Università degli Studi di Padova
Dipartimento di Fisica ed Astronomia
Scuola di Dottorato di Ricerca in Fisica
Ciclo XXIX

Search for a Higgs boson decaying to a pair of b quarks in the forward region of pp collisions with the LHCb detector.

Thesis presented by
Lorenzo Sestini

Work prepared in collaboration with
Istituto Nazionale di Fisica Nucleare

Director of the Doctoral School
Prof. Gianguido Dall'Agata

Supervisor
Prof. Donatella Lucchesi

Contents

Introduction

In 2013 the ATLAS [1] and CMS [2] experiments at the Large Hadron Collider (LHC) [3] observed a boson compatible with the Standard Model (SM) Higgs, with a mass of about 125 GeV/c^2 [4]. This observation has been a fundamental step in the understanding of the mechanism which gives mass to the vector bosons, the carriers of the electroweak force, and to the fermions.

The precise measurements of the Higgs coupling with fermions is a key point in contemporary particle physics, since they can be used to test the SM and to seek for new physics hints. ATLAS and CMS found evidence of the Higgs coupling with fermions studying the leptonic decay channels [5, 6]. Neverthless the decay of the Higgs to a pair of b quarks has not been clearly observed yet by ATLAS and CMS [5] [6]. For this decay channel the experiments found a combined significance of 2.6σ by studying the Higgs associated production with a W or Z bosons. Direct experimental limit on the inclusive production of $H^0 \to b\bar{b}$ does not exist yet, but it is of great interest since it can be a probe of production mechanisms beyond the SM.

In the picture of LHC experiments, the LHCb detector [7] has a complementary role with respect to ATLAS and CMS. LHCb is an experiment initially designed for heavy flavour physics, located at the LHC. It is a forward spectrometer which primary goal has been to look for evidence of new physics in CP violation and rare decays of beauty and charm hadrons in proton-proton collisions. With the large expected $b\bar{b}$ cross section of about 500 μb at a centre-of-mass energy of 14 TeV, the LHC is the most copious source of b quarks and B mesons in the world. The LHCb detector is able to exploit this large number of b hadrons, thanks to its efficient, robust and flexible trigger, specifically designed for the hadronic environment. Excellent vertex and momentum resolution are necessary to identify B mesons and reduce the combinatorial background.

In the last years, the development of jets reconstruction and heavy flavour tagging algorithms optimized for LHCb has qualified the experiment as a General Forward Detector. Thanks to its unique features LHCb is able to perform electroweak and jets measurements in a complementary phase space with respect to the General Purpose Detectors (GPD) at LHC, ATLAS and CMS. At LHCb the jet energy is measured by the combination of the precise tracking system and of the calorimeters. Moreover, the jet heavy flavour is efficiently identified through the secondary vertex finding system.

In general the search for new $b\bar{b}$ resonances is possible at almost any energy at LHCb, not only above the Higgs, thanks to the trigger with low energy thresholds for jets. Several extensions of the SM predict new heavy particles accessible at LHC energies, that decay into two energetic b-quarks. Such new states may include an excited composite quark q^*, exemplifiying quark substructure [8], which can be studied looking for the production of a dijet in association with a photon, a massive axigluon (which current lower bound on mass is of about 1 TeV/c^2) predicted by chiral color models [9] that could extend QCD to a gauge group which is $SU(3)^L \times SU(3)^R$ or many other exotics final states.

In this thesis techniques to identify and reconstruct $b\bar{b}$ resonances with the LHCb detector are developed. The LHCb data are analyzed to measure SM processes such as $Z \rightarrow b\bar{b}$ and $W + b\bar{b}$ to validate the reconstruction tools. The goal of this work is to set experimental limit to the $H^0 \rightarrow b\bar{b}$ for different production mechanisms. The thesis is organized as follow:

- Chapter 1: an overview of the Electroweak theory and of the Brout-Englert-Higgs mechanism is given. The Higgs production at the LHC is explained and the experimental results on the Higgs are discussed;

- Chapter 2: after a brief introduction of the LHC accelerator facility, the sub-systems forming the LHCb detector are described and their performance are discussed. The LHCb trigger system is also presented. Finally an overview of the offline reconstruction algorithms and software is reported;

- Chapter 3: the jet reconstruction algorithm used and tuned by LHCb during the Run I data taking is presented and discussed in details. Heavy flavour identification algorithms used at LHCb are also presented with the measurements of their performance;

- Chapter 4: the data collected by LHCb during the Run I data taking are analyzed to identify the $Z \rightarrow b\bar{b}$ decay and to measure its cross section. The Jet Energy Scale is also measured by studying the $Z \rightarrow b\bar{b}$ sample, which is fundamental to validate the LHCb b-jet reconstruction techniques;

- Chapter 5: the data collected by LHCb during the Run I data taking are analyzed to search for the Higgs decaying to a pair of b or c quarks in association with a W or Z boson. In this Chapter the production cross sections of $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$, which are the main background in the Higgs search, are measured.

