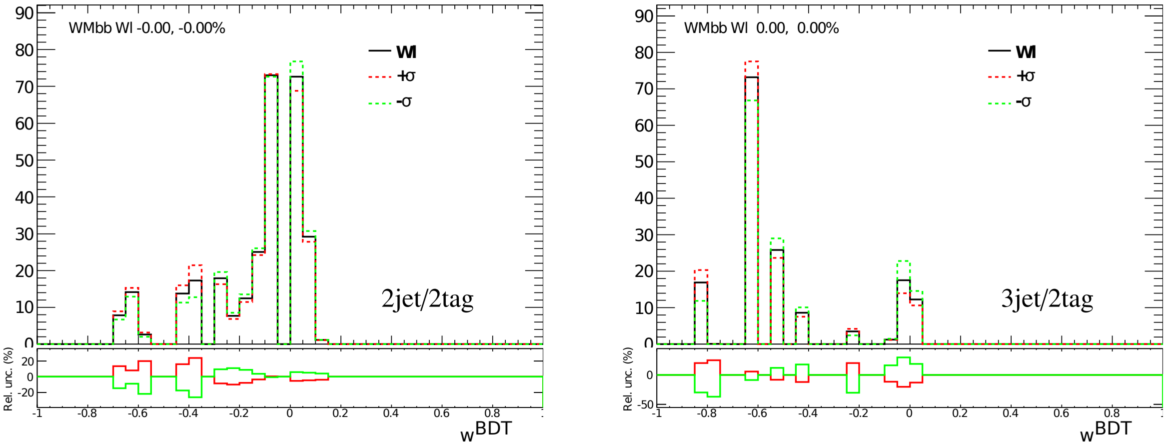


Figure 9.9: The remaining W+light Monte Carlo in the 2jet/2tag and 3jet/2tag regions and its behavior under variation of the **WMbb** nuisance parameter.

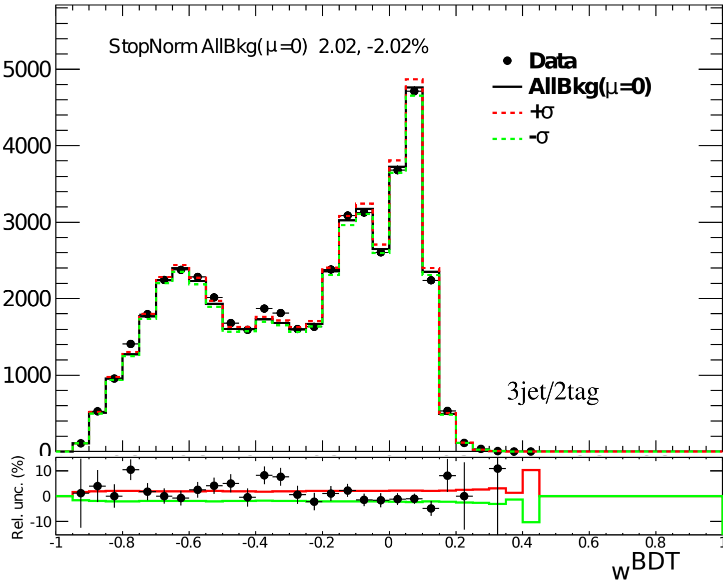


Figure 9.10: The change in the overall background normalization in the 3jet/2tag region under variation of the **StopNorm** nuisance parameter.

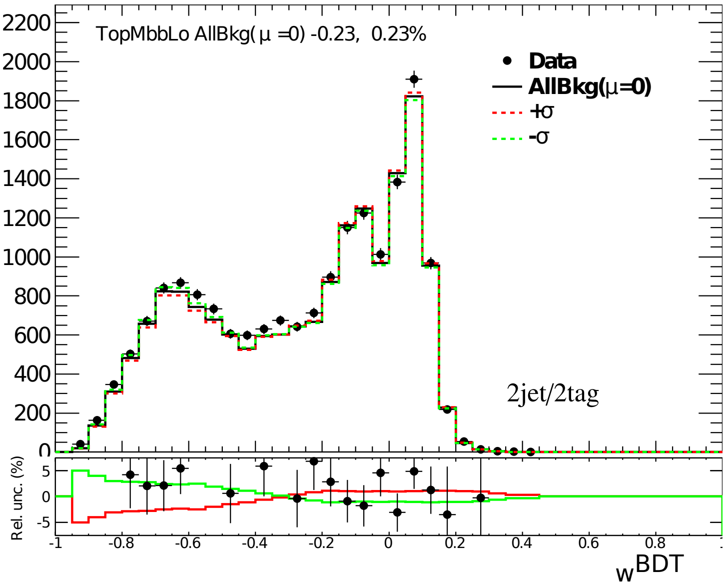


Figure 9.11: The change in the overall background shape in the 2jet/2tag region under variation of the **TopMbbLo** nuisance parameter.

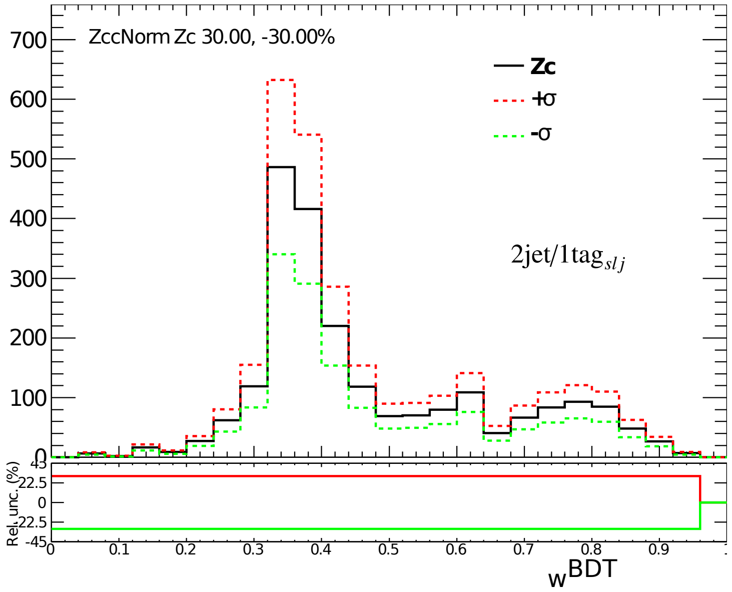


Figure 9.12: The change in the $Z + c$ background normalization in the 2jet/1tag$_{slj}$ region under variation of the **ZccNorm** nuisance parameter.

9.1.2 Post-fit background scale factors and data/MC agreement

As described in Section 8.3, the normalizations of the most important background processes for the WH channel are left floating in the fit. The scale factors which are obtained are listed in Table 9.1, where also the corresponding scale factors resulting from the multijet pre-fit are shown for comparison. No uncertainties are quoted for the latter since the multijet pre-fit uses several statistically correlated distributions, see Section 7.9 for details. It should be noted that only the multijet normalization, which is assigned an ad-hoc uncertainty of 30%, is propagated from the multijet-fit to the profile-likelihood fit. All other scale factors are reverted to their nominal values and left floating in the profile-likelihood fit. Taking into account that the multijet pre-fit does not include any nuisance parameters to cope with the systematic uncertainties, a good agreement between the two fit results is observed.

It is clear that the uncertainties associated with the background Monte Carlo scale factors have an immediate impact on the sensitivity of the analysis. The expected limit (see Section 9.2) is in particular affected by the precision with which the $W+b$ normalization can be determined. Even with the 3jet/2tag control region, there remains an ambiguity between the $W + b$ and $t\bar{t}$ scale factors which increases the corresponding uncertainties. Further constraining the $t\bar{t}$ background with a 4jet/2tag region has been tested but this approach was finally discarded due to additional difficulties in the more complicated extrapolation across three regions of different jet multiplicity. Including the ZH channel in the global fit would provide further $t\bar{t}$ control regions which are almost free of W+jets background and would thus help to constrain the Monte Carlo scale factors further.

Figure 9.13 shows a comparison of the data to the Monte Carlo prediction using the nuisance parameter values and scale factors resulting from the maximization of the profile likelihood in the global fit. Excellent agreement is observed. The dashed blue line indicates the pre-fit Monte Carlo expectation with all nuisance parameters and scale factors fixed to their nominal values. A list of the post-fit event counts for data and all Monte Carlo samples can be found in Appendix A.

Monte Carlo scale factors		
Background	multijet pre-fit	global fit
W + light	1.05	1.00 ± 0.08
$W + c$	0.71	0.79 ± 0.13
$W + b$	1.30	1.41 ± 0.16
$t\bar{t}$	0.99	0.94 ± 0.10

Table 9.1: Monte Carlo scale factors obtained from the global fit and the multijet pre-fit.

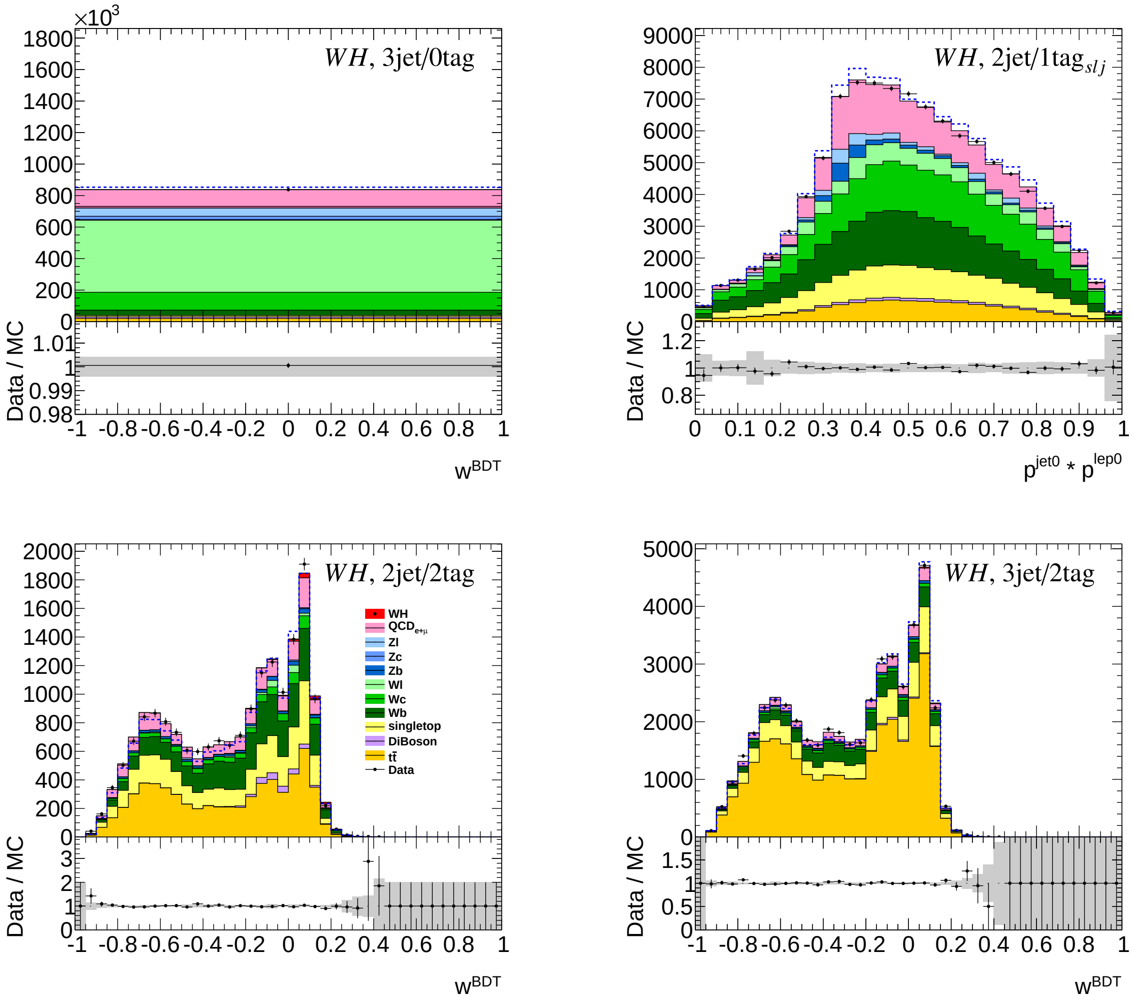


Figure 9.13: The distributions used in the global fit (WH channel) with all post-fit scale factors and optimized nuisance parameter values applied.

9.2 Upper limit on the Higgs boson production cross-section and measured signal strength

Using the profile-likelihood approach introduced in Section 8.1, a one-sided upper limit on the Standard Model Higgs boson production cross section is calculated at the 95% confidence level (CL) for different values of m_H. The result is shown in Figure 9.14, the corresponding numbers are detailed in Table 9.2. For $m_H = 125\,\text{GeV}$, the expected limit in the absence of signal is $2.06 \times \sigma_{\text{SM}}$, while the observed limit is $3.15 \times \sigma_{\text{SM}}$. The fitted value of the signal strength parameter at this mass is $\mu = 1.40 \pm 0.95$. This excess corresponds to a local p_0 of 0.14 (1.10σ) for obtaining a result at least as signal-like as observed in the absence of a signal. It is thus *not* significant and no observation of a signal can be claimed. The corresponding expected value of p_0 in the presence of a Standard Model Higgs boson at that mass is 0.15 (1.04σ). A comparison of the observed/expected local p_0 for other values of m_H is presented in Figure 9.15. The seemingly contradictory trends in the curves for observed and expected p_0 can be understood in the following way: the branching fraction for $H \to b\bar{b}$ and thus the expected signal strength decreases rapidly towards higher m_H. Therefore the sensitivity of the analysis is reduced as the mass of a hypothetical Higgs boson is increased, which is reflected by the curve which represents the expected p_0 in Figure 9.15. If on the other hand an excess is observed in the $H \to b\bar{b}$ channel, it is visible across the entire mass range considered for this analysis due to the poor $b\bar{b}$ mass resolution of $\sim 10\%$ (see Section 7.3.6). Interpreting this excess under different mass hypotheses will now result in a significance which *increases* with the mass since the corresponding signal expectation *decreases* at the same time. This behavior is demonstrated by the curve which represents the observed p_0 in Figure 9.15.