- Chapter 6: the $b\bar{b}$ sample used in Chapter 3 for the $Z \rightarrow b\bar{b}$ measurement is studied to set an upper limit on the inclusive production of $H^0 \rightarrow b\bar{b}$. Future prospects on the $H^0 \rightarrow b\bar{b}$ observation are also discussed.

Chapter 1

Higgs boson: theory and experimental results

1.1 Electroweak theory

The electroweak (EW) theory [10–12] unifies the electromagnetic and the weak interactions as a manifestation of the same force. It follows the gauge invariant structure of the Quantum Electron Dynamics (QED) [13], preserving the local phase invariance $U_Y(1)$ of the electromagnetic interaction and the isospin invariance $SU(2)_L$ of the weak interaction. Therefore the EW lagrangian is invariant under the group transformation:

$$G = U_Y(1) \times SU(2)_L. \tag{1.1}$$

The $SU(2)_L$ invariance reflects the non-conservation of the parity experimentally observed in the weak interaction [14]. It implies that the weak current couples only to left handed fermions. The left-handed and right-handed spinors are respectively defined as

$$\begin{aligned} \psi_L &= P_L\psi = \frac{1+\gamma^5}{2}\psi, \\ \psi_R &= P_R\psi = \frac{1-\gamma^5}{2}\psi, \\ \psi &= \psi_L + \psi_R, \end{aligned} \tag{1.2}$$

where ψ is the spinor, P_L and P_R are the chirality operators and γ^5 is the product of the four Dirac matrices. In the EW lagrangian the fermions appear as left-handed doublets and right-handed singlets. For the quarks they are:

$$q_L^i = \begin{pmatrix} u \\ d \end{pmatrix}_L, \ \begin{pmatrix} c \\ s \end{pmatrix}_L, \ \begin{pmatrix} t \\ b \end{pmatrix}_L, \tag{1.3}$$

$$\begin{aligned} u_R^i &= u_R, \quad c_R, \quad t_R, \\ d_R^i &= d_R, \quad s_R, \quad b_R, \end{aligned} \tag{1.4}$$

and for the leptons:

$$\ell^i_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \tag{1.5}$$

$$e^i_R = e_R, \quad \mu_R, \quad \tau_R, \tag{1.6}$$

where the index i indicates one of the three quarks or leptons families. Under $SU(2)_L$ the left-handed fields transform as doublets while the right-handed fields do not transform. The EW gauge lagrangian is expressed by:

$$\mathcal{L}_{gauge} = -\frac{1}{4}F^i_{\mu\nu}F^{\mu\nu}_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.7}$$

where the field strength tensors $F^i_{\mu\nu}$ and $B_{\mu\nu}$ are defined as

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon_{ijk}W^j_\mu W^i_\nu, \end{aligned} \tag{1.8}$$

where B_μ and $W^i_{\mu\nu}$, with $i = 1, 2, 3$, are the fields, g is a coupling constant and ϵ_{ijk} is the total antisymmetric symbol. The fermion lagrangian is written as:

$$\mathcal{L}_f = \bar{q}^j_L i\not{D} q_{jL} + \bar{\ell}^j_L i\not{D} \ell_{jL} + \bar{u}_R i\not{D} u_R + \bar{d}_R i\not{D} d_R + \bar{e}_R i\not{D} e_R, \tag{1.9}$$

where D_μ is the gauge covariant derivative:

$$D_\mu = \partial_\mu + ig\frac{\tau_j}{2}W^j_\mu + \frac{ig'}{2}B_\mu, \tag{1.10}$$

g' is a coupling constant and τ^j are the Pauli matrices. The EW theory requires the existence of four massless force carriers, two with electric charge and the other two with neutral charge. Experimentally the weak interaction is characterized by a short range, therefore its carriers must have mass. This is the reason why the underlying EW symmetry should be broken by some mechanism, which generates the massive vector bosons [15–20]. The three vector bosons are W^+ and W^-, with electric charge, and the Z, with neutral charge, and they are exchanged in weak interactions. The photon, exchanged in electromagnetic interactions, is not involved in the symmetry breaking mechanism and remains massless. The complete EW lagrangian in the SM can be written as:

$$\mathcal{L} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_\Phi + \mathcal{L}_Y, \tag{1.11}$$

where $\mathcal{L}_\Phi$ is associated to the scalar field responsible of the electroweak symmetry breaking (EWSB) and $\mathcal{L}_Y$ is the Yukawa lagrangian. These two terms are discussed in the next sections.

1.2 Brout-Englert-Higgs mechanism

The $\mathcal{L}_\Phi$ lagrangian can be written as

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger D^\mu \Phi + V(\Phi), \tag{1.12}$$

where Φ is the scalar field doublet:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.13}$$

with $\phi_{1,2,3,4}$ real scalar functions, $V(\Phi)$ is the field potential

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2, \tag{1.14}$$

and λ and μ are real constants. For $\lambda > 0$ the "mexican hat" potential is obtained: it is characterized by having a continuum set of field configurations minimizing the energy of the system. Therefore the choice of the ground state is arbitrary and without losing generalities one can assume (*spontaneous symmetry breaking*):

$$\begin{aligned} &\langle 0|\phi_i|0\rangle = 0 \text{ for } i = 1,2,4 \\ &\langle 0|\phi_3|0\rangle = \nu, \end{aligned} \tag{1.15}$$

where ν is a constant (vacuum expectation value). The linear perturbations near the ground state can be expressed as

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix} \tag{1.16}$$

where H is a real scalar field called *Higgs field*. It is convenient to rotate the W_μ^j and the B_μ fields:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\ Z_\mu &= \cos(\theta_W) W_\mu^1 - \sin(\theta_W) B_\mu, \\ A_\mu &= \sin(\theta_W) W_\mu^1 + \cos(\theta_W) B_\mu, \end{aligned} \tag{1.17}$$

with

$$\begin{aligned} \sin(\theta_W) &= \frac{g}{\sqrt{g^2 + g'^2}}, \\ \cos(\theta_W) &= \frac{g'}{\sqrt{g^2 + g'^2}}. \end{aligned} \tag{1.18}$$

Substituting 1.16 and 1.17 in 1.12 and taking the quartic order of $\mathcal{L}_\Phi$, the Higgs lagrangian is obtained:

$$\mathcal{L}_H = \frac{1}{2}\partial_\mu H \partial^\mu H + (\nu + H)^2 \left[\frac{g^2}{4} W_\mu^\dagger W^\mu + \frac{1}{8}(g^2 + g'^2) Z_\mu^\dagger Z^\mu\right] - \lambda\nu^2 H^2 - \lambda\nu H^3 - \frac{1}{4}H^4. \tag{1.19}$$

After the EWSB the W and Z bosons obtain their masses, which are given by the factors that multiply the $W_\mu^\dagger W^\mu$ and $Z_\mu^\dagger Z^\mu$ terms:

$$\begin{aligned} m_W &= \frac{1}{2}\nu g, \\ m_Z &= \frac{1}{2}\nu\sqrt{g^2 + g'^2}. \end{aligned} \tag{1.20}$$

The Higgs boson is generated by the quantization of the H field and its mass is given by

$$m_H = \sqrt{2\lambda\nu^2}. \tag{1.21}$$

The terms with $HW_\mu^\dagger W^\mu$ and $HZ_\mu^\dagger Z^\mu$ represent respectively the HWW and HZZ vertices. The mechanism described in this section that generates the EWSB is known as *Brout-Englert-Higgs mechanism* [15–20]. The vacuum expectation value is given by

$$\nu = \frac{2m_W}{g'} \simeq (\sqrt{2}G_F)^{-1/2} \simeq 246 \text{ GeV}, \tag{1.22}$$

derived from the Fermi constant G_F measured in muon decays [21].

1.3 Yukawa couplings

The mass term of fermions, according to the Dirac equation, is given by

$$-m\bar{\psi}\psi = -m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R). \tag{1.23}$$