9.2.1 Extrapolation of the Monte Carlo statistics

During the preparation of this thesis, the statistics of the used Monte Carlo samples was increased. While the latest ATLAS publication [85] already made use of the larger samples, they could not yet be incorporated in this thesis. As a consequence, the result reported here suffers from an additional systematic uncertainty compared to the cut-based ATLAS result due to the limited MC statistics. Extrapolating the statistical uncertainties assigned to the Monte Carlo to the full sample size employed in [85], the expected limit at $m_H = 125\,\text{GeV}$ decreases to $1.88 \times \sigma_{\text{SM}}$ in the WH channel, which corresponds to an improvement of 11.3% over the result ($2.12 \times \sigma_{\text{SM}}$) presented in [85]. Furthermore, recent studies with larger Monte Carlo samples indicate that also the performance of the boosted decision trees can be improved by $\sim 10\%$ with larger training samples. It is thus fair to say that the method presented in this thesis has the potential to improve on the results presented in [85] by 10 – 20%.

Expected and observed upper limits on σ^H_{SM}						
Higgs mass	-2σ	-1σ	**exp.**	$+1\sigma$	$+2\sigma$	**obs.**
115	0.74	1.00	1.39	1.93	2.59	1.35
120	0.92	1.24	1.72	2.39	3.21	1.89
125	1.10	1.48	2.06	2.86	3.84	3.15
130	1.28	1.71	2.38	3.31	4.43	4.20
135	1.38	1.86	2.58	3.59	4.81	4.98
140	2.27	3.05	4.23	5.89	7.90	6.86

Table 9.2: Upper limit on the Standard Model Higgs boson production cross section obtained in the WH channel

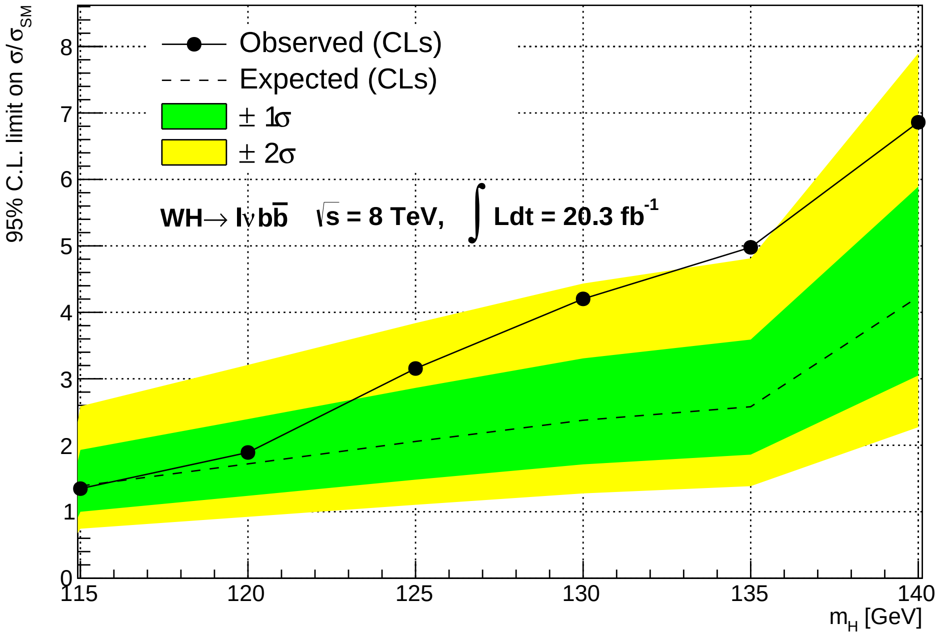


Figure 9.14: The upper limit on the Standard Model Higgs boson production cross section obtained in the WH channel as a function of the Higgs boson mass m_H.

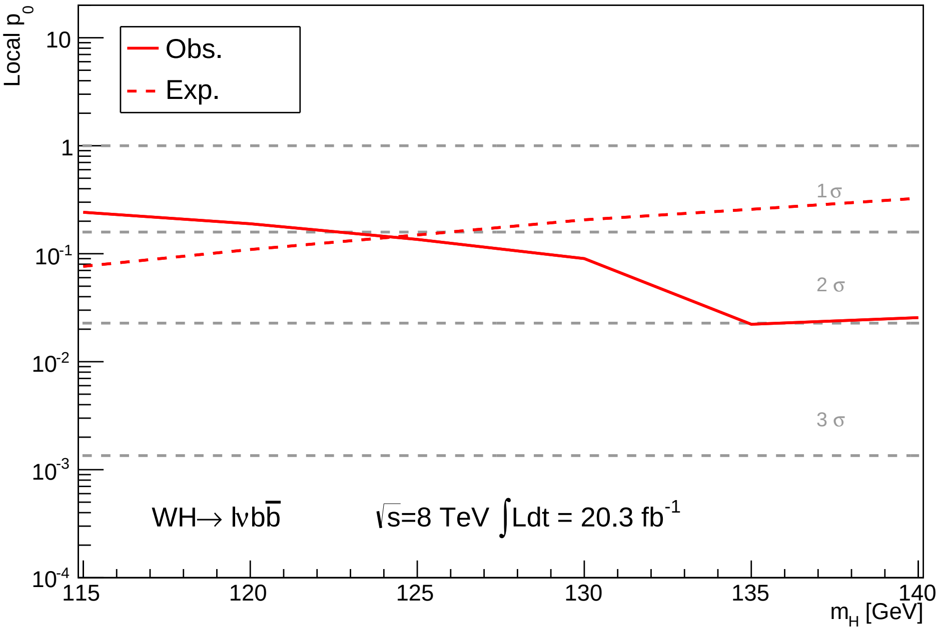


Figure 9.15: The observed local p_0 value obtained in the WH channel as a function of the Higgs boson mass m_H. The dashed line indicates the expected local p_0 value under the hypothesis of a Standard Model Higgs signal.

“No book can ever be finished. While working on it we learn just enough to find it immature the moment we turn away from it”

KARL POPPER

CHAPTER 10

Summary

In this thesis, a novel search for the Standard Model Higgs boson in the decay channel $H \to b\bar{b}$ is presented, using the entire data-set recorded by the ATLAS experiment in 2012 amounting to 20.3 fb^{-1} of data. The search is focused on Higgs production in association with a leptonically decaying W or Z boson to allow for an effective event selection based on high p_{T} lepton triggers.

The analysis makes use of a machine learning approach based on boosted decision trees. First, this algorithm is used to guide the construction of a set of pre-selection cuts which provide high signal efficiency and good multijet rejection. Second, it is used to infer an optimal discriminant function which serves as the input to the final hypothesis test from which an upper limit on the Higgs boson production cross-section is obtained.

The boosted decision trees which are used to compute the final discriminant function operate on a variable set which is constructed from Lorentz invariant quantities and angles, providing a minimal, robust and physically interpretable parametrization of the VH system.

The complex background composition in the $H \to b\bar{b}$ decay channel is controlled by dedicated control regions of different jet and b-tag multiplicities which provide constraints on the normalization of the V+jets and $t\bar{t}$ background. Although the event selection is designed to provide a good multijet rejection, a non-negligible amount of multijet background remains in the WH channel. A data-driven procedure to construct templates for this background is established to overcome the absence of an adequate simulation. The aptitude of the resulting templates is demonstrated in a multi-parameter fit to data.

All relevant systematic uncertainties, be they of experimental or theoretical origin or motivated by the observation of deficiencies in the Monte Carlo simulation, are taken into account through additional nuisance parameters in the global fit model for the signal hypothesis test.

The calculation of an upper limit on the Standard Model Higgs production cross section is performed for the WH channel of the analysis. No significant excess above the Standard Model background expectation is observed. For a Higgs boson mass of $m_H = 125\,\text{GeV}$, the expected limit in the absence of signal is $2.06 \times \sigma_{\text{SM}}$, while the observed limit is $3.15 \times \sigma_{\text{SM}}$. This excess corresponds to a local significance of 1.1 standard deviations. The ratio of the measured signal strength to the Standard Model expectation is $\mu = 1.40 \pm 0.95$.

Extrapolating the statistical uncertainties assigned to the Monte Carlo to the full sample size produced for the latest ATLAS publication [85], the approach presented in this thesis could improve on the expected limit at $m_H = 125\,\text{GeV}$ by 10 – 20%.

“Measure what can be measured, and make measurable what cannot be measured!”

Galileo Galilei

CHAPTER 11

Outlook

The analysis presented in this thesis incorporates many novel ideas that have been developed in the last two years. However, the full potential of the methods introduced is not yet fully exploited. Several possible improvements could not be considered in the given time frame:

1. The fit model presented in the analysis was designed with the combination of WH and ZH channel in mind, but the limit calculation was only performed in the WH channel. Incorporating the ZH channel in the fit would significantly improve the sensitivity of the search.
2. The optimization of the analysis with regard to the event pre-selection and the choice of the discriminating variables is mainly driven by studies in the WH channel. A more careful consideration of the ZH channel could probably improve its sensitivity, in particular taking into account the information about the transverse boost of the ZH system.
3. The modeling problems observed in the 2jet/0tag and 2jet/1tag region should be investigated further. Although there are good arguments to conclude that ISR/FSR contamination plays an important role in these effects, this is probably only a part of the explanation.
4. Studies with training data sets of increasing size suggest that the performance of the BDT classifier used in this analysis could be improved by more than 10% by using a lager Monte Carlo training sample, in particular for the signal process.
5. Increased Monte Carlo statistics would also benefit the limit calculation as additional nuisance parameters representing the statistical uncertainty of the Monte Carlo expectation could be dropped.
6. The fit model for the global fit can probably be simplified by removal of insignificant nuisance parameters, improving stability and speed of computation.
7. Several changes in the analysis are necessary to evaluate the data recorded by ATLAS in 2011. A combination with the 2012 data is desirable.
8. As a long-term goal, a consistent treatment of the $ZH \to \nu\nu b\bar{b}$ channel needs to be developed to optimally exploit $H \to b\bar{b}$ in associated production. Judging based on the cut-based ATLAS analysis, this channel is expected to yield a similar sensitivity than the WH channel.

APPENDIX A

Post-fit event counts in the WH channel

Post-fit event gain WH				
process	3jet/0tag	2jet/1tag$_{slj}$	**2jet/2tag**	3jet/2tag
multijet ($e + \mu$)	123922.0	18346.2	2101.5	2472.2
Z+light	41796.2	1233.8	10.9	0.0
$Z + c$	19022.4	1292.6	44.7	82.7
$Z + b$	3238.9	2272.4	326.3	525.9
W+light	444084.0	8027.5	175.9	77.0
$W + c$	129282.0	23036.9	783.1	798.7
$W + b$	37930.3	23784.5	3833.2	4686.2
single-top	9456.3	15313.5	3757.0	6896.3
diboson	9571.5	1158.9	261.7	188.8
$t\bar{t}$	19091.0	9185.7	6290.7	30653.8
$\sum$ bkg.	837394.0	103652.0	17585.0	46381.8
WH	17.5	47.5	108.6	60.6
data	837427.0	103706.0	17677.0	46447.0

Table A.1: Expected and observed number of events in the WH channel after the global fit

APPENDIX B

Variable distributions in the ZH channel

Just like in the *WH* channel (see Section 7.8), the distributions of the discriminating variables are studied in various control regions in the *ZH* channel. On the whole, a good data/MC agreement is observed. Using the scale factors determined in the normalization fit described in Section 7.9, the plots in Figure B.1- B.6 are obtained.

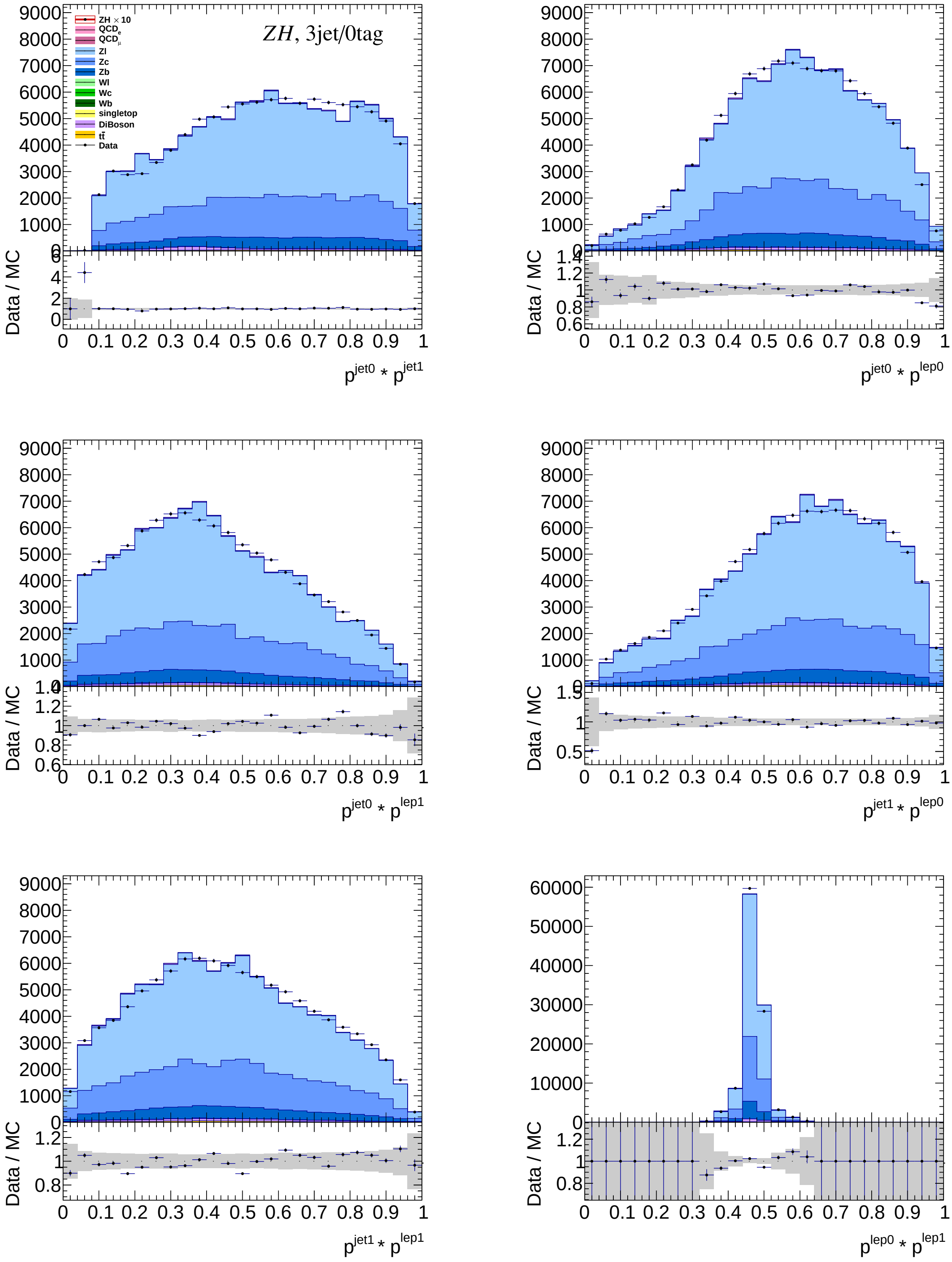


Figure B.1: The Lorentz invariant variables given by the four-vector products of the final state objects in the 3jet/0tag control region of the *ZH* channel.

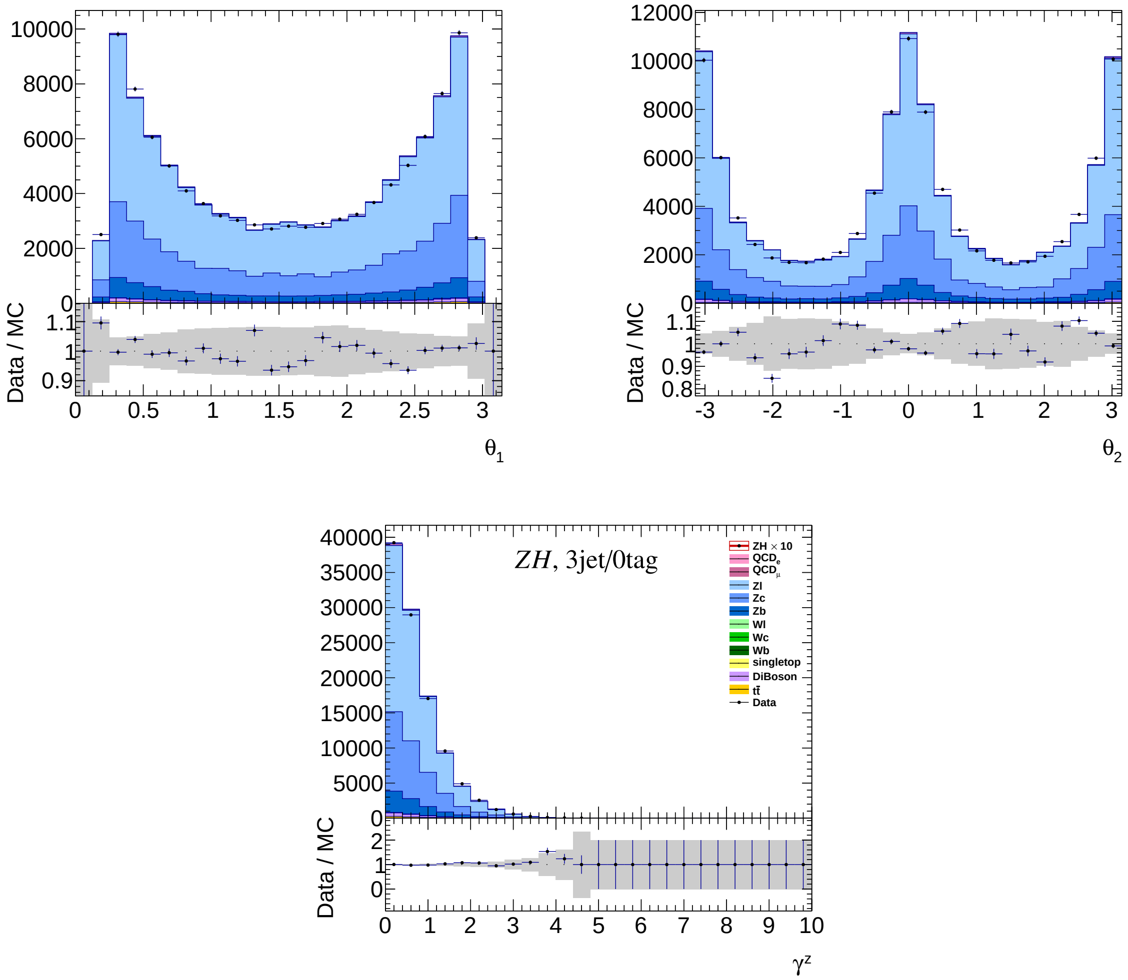


Figure B.2: The angular variables and the longitudinal boost in the 3jet/0tag control region of the *ZH* channel.

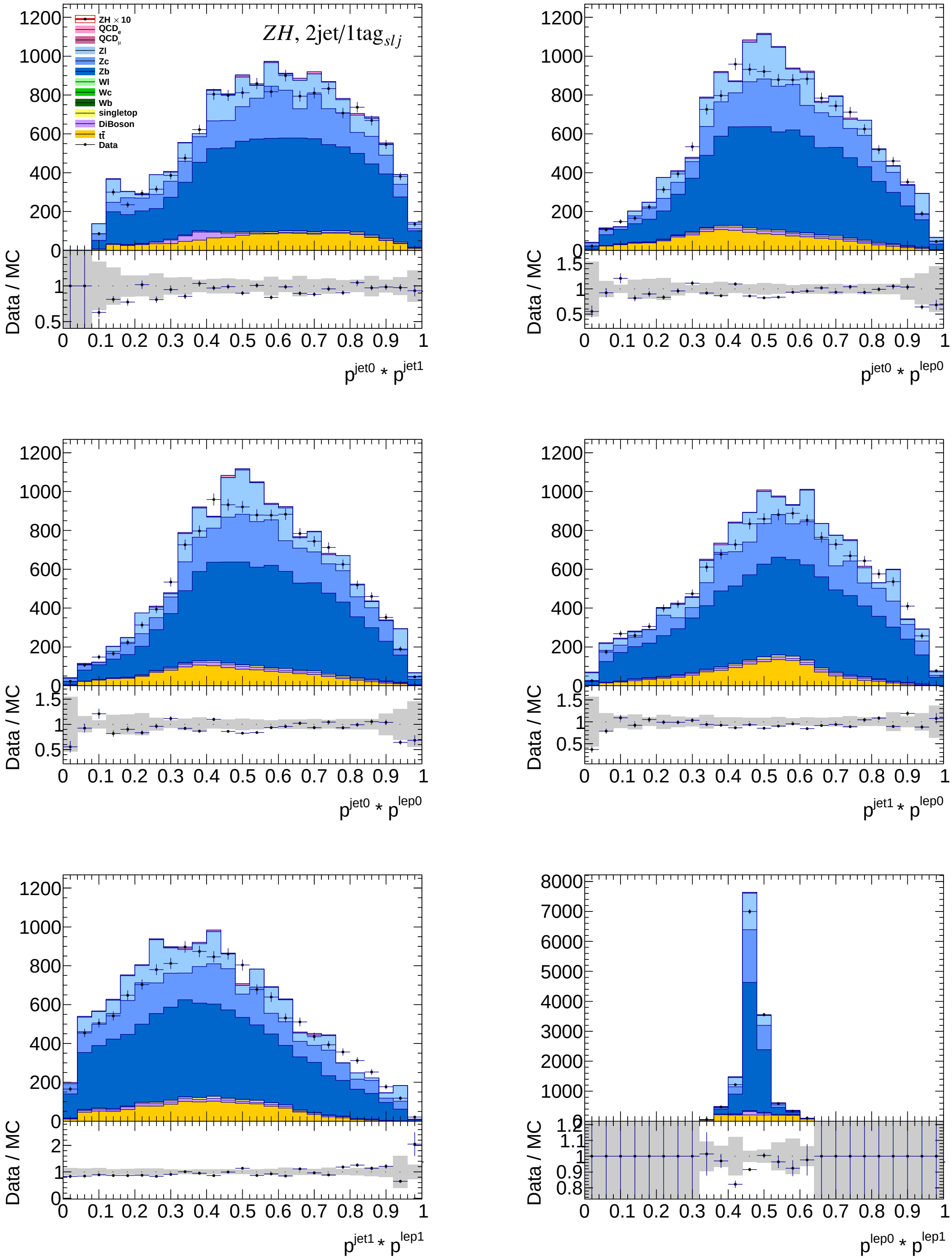


Figure B.3: The Lorentz invariant variables given by the four-vector products of the final state objects in the 2jet/1tag$_{slj}$ control region of the *ZH* channel.

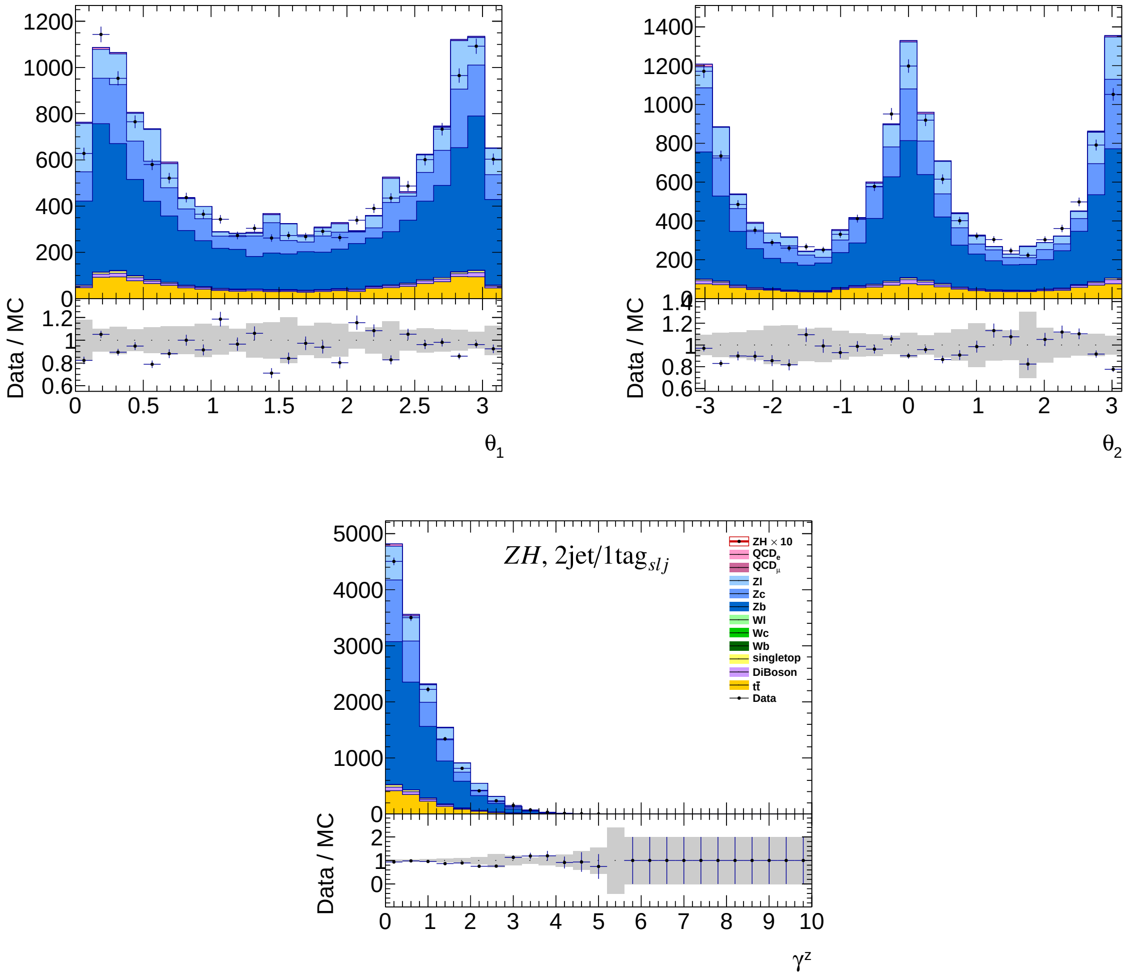


Figure B.4: The angular variables and the longitudinal boost in the 2jet/1tag$_{slj}$ control region of the *ZH* channel.

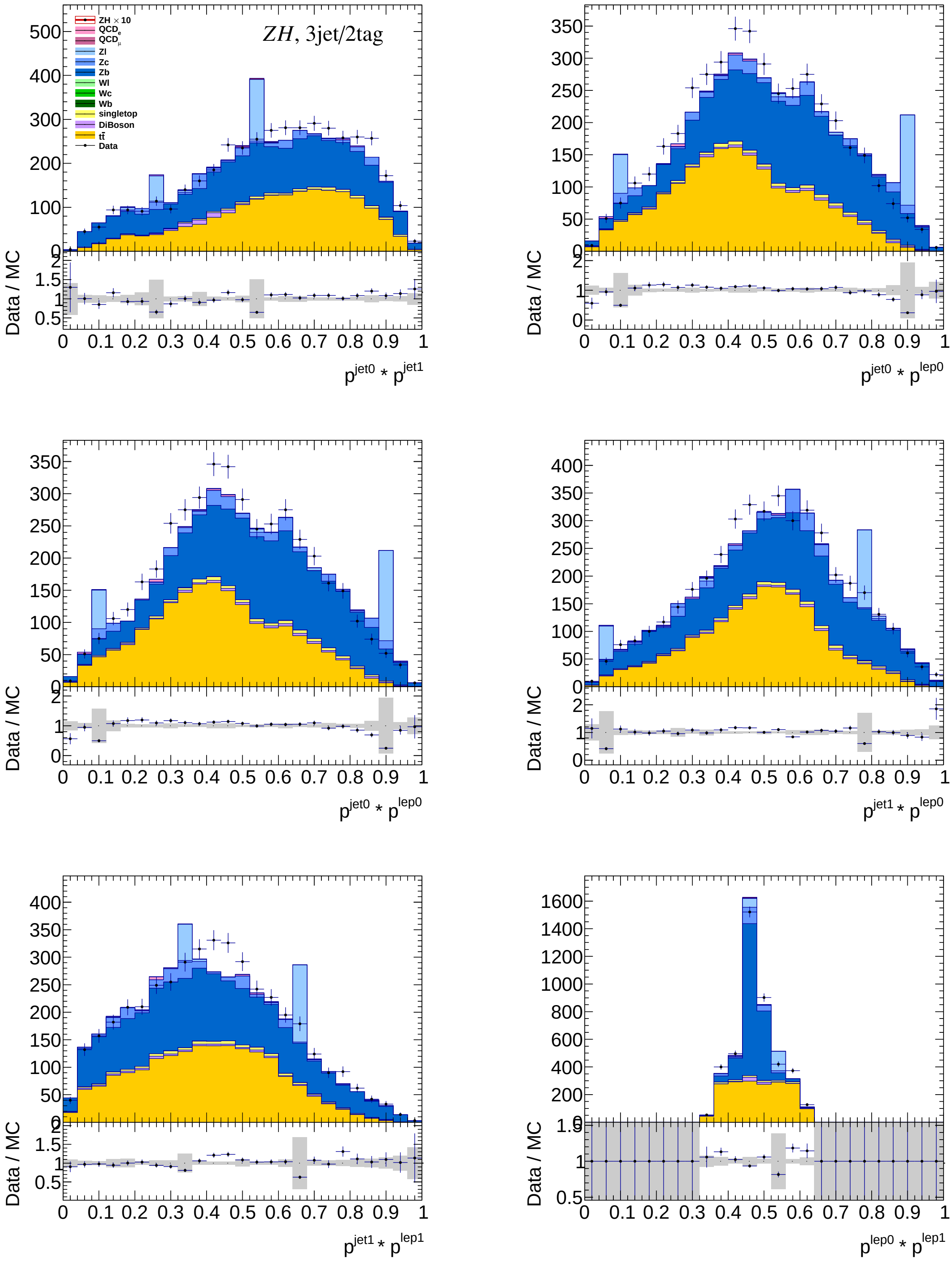


Figure B.5: The Lorentz invariant variables given by the four-vector products of the final state objects in the 3jet/2tag control region of the *ZH* channel.

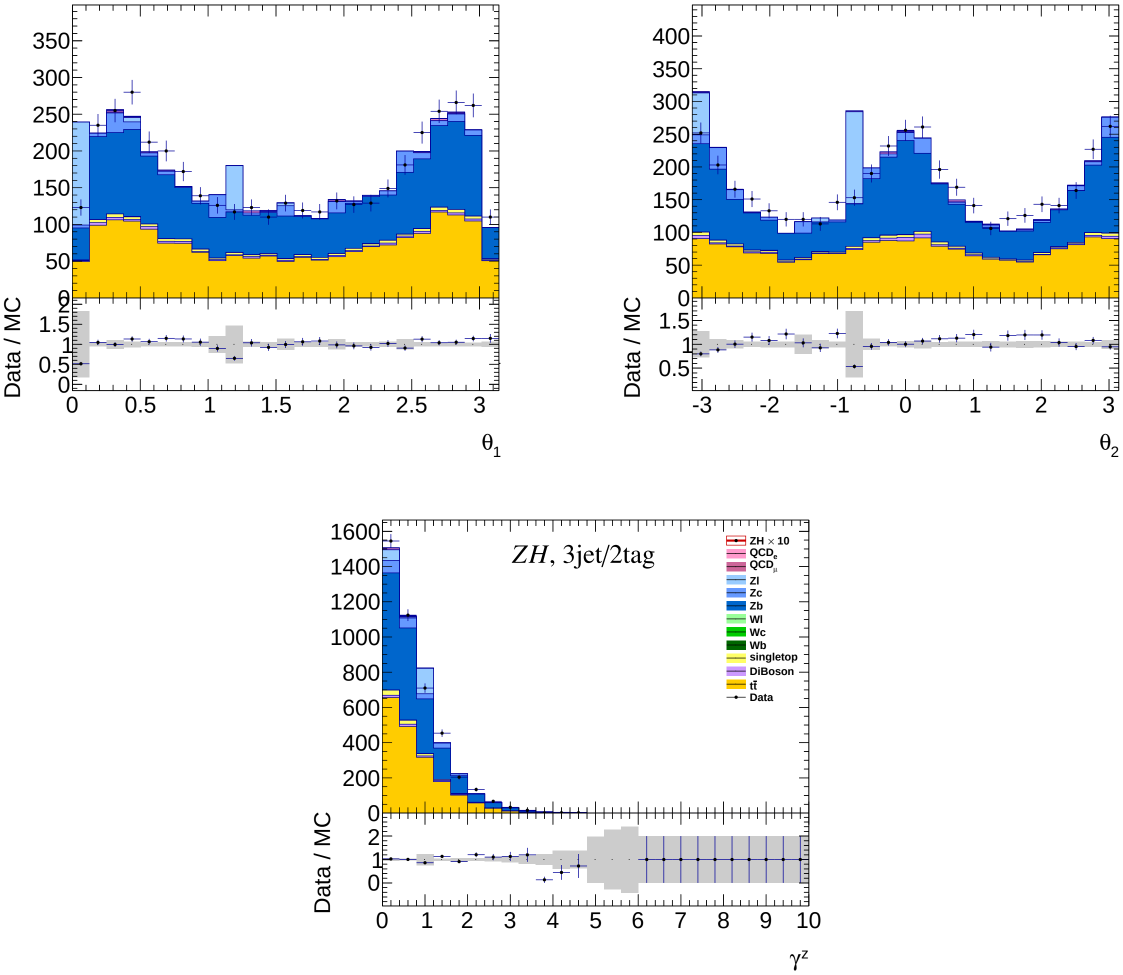


Figure B.6: The angular variables and the longitudinal boost in the 3jet/2tag control region of the *ZH* channel.

APPENDIX C

Overtraining control plots

As discussed in Chapter 6, flexible classification algorithms like boosted decision trees may exhibit overtraining if their parameters are incorrectly tuned or if the amount of training events is insufficient. Although this does *not* result in a bias if the classifier is evaluated on an independent sample, the performance of the classifier will generally degrade in the presence of overtraining. To detect possible overtraining in the analysis presented here, the BDT output distribution obtained on the training sample is overlayed with the BDT output obtained on an independent test sample. The BDT configuration presented in Section 7.6 is used. Figure C.1 shows the resulting distributions for both BDT training algorithms that are employed in the analysis. No evidence for overtraining is observed.

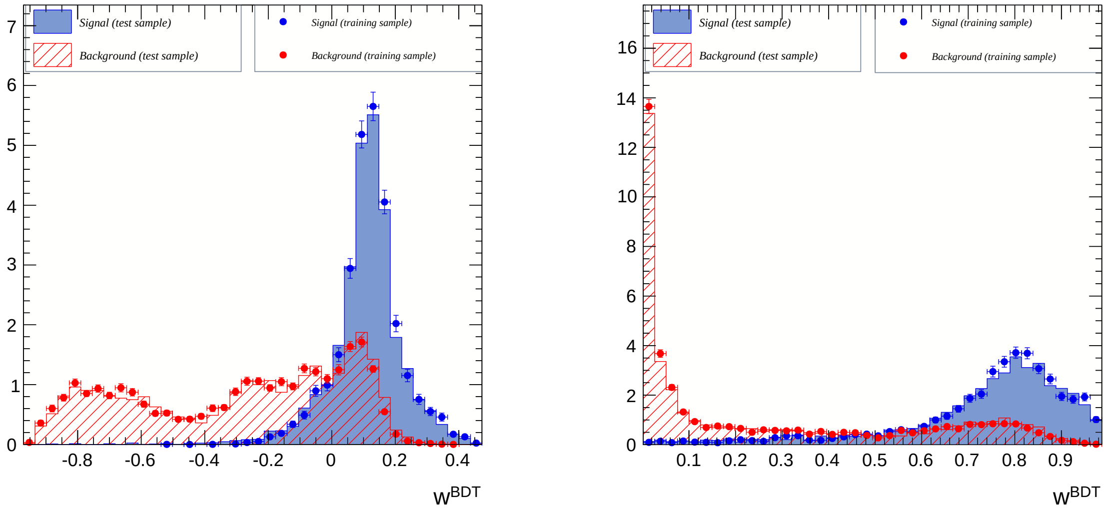


Figure C.1: Overlay of the BDT output obtained for the training sample and an independent test sample in the WH channel. Top: Adaptive boosting. Bottom: Gradient boosting.

APPENDIX D

Effect of systematic uncertainties on the BDT output in the WH signal region

The plots on the following pages show the effect of a ±1σ variation for all nuisance parameters in the global fit. Note that all Monte Carlo scale factors are fixed at 1.0 here, which means that these plots do not reflect the actual data/MC agreement achieved.

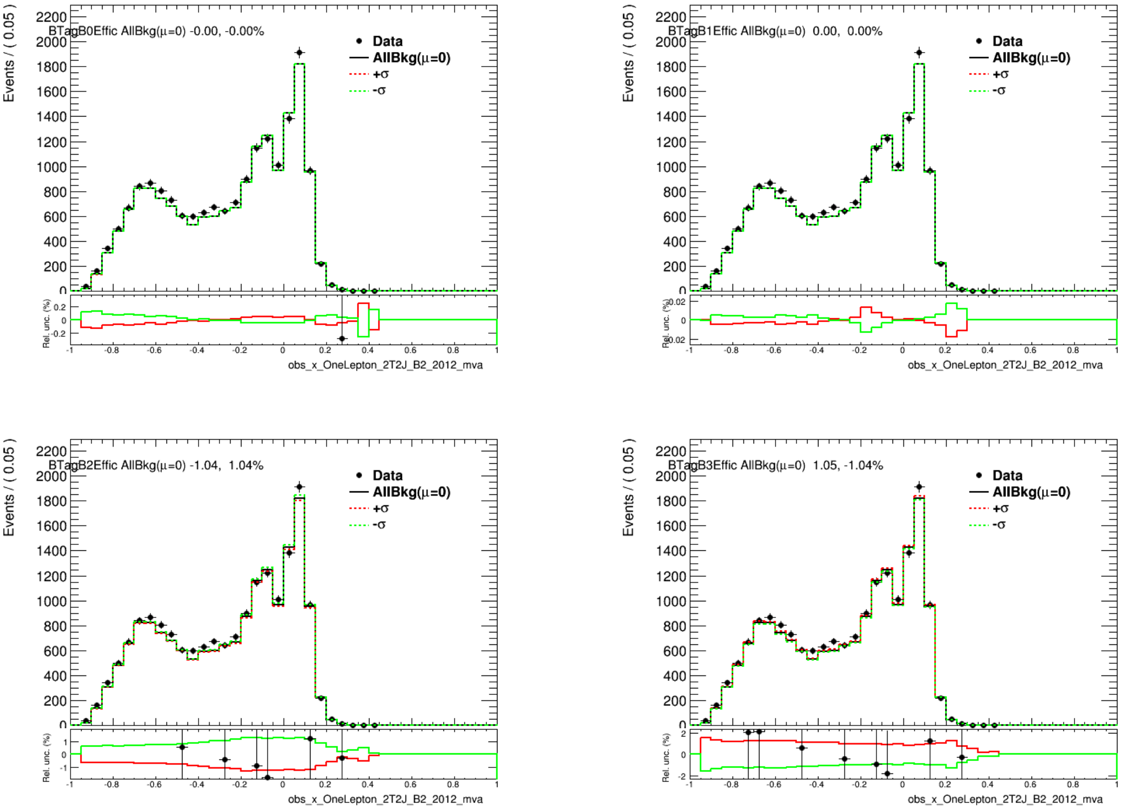


Figure D.1: First set of systematic variations in the signal region of the WH channel.

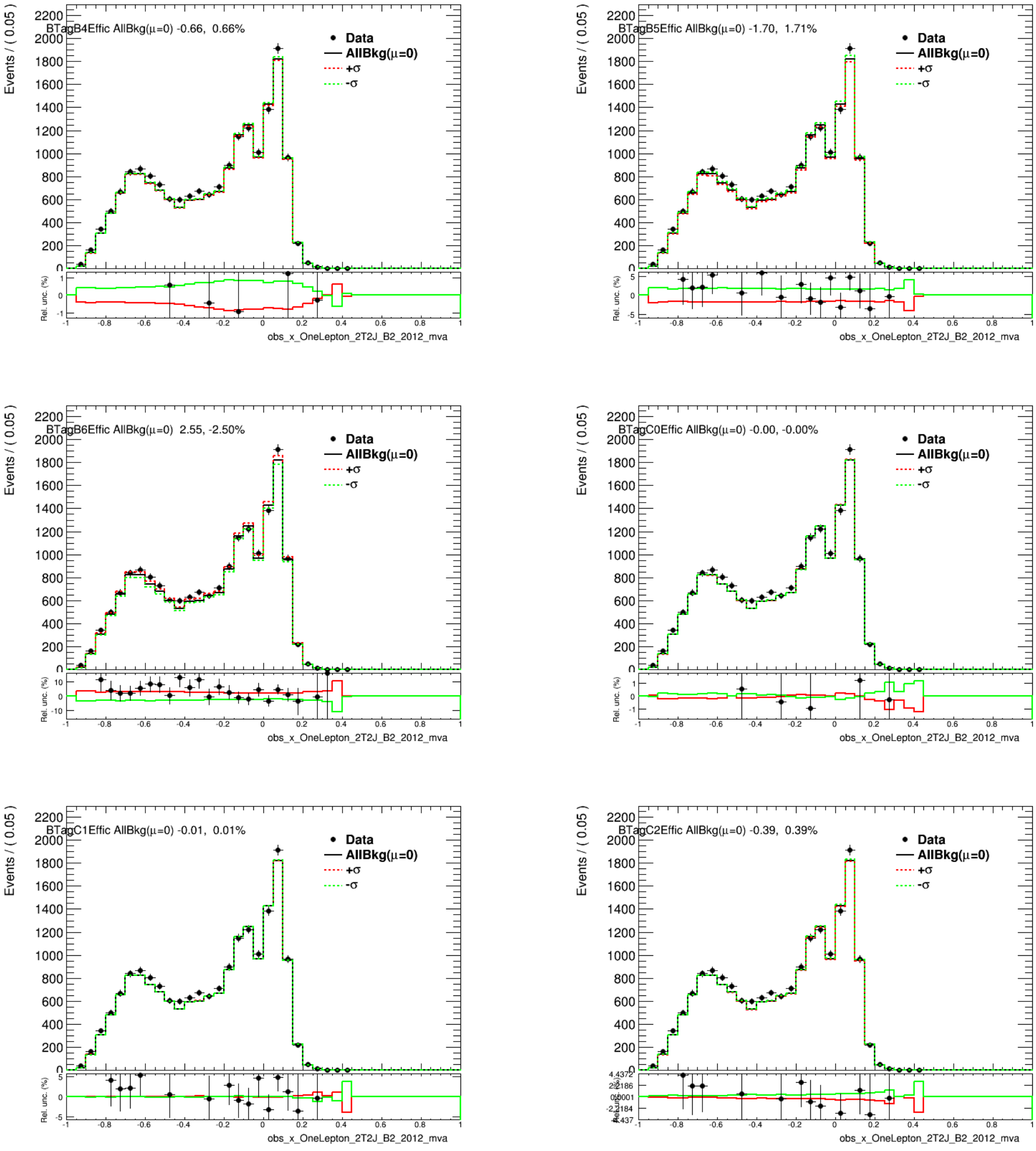


Figure D.2: Second set of systematic variations in the signal region of the *WH* channel.

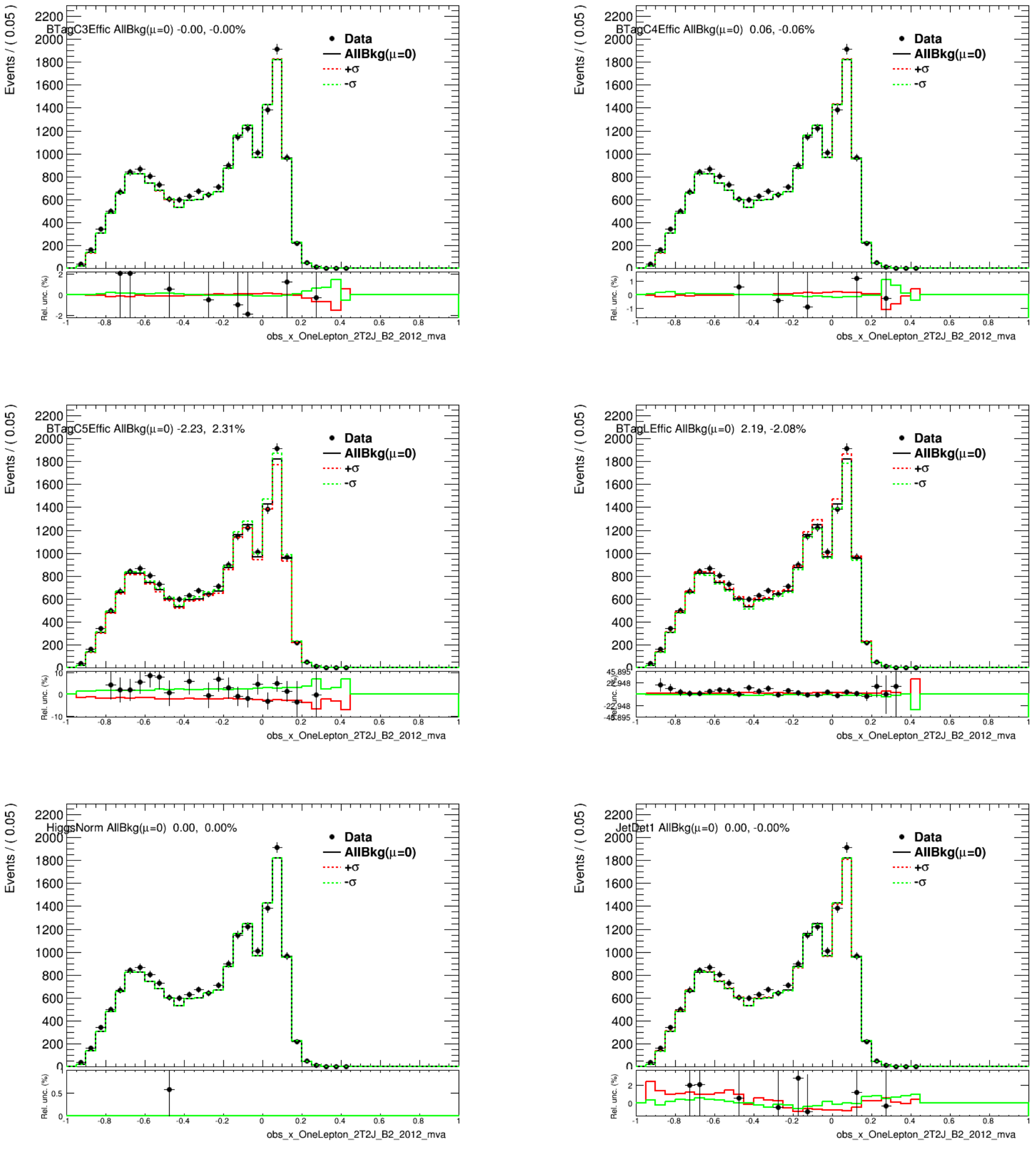


Figure D.3: Third set of systematic variations in the signal region of the WH channel.

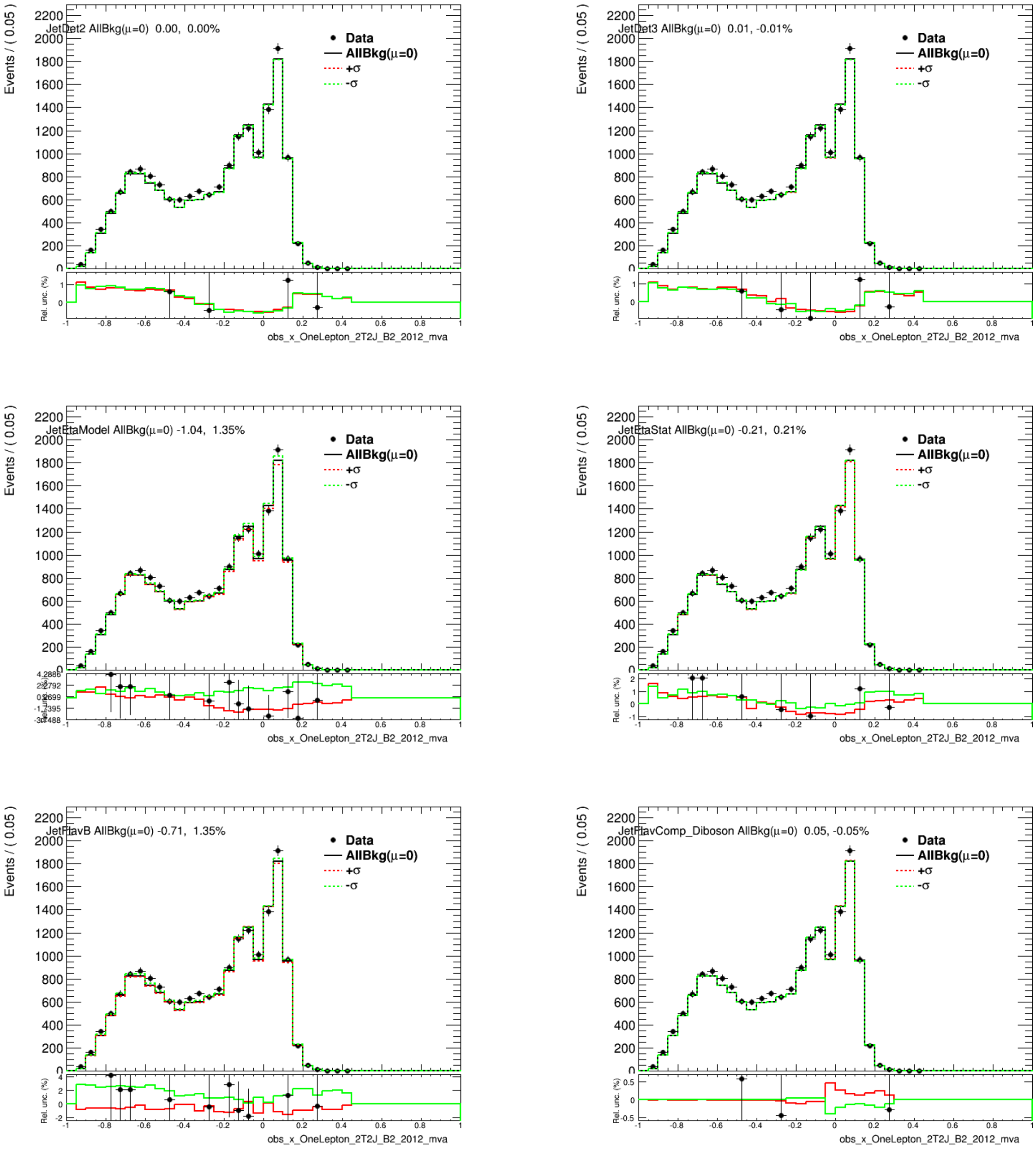


Figure D.4: Fourth set of systematic variations in the signal region of the *WH* channel.

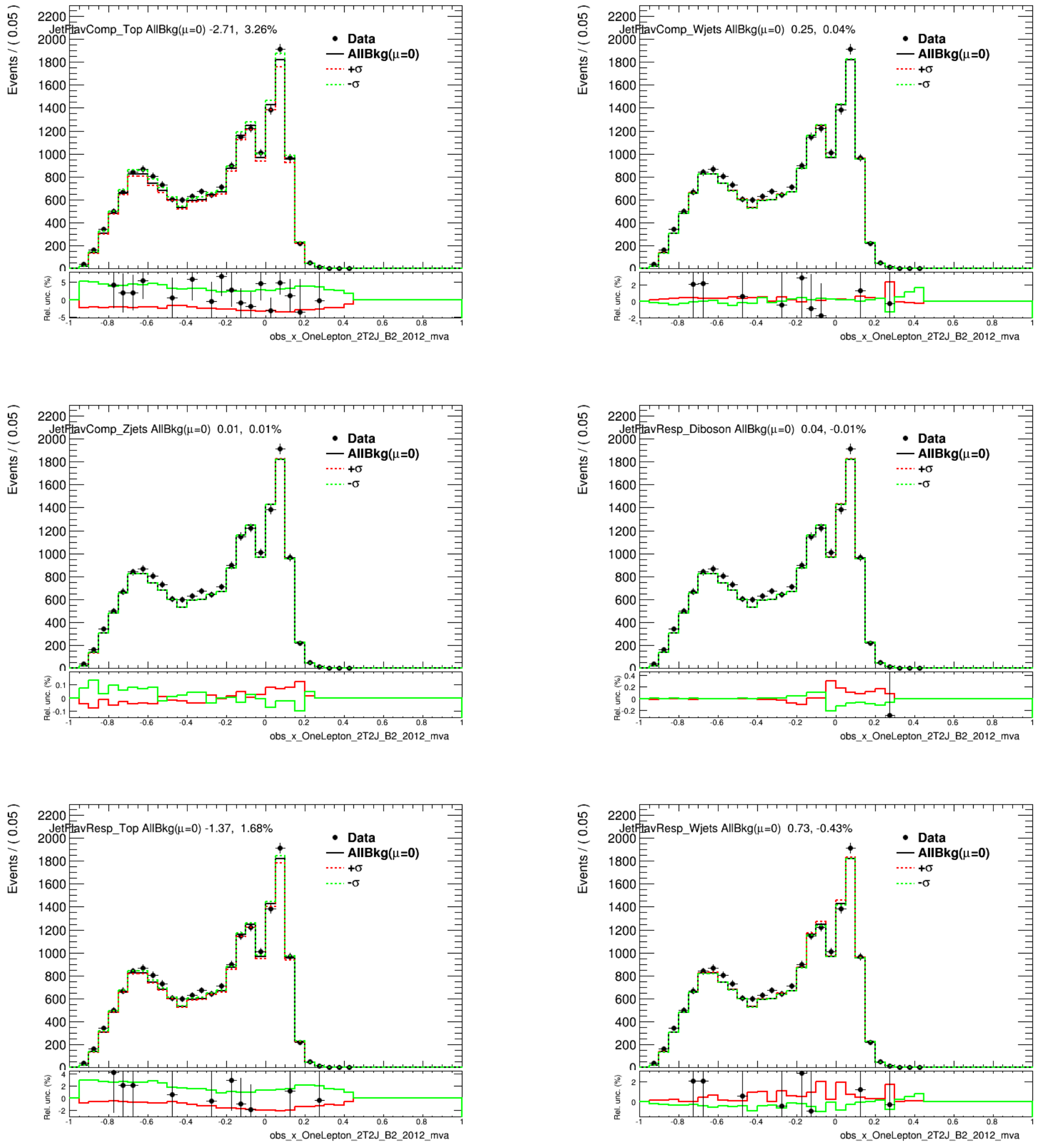


Figure D.5: Fifth set of systematic variations in the signal region of the WH channel.

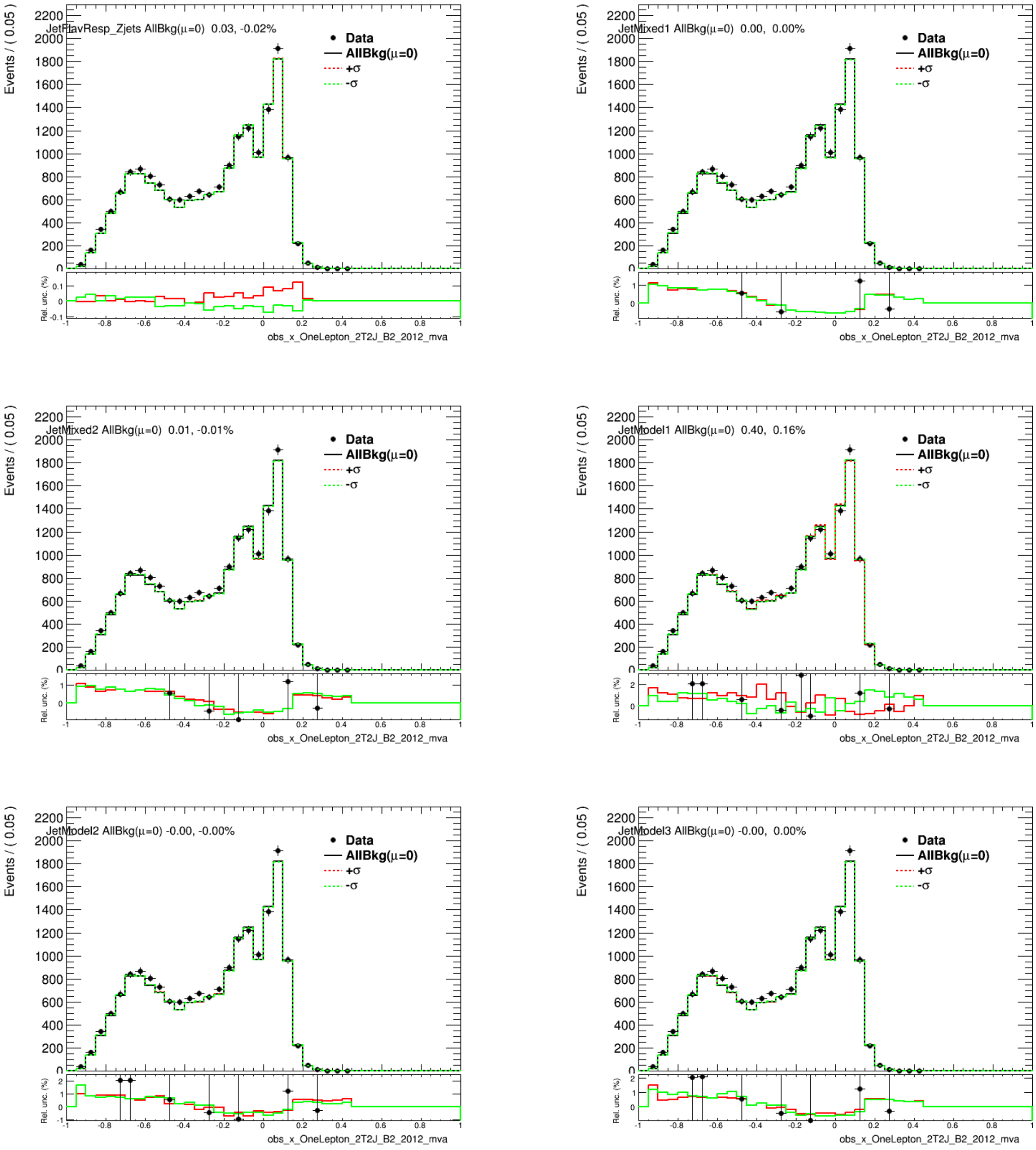


Figure D.6: Sixth set of systematic variations in the signal region of the *WH* channel.

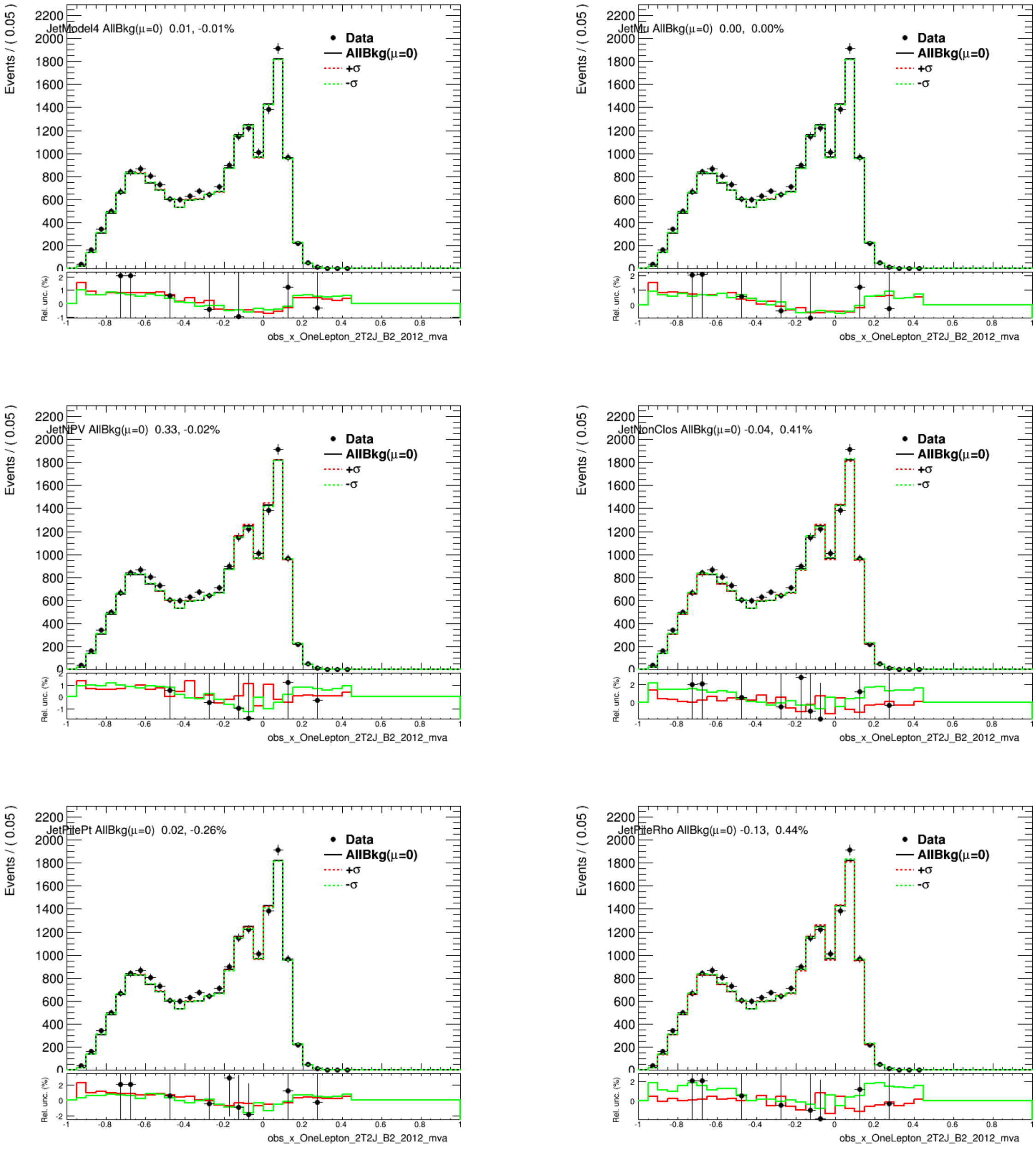


Figure D.7: Seventh set of systematic variations in the signal region of the WH channel.

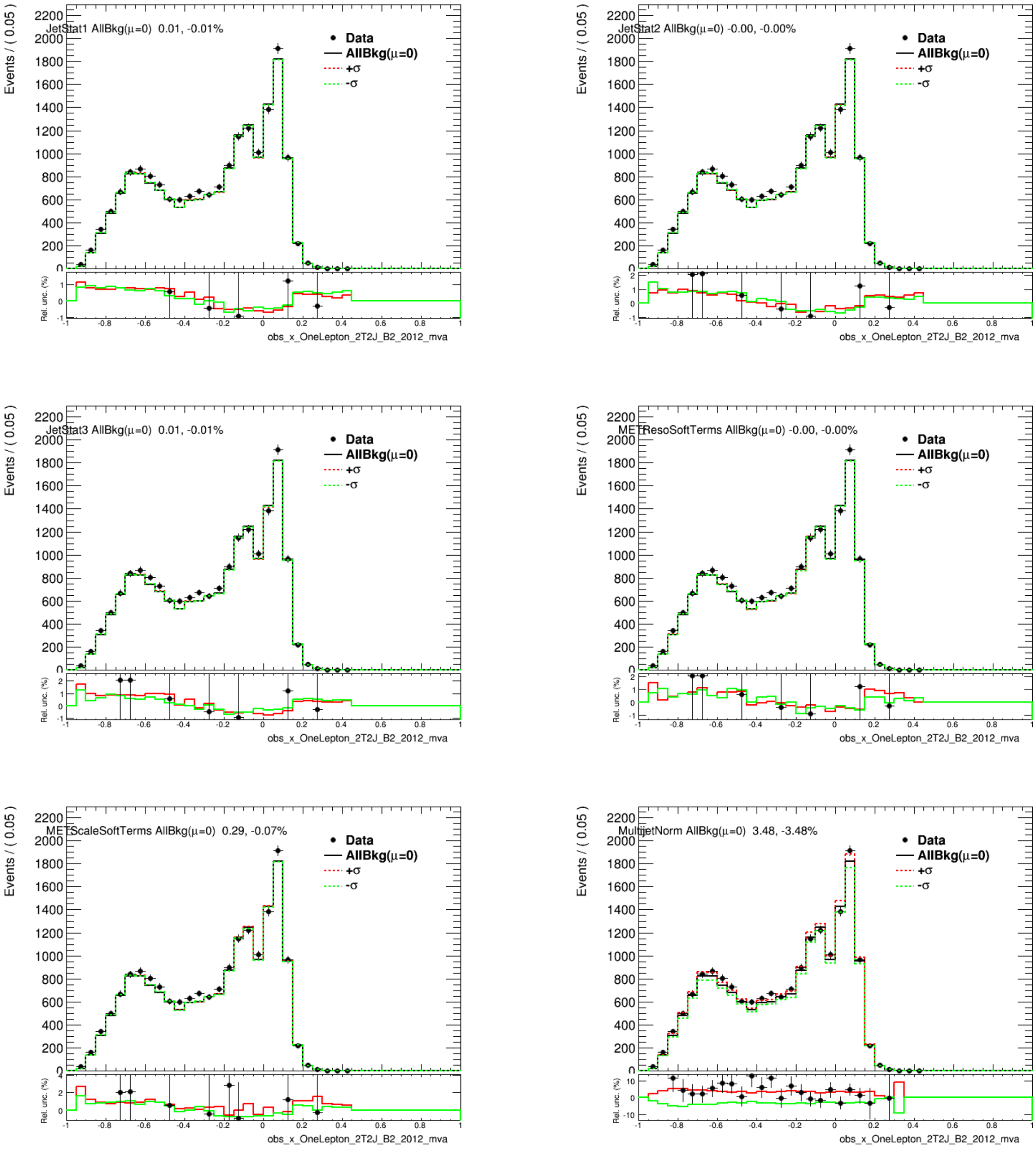


Figure D.8: 8th set of systematic variations in the signal region of the WH channel.

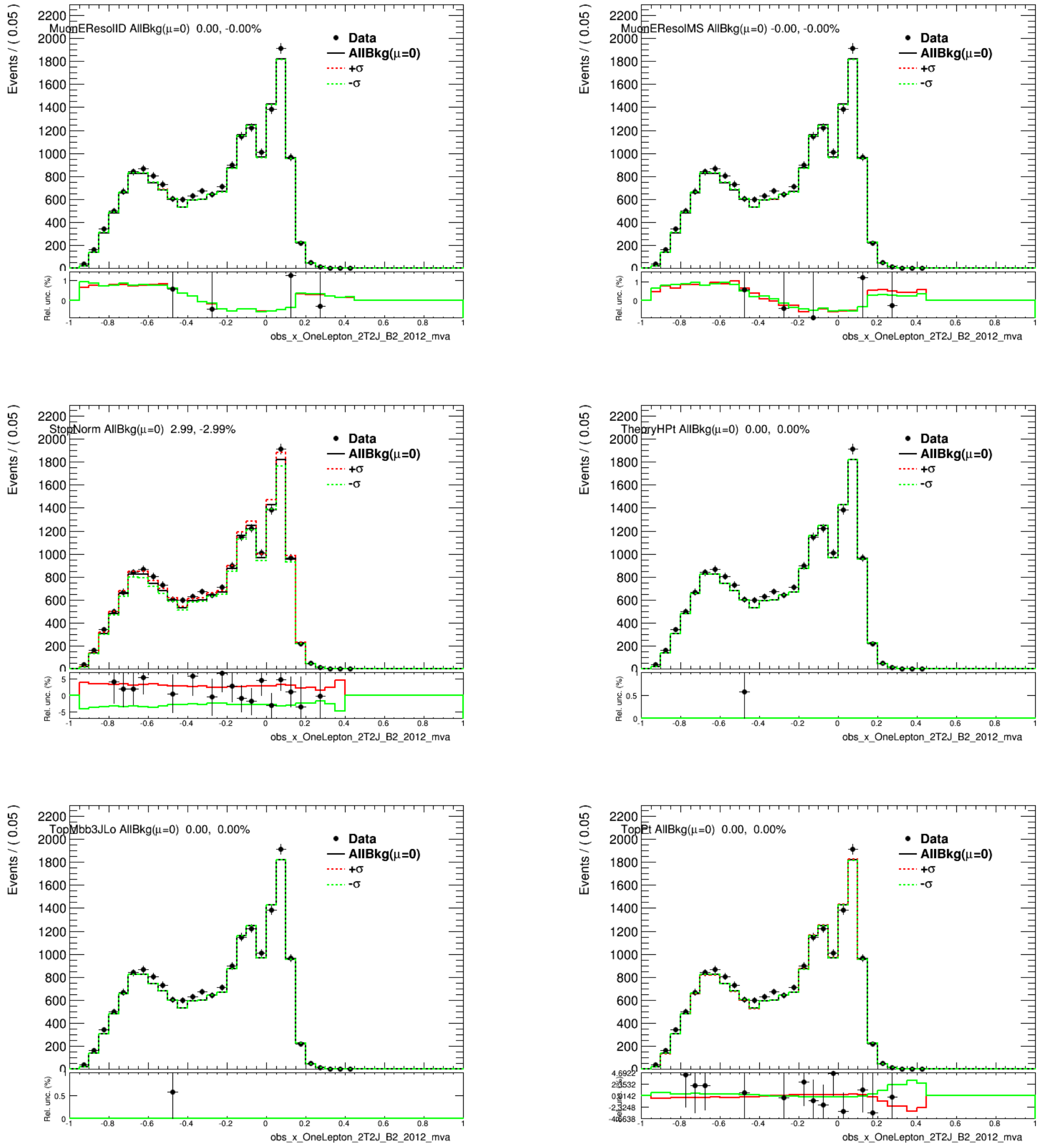


Figure D.9: 9th set of systematic variations in the signal region of the WH channel.

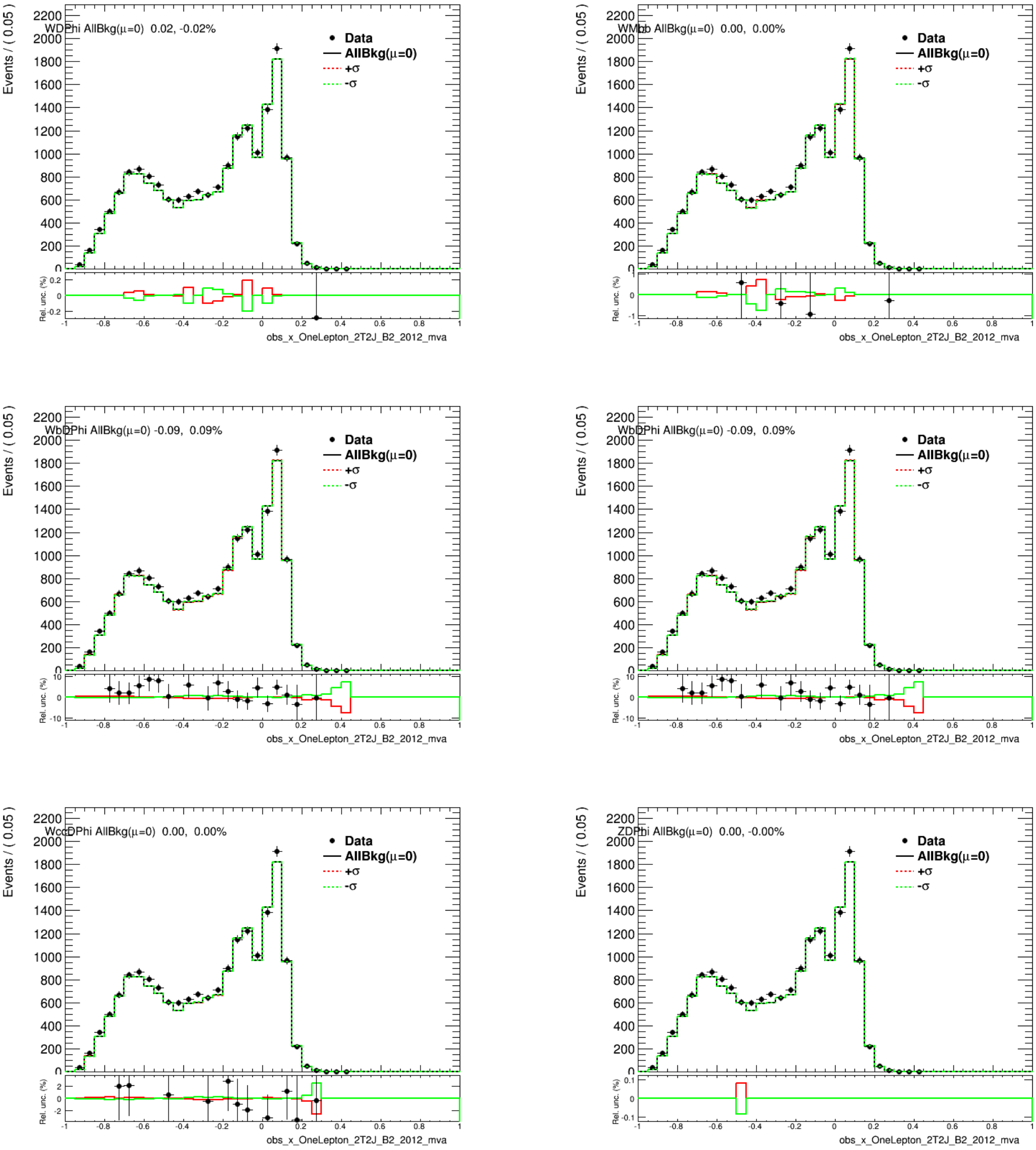


Figure D.10: 10th set of systematic variations in the signal region of the *WH* channel.

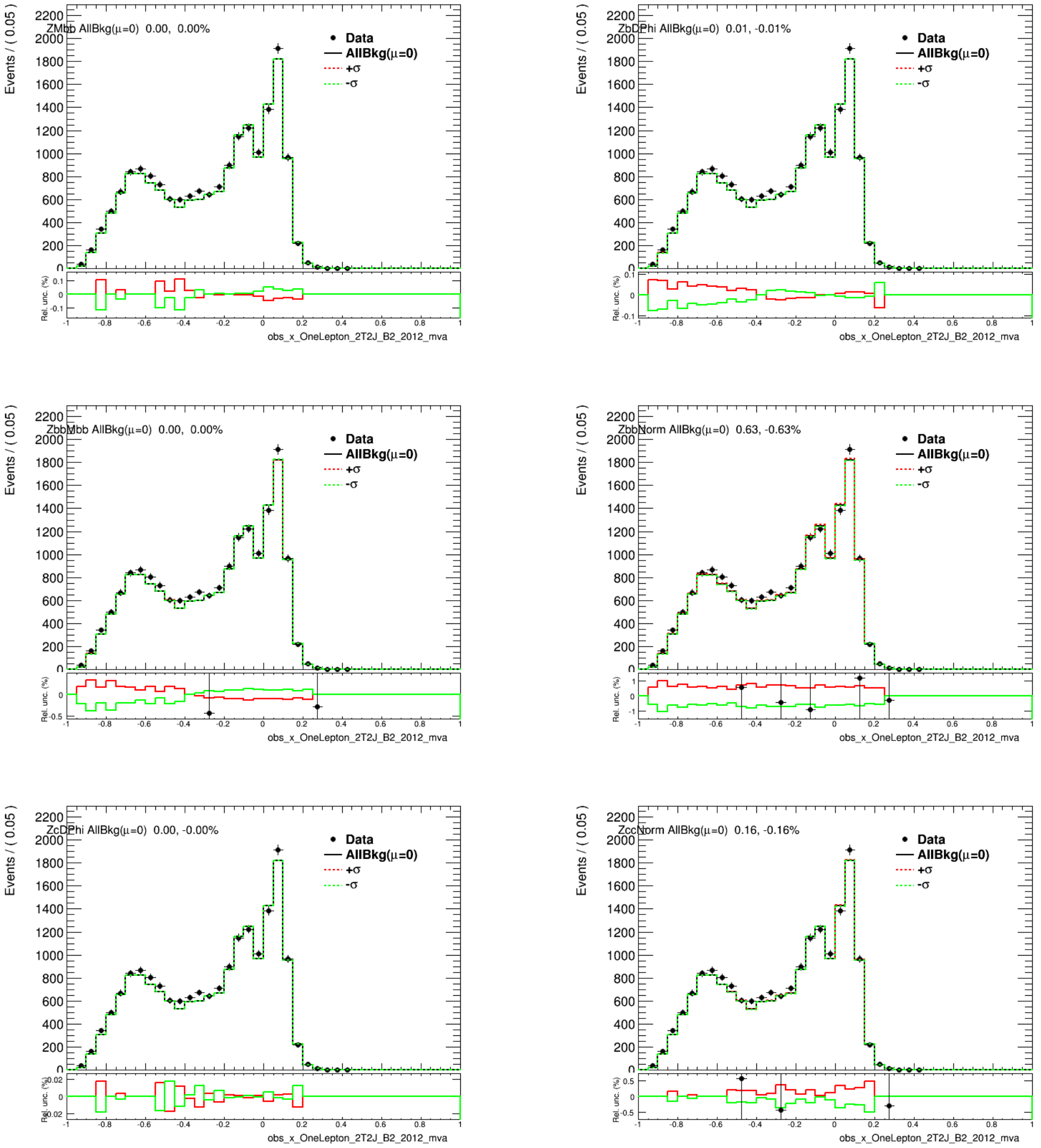


Figure D.11: 11th set of systematic variations in the signal region of the WH channel.

Bibliography

[1] S. Novaes, “Standard model: An Introduction” (1999), eprint: `hep-ph/0001283`.

[2] A. Djouadi,
“The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model”,
Phys.Rept. 457 (2008) 1–216, DOI: `10.1016/j.physrep.2007.10.004`,
eprint: `hep-ph/0503172`.

[3] J. Goldstone, A. Salam and S. Weinberg, “Broken Symmetries”, *Phys. Rev.* 127 (3 1962) 965–970, DOI: `10.1103/PhysRev.127.965`,
URL: `http://link.aps.org/doi/10.1103/PhysRev.127.965`.

[4] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys.Rev.Lett.* 13 (1964) 508–509, DOI: `10.1103/PhysRevLett.13.508`.

[5] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys.Lett.* 12 (1964) 132–133, DOI: `10.1016/0031-9163(64)91136-9`.

[6] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys.Rev.Lett.* 13 (1964) 321–323, DOI: `10.1103/PhysRevLett.13.321`.

[7] G. Guralnik, C. Hagen and T. Kibble, “Global Conservation Laws and Massless Particles”, *Phys.Rev.Lett.* 13 (1964) 585–587, DOI: `10.1103/PhysRevLett.13.585`.

[8] “Mass appeal”, *Nature* 483 (2012) 374, DOI: `10.1038/483374a`.

[9] F. Gianotti on behalf of the ATLAS collaboration,
“Status of Standard Model Higgs Searches in ATLAS”,
Presentation given at CERN on July 4, 2012.

[10] J. Incandela on behalf of the CMS collaboration, “Status of the CMS SM Higgs Search”, Presentation given at CERN on July 4, 2012.

[11] G. Aad et al., “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys.Lett.* B716 (2012) 1–29,
DOI: `10.1016/j.physletb.2012.08.020`, eprint: `1207.7214`.

[12] S. Chatrchyan et al.,
“Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”,
Phys.Lett. B716 (2012) 30–61, DOI: `10.1016/j.physletb.2012.08.021`,
eprint: `1207.7235`.

[13] B. W. Lee, C. Quigg and H. Thacker,
“Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass”,
Phys.Rev. D16 (1977) 1519, DOI: `10.1103/PhysRevD.16.1519`.

[14] W. J. Marciano and S. S. Willenbrock, "Radiative corrections to heavy Higgs scalar production and decay", *Phys.Rev.* D37 (1988) 2509, DOI: `10.1103/PhysRevD.37.2509`.

[15] K. Riesselmann and S. Willenbrock, "Ruling out a strongly interacting standard Higgs model", *Phys.Rev.* D55 (1997) 311–321, DOI: `10.1103/PhysRevD.55.311`, eprint: `hep-ph/9608280`.

[16] M. J. Duncan, R. Philippe and M. Sher, "Theoretical ceiling on quark masses in the standard model", *Phys.Lett.* B153 (1985) 165, DOI: `10.1016/0370-2693(85)91421-2`.

[17] M Gockeler et al., "Scaling analysis of the O(4)-symmetrical phi-4-theory in the broken phase", English, *NUCLEAR PHYSICS B* 404.1-2 (1993) 517–555, ISSN: 0550-3213.

[18] N. Cabibbo et al., "Bounds on the Fermions and Higgs Boson Masses in Grand Unified Theories", *Nucl.Phys.* B158 (1979) 295–305, DOI: `10.1016/0550-3213(79)90167-6`.

[19] M. Lindner, M. Sher and H. W. Zaglauer, "Probing Vacuum Stability Bounds at the Fermilab Collider", *Phys.Lett.* B228 (1989) 139, DOI: `10.1016/0370-2693(89)90540-6`.

[20] T. van Ritbergen and R. G. Stuart, "On the precise determination of the Fermi coupling constant from the muon lifetime", *Nucl.Phys.* B564 (2000) 343–390, DOI: `10.1016/S0550-3213(99)00572-6`, eprint: `hep-ph/9904240`.

[21] S. Heinemeyer et al., "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties" (2013), eprint: `1307.1347`.

[22] J. Gillies, *LHC 2012 proton run extended by seven weeks*, CERN bulletin, July 23, 2012.

[23] "Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV", tech. rep. CMS-PAS-HIG-12-045, CERN, 2012.

[24] "An update of combined measurements of the new Higgs-like boson with high mass resolution channels", tech. rep. ATLAS-CONF-2012-170, CERN, 2012.

[25] G. Aad et al., "Evidence for the spin-0 nature of the Higgs boson using ATLAS data" (2013), eprint: `1307.1432`.

[26] S. Chatrchyan et al., "Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs", *Phys. Rev. Lett.* 110 (8 2013) 081803, DOI: `10.1103/PhysRevLett.110.081803`, URL: `http://link.aps.org/doi/10.1103/PhysRevLett.110.081803`.

[27] M. Mangano et al., "b anti-b final states in Higgs production via weak boson fusion at the LHC", *Phys.Lett.* B556 (2003) 50–60, DOI: `10.1016/S0370-2693(03)00078-9`, eprint: `hep-ph/0210261`.

[28] O. Brein, A. Djouadi and R. Harlander, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders", *Phys.Lett.* B579 (2004) 149–156, DOI: `10.1016/j.physletb.2003.10.112`, eprint: `hep-ph/0307206`.

[29] S. Dittmaier et al., "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables" (2011), eprint: `1101.0593`.

[30] M. Ciccolini, S. Dittmaier and M. Kramer, "Electroweak radiative corrections to associated WH and ZH production at hadron colliders", *Phys.Rev.* D68 (2003) 073003, DOI: `10.1103/PhysRevD.68.073003`, eprint: `hep-ph/0306234`.

[31] R. Kleiss, Z. Kunszt and W. J. Stirling, "Discovering a light intermediate mass Higgs in proton proton collisions using associated production with W bosons", *Phys.Lett.* B253 (1991) 269–274, DOI: `10.1016/0370-2693(91)91396-D`.

[32] B. A. Kniehl, "Associated production of Higgs and Z bosons from gluon fusion in hadron collisions", *Phys.Rev.* D42 (1990) 2253–2258, DOI: `10.1103/PhysRevD.42.2253`.

[33] R. Lafaye et al., "Measuring the Higgs Sector", *JHEP* 0908 (2009) 009, DOI: `10.1088/1126-6708/2009/08/009`, eprint: `0904.3866`.

[34] A. Team, "The four main LHC experiments", 1999.

[35] A. Team, "Diagram of an LHC dipole magnet.", 1999.

[36] G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* 3 (2008) S08003, DOI: `10.1088/1748-0221/3/08/S08003`.

[37] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008.

[38] J. Pequenao, "Computer generated image of the ATLAS inner detector", 2008.

[39] G. Aad et al., "Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC", *New J.Phys.* 13 (2011) 053033, DOI: `10.1088/1367-2630/13/5/053033`, eprint: `1012.5104`.

[40] W. Lampl et al., "Calorimeter clustering algorithms: Description and performance" (2008).

[41] M. Cacciari, G. P. Salam and G. Soyez, "The Anti-k(t) jet clustering algorithm", *JHEP* 0804 (2008) 063, DOI: `10.1088/1126-6708/2008/04/063`, eprint: `0802.1189`.

[42] "A measurement of the muon reconstruction efficiency in 2010 ATLAS data using jpsi decays", tech. rep. ATLAS-CONF-2012-125, CERN, 2012.

[43] G. Aad et al., "Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics" (2009), eprint: `0901.0512`.

[44] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys.Rept.* 504 (2011) 145–233, DOI: `10.1016/j.physrep.2011.03.005`, eprint: `1101.2599`.

[45] P. Skands, "Introduction to QCD" (2012), eprint: `1207.2389`.

[46] F. James, "Monte Carlo theory and practice", *Rept.Prog.Phys.* 43 (1980) 1145, DOI: `10.1088/0034-4885/43/9/002`.

[47] J. C. Collins and D. E. Soper, "Parton Distribution and Decay Functions", *Nucl.Phys.* B194 (1982) 445, DOI: `10.1016/0550-3213(82)90021-9`.

[48] R. Brock et al., "Handbook of perturbative QCD; Version 1.1: September 1994", *Rev. Mod. Phys.* (1994).

[49] A. Cooper-Sarkar, “What did HERA teach us about the structure of the proton?”, *J.Phys.* G39 (2012) 093001, DOI: `10.1088/0954-3899/39/9/093001`, eprint: `1206.0894`.

[50] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl.Phys.* B126 (1977) 298, DOI: `10.1016/0550-3213(77)90384-4`.

[51] T. Plehn, “LHC Phenomenology for Physics Hunters” (2008) 125–180, eprint: `0810.2281`.

[52] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* 1106 (2011) 128, DOI: `10.1007/JHEP06(2011)128`, eprint: `1106.0522`.

[53] T. Kinoshita, “Mass Singularities of Feynman Amplitudes”, *Journal of Mathematical Physics* 3.4 (1962) 650–677, DOI: `10.1063/1.1724268`, URL: `http://link.aip.org/link/?JMP/3/650/1`.

[54] T. D. Lee and M. Nauenberg, “Degenerate Systems and Mass Singularities”, *Phys. Rev.* 133 (6B 1964) B1549–B1562, DOI: `10.1103/PhysRev.133.B1549`, URL: `http://link.aps.org/doi/10.1103/PhysRev.133.B1549`.

[55] W. Giele, D. Kosower and P. Skands, “Higher-Order Corrections to Timelike Jets”, *Phys.Rev.* D84 (2011) 054003, DOI: `10.1103/PhysRevD.84.054003`, eprint: `1102.2126`.

[56] G. Corcella et al., “HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)”, *JHEP* 0101 (2001) 010, DOI: `10.1088/1126-6708/2001/01/010`, eprint: `hep-ph/0011363`.

[57] S. Catani et al., “QCD matrix elements + parton showers”, *JHEP* 0111 (2001) 063, eprint: `hep-ph/0109231`.

[58] S. Frixione and B. R. Webber, “The MC and NLO 3.4 Event Generator” (2008), eprint: `0812.0770`.

[59] T. Sjostrand, S. Mrenna and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput.Phys.Commun.* 178 (2008) 852–867, DOI: `10.1016/j.cpc.2008.01.036`, eprint: `0710.3820`.

[60] S. Frixione, P. Nason and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* 0711 (2007) 070, DOI: `10.1088/1126-6708/2007/11/070`, eprint: `0709.2092`.

[61] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* 0902 (2009) 007, DOI: `10.1088/1126-6708/2009/02/007`, eprint: `0811.4622`.

[62] X. Artru and G. Mennessier, “String model and multiproduction”, *Nucl.Phys.* B70 (1974) 93–115, DOI: `10.1016/0550-3213(74)90360-5`.

[63] B. Andersson, S. Mohanty and F. Soderberg, “Recent developments in the Lund model” (2002), eprint: `hep-ph/0212122`.

[64] G. Aad et al., “Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s} = 900$ GeV and 7 TeV with the ATLAS detector”, *Phys.Rev.* D83 (2011) 112001, DOI: `10.1103/PhysRevD.83.112001`, eprint: `1012.0791`.

[65] R. Field, “Min-Bias and the Underlying Event at the LHC”, *Acta Phys.Polon.* B42 (2011) 2631–2656, DOI: `10.5506/APhysPolB.42.2631`, eprint: `1110.5530`.

[66] T. Sjostrand and M. van Zijl, “A Multiple Interaction Model for the Event Structure in Hadron Collisions”, *Phys.Rev.* D36 (1987) 2019, DOI: `10.1103/PhysRevD.36.2019`.

[67] S. Agostinelli et al., “GEANT4: A Simulation toolkit”, *Nucl.Instrum.Meth.* A506 (2003) 250–303, DOI: `10.1016/S0168-9002(03)01368-8`.

[68] G. Aad et al., “The ATLAS Simulation Infrastructure”, *Eur.Phys.J.* C70 (2010) 823–874, DOI: `10.1140/epjc/s10052-010-1429-9`, eprint: `1005.4568`.

[69] P. Golonka and Z. Was, “PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays”, *Eur.Phys.J.* C45 (2006) 97–107, DOI: `10.1140/epjc/s2005-02396-4`, eprint: `hep-ph/0506026`.

[70] Z. Was, “TAUOLA the library for tau lepton decay, and KKMC / KORALB / KORALZ /... status report”, *Nucl.Phys.Proc.Suppl.* 98 (2001) 96–102, DOI: `10.1016/S0920-5632(01)01200-2`, eprint: `hep-ph/0011305`.

[71] “New ATLAS event generator tunes to 2010 data” (2011).

[72] “Search for the Standard Model Higgs boson in produced in association with a vector boson and decaying to bottom quarks with the ATLAS detector”, tech. rep. ATLAS-CONF-2012-161, CERN, 2012.

[73] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2009.

[74] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses”, English, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933) pp. 289–337, ISSN: 02643952, URL: `http://www.jstor.org/stable/91247`.

[75] C. M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 2006.

[76] D. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003, URL: `http://www.inference.phy.cam.ac.uk/mackay/itila/book.html`.

[77] J. Therhaag, “Introduction to neural networks in high energy physics”, *European Physical Journal Web of Conferences*, vol. 55, European Physical Journal Web of Conferences, July 2013, DOI: `10.1051/epjconf/20135502003`.

[78] B. P. Roe et al., “Boosted decision trees as an alternative to artificial neural networks for particle identification”, *Nuclear Instruments and Methods in Physics Research A* 543 (May 2005) 577–584, DOI: `10.1016/j.nima.2004.12.018`, eprint: `arXiv:physics/0408124`.

[79] L. Breiman et al., *Classification and Regression Trees*, Belmont, CA: Wadsworth, 1984.

[80] C. Gini, “Variabilita e mutabilita”, *Memorie di metodologia statistica* (1912).

[81] Y. Freund and R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, *Journal of Computer and System Sciences* 55.1 (1997) 119 –139, ISSN: 0022-0000.

[82] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine”, *Annals of Statistics* 29 (2000) 1189–1232.

[83] J. H. Friedman, “Stochastic Gradient Boosting”, *Computational Statistics and Data Analysis* 38 (1999) 367–378.

[84] A. Hoecker et al., “TMVA: Toolkit for Multivariate Data Analysis”, *PoS* ACAT (2007) 040, eprint: `physics/0703039`.

[85] “Search for the bb decay of the Standard Model Higgs boson in associated W/ZH production with the ATLAS detector”, tech. rep. ATLAS-CONF-2013-079, CERN, 2013.

[86] G. Aad et al., “Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector at the LHC”, *Eur.Phys.J.* C71 (2011) 1630, DOI: `10.1140/epjc/s10052-011-1630-5`, eprint: `1101.2185`.

[87] S. Hageboeck, private communication, 2012.

[88] H Gray, G Piacquadio and R Goncalo, “Search for the Standard Model Higgs boson in associated production with a vector boson and decaying to bottom quarks with the ATLAS detector”, tech. rep. ATL-COM-PHYS-2012-1081, CERN, 2012.

[89] “Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data”, tech. rep. ATLAS-CONF-2013-088, CERN, 2013.

[90] G. Aad et al., “Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data”, *The European Physical Journal C* 72.3 (2012) 1–46, ISSN: 1434-6044.

[91] “b-jet tagging calibration on c-jets containing D^{*+} mesons” (2012).

[92] G Aad et al., “B-tagging for VH, with $H \to b\bar{b}$”, tech. rep. ATL-COM-PHYS-2013-697, CERN, 2013.

[93] V. Kostyukhin, *b-jet energy correction to parton level and validation in $t\bar{t}$*, Presentation in ATLAS meeting on jet calibration and resolution, 26 February 2013, 2013.

[94] G Aad et al., “Invariant Mass Studies for the $H \to b\bar{b}$ Measurements for LHCP”, tech. rep. ATL-COM-PHYS-2013-449, CERN, 2013.

[95] W. Verkerke and D. Kirkby, “The RooFit toolkit for data modeling” (2003), eprint: `arXiv:physics/0306116`.

[96] V. Kostyukhin, *Electron QCD shape problem*, Presentation in the Bonn $H \to b\bar{b}$ meeting, 23 May 2013, 2013.

[97] “Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks for LHCp 2013”, tech. rep. CMS-PAS-HIG-13-012, CERN, 2013.

[98] J. Gallicchio et al., “Multivariate discrimination and the Higgs + W/Z search”, *JHEP* 1104 (2011) 069, DOI: `10.1007/JHEP04(2011)069`, eprint: `1010.3698`.

[99] V. Kostyukhin, *A possibility to gain $H \to b\bar{b}$ sensitivity by looser selection and Lorentz invariant variables*, Presentation in the ATLAS $H \to b\bar{b}$ MVA working group meeting, 26 February 2013, 2013.

[100] T. Fawcett, “ROC graphs: Notes and practical considerations for researchers”, *Machine learning* 31 (2004).

[101] D. Buescher, *Testing statistics for MVA*, Presentation in the ATLAS $H \rightarrow b\bar{b}$ Moriond analysis meeting, 30 November 2012, 2012.

[102] S. Hoeche et al., “QCD matrix elements + parton showers: The NLO case”, *JHEP* 1304 (2013) 027, DOI: `10.1007/JHEP04(2013)027`, eprint: `1207.5030`.

[103] G. Aad et al., “Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* 1307 (2013) 032, DOI: `10.1007/JHEP07(2013)032`, eprint: `1304.7098`.

[104] V. Kostyukhin, *Solution for MC problems*, Presentation in the ATLAS HSG5 analysis meeting, 26 March 2013, 2013.

[105] V. Kostyukhin, *ISR in W+jets*, Presentation in the ATLAS HSG5 analysis meeting, 9 April 2013, 2013.

[106] W Bell et al., “Measurement of top quark pair differential cross section with ATLAS in pp collisions at sqrt(s)=7 TeV: Measurement of top quark pair differential cross section with ATLAS using 5/fb of data collected in 2011 and the rel.17 version of the Atlas software”, tech. rep. ATL-COM-PHYS-2012-1137, CERN, 2012.

[107] V. Kostyukhin, *DPhi correction check*, Presentation in the ATLAS HSG5 analysis meeting, 14 May 2013, 2013.

[108] G. Aad et al., “Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC”, *Eur.Phys.J.* C73 (2013) 2518, DOI: `10.1140/epjc/s10052-013-2518-3`, eprint: `1302.4393`.

[109] “Jet energy scale and its systematic uncertainty in proton-proton collisions at sqrt(s)=7 TeV with ATLAS 2011 data”, tech. rep. ATLAS-CONF-2013-004, CERN, 2013.

[110] F Ahmadov et al., “Search for the Higgs boson in the $WH \rightarrow l\nu b\bar{b}$ and $ZH \rightarrow llb\bar{b}$ channels in ATLAS using 5fb-1 of LHC collision data”, tech. rep. ATL-COM-PHYS-2011-1648, CERN, 2011.

[111] S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, *The Annals of Mathematical Statistics* 9.1 (1938) 60–62.

[112] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *European Physical Journal C* 71 (Feb. 2011) 1554, eprint: `1007.1727`.

[113] L. Lista, “Setting limits and application to Higgs boson search”, *EPJ Web of Conferences* 55 (2013).

List of Figures

List of Tables