The $SU(2)_L$ group acts only on the left components of the fermion field: this leads to the fundamental problem in which the mass term of the Dirac field is not $SU(2)_L$ invariant, in contrast with the EW theory. This inconsistency may be solved by assuming a *Yukawa coupling* of the fermions with a scalar field. For a lepton, the $SU(2)_L$ invariant Yukawa lagrangian can be written as

$$\mathcal{L}_Y^i = -g^i \left(\bar{\ell}_L^i \Phi e_R^i + \ell_L^i \Phi^\dagger e_L^i\right), \tag{1.24}$$

where g^i is a coupling constant and Φ is the scalar doublet defined in equation 1.13. The label i indicates the lepton flavour, but neutrinos are not included, since they are massless in the SM. After the EWSB, if interactions above the ground state are considered and the Higgs field expectation value is different than 0, the lepton Yukawa lagrangian becomes

$$\mathcal{L}_Y^i = -\frac{g^i\nu}{\sqrt{2}}\left(\bar{e}_L^i e_R^i + \bar{e}_R^i e_L^i\right) - \frac{g^i}{\sqrt{2}}\left(\bar{e}_L^i e_R^i + \bar{e}_R^i e_L^i\right) H = -\frac{g^i\nu}{\sqrt{2}}\bar{e}^i e^i - \frac{g^i}{\sqrt{2}}\bar{e}^i e^i H. \tag{1.25}$$

The mass of the lepton is the factor that multiplies the $\bar{e}^i e^i$ term:

$$m_i = \frac{g_i\nu}{\sqrt{2}}. \tag{1.26}$$

The $\bar{e}^i e^i H$ term describes the interaction between the lepton and the Higgs.

The interaction between the quarks and the scalar field, before the EWSB, is given by

$$\mathcal{L}_Y^q = -Y_{ij}^u \bar{Q}_L^i \Phi u_R^j - Y_{ij}^d \bar{Q}_L^i \Phi_C d_R^j + h.c. \tag{1.27}$$

where i and j indicate the quark flavour, Φ_C is defined as $\Phi_C = i\tau_2 \Phi$, Y^u and Y^d are 3×3 complex Yukawa matrices which elements have to be measured in experiments. After the EWSB the quarks acquire their mass that can be obtained by diagonalizing Y^u and Y^d with a unitary transformation V:

$$M^u = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} = \frac{\nu}{\sqrt{2}} V_L^u Y^u V_R^{u\dagger}, \tag{1.28}$$

$$M^d = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} = \frac{\nu}{\sqrt{2}} V_L^d Y^u V_R^{d\dagger}. \tag{1.29}$$

The matrix that connects the interaction eigenstates with the mass eigenstates is called Cabibbo-Kobayashi-Maskawa matrix [22, 23] and it is given by

$$V_{CKM} = V_L^d V_L^{u\dagger}. \tag{1.30}$$

One of the direct consequences of the Yukawa lagrangian is that the strength of the coupling of the Higgs with a fermion (g_f) depends from the fermion mass:

$$g_f = \sqrt{2} \frac{m_f}{\nu}. \tag{1.31}$$

1.4 Higgs production in proton-proton collisions

In proton-proton collisions the Higgs boson can be directly produced in the hard interaction. The production cross section is given by the following expression [24]:

$$\sigma_{pp \to H^0 + X} = \sum_{a,b} \int dx_a dx_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \hat{\sigma}_{ab \to H^0 + X}(Q^2),$$

where $a(b)$ indicates a parton in the proton $A(B)$, $f_{a/A}(f_{b/B})$ is the Parton Distribution Function that represents the probability density for the parton to have a proton fraction momentum $x_a(x_b)$ and a transferred momentum square Q^2, $\hat{\sigma}_{ab \to H^0 + X}$ is the partonic cross section. $\hat{\sigma}_{ab \to H^0 + X}$ can be obtained from the matrix element of the considered process. The leading $ab \to H^0 + X$ processes in proton-proton collisions are:

- $gg \to H^0$: Higgs produced via gluon-gluon fusion, mediated by a heavy quark loop in which the top is the main contribution;

- $qq' \rightarrow qq'H^0$: Higgs produced via vector bosons fusion, where two interacting quarks emit vector bosons (W^+W^- or ZZ) that fuse together in a Higgs boson;
- $q\bar{q}' \rightarrow W/ZH^0$: Higgs-strahlung, where one quark and one anti-quark produce an excited vector boson (Z^* or W^*) which de-excites emitting a Higgs boson;
- $gg \rightarrow t\bar{t}H^0$: where two gluons split in $t\bar{t}$ couples, and the Higgs is produced in the fusion of one top and one anti-top.

The Feynman diagrams for these processes are shown in figure 1.1. The estimated cross sections for the production of a Higgs boson with a mass of 125.0 GeV/c^2 in proton-proton collisions at centre-of-mass energy ($\sqrt{s}$) of 8 TeV are reported in table 1.1 [25]. They are computed at next-to-next-to-leading order (NNLO) in the QCD theory and at next-to-leading order (NLO) in the EW theory, with the exception of the $pp \rightarrow t\bar{t}H^0$ cross section which is computed at NLO in QCD.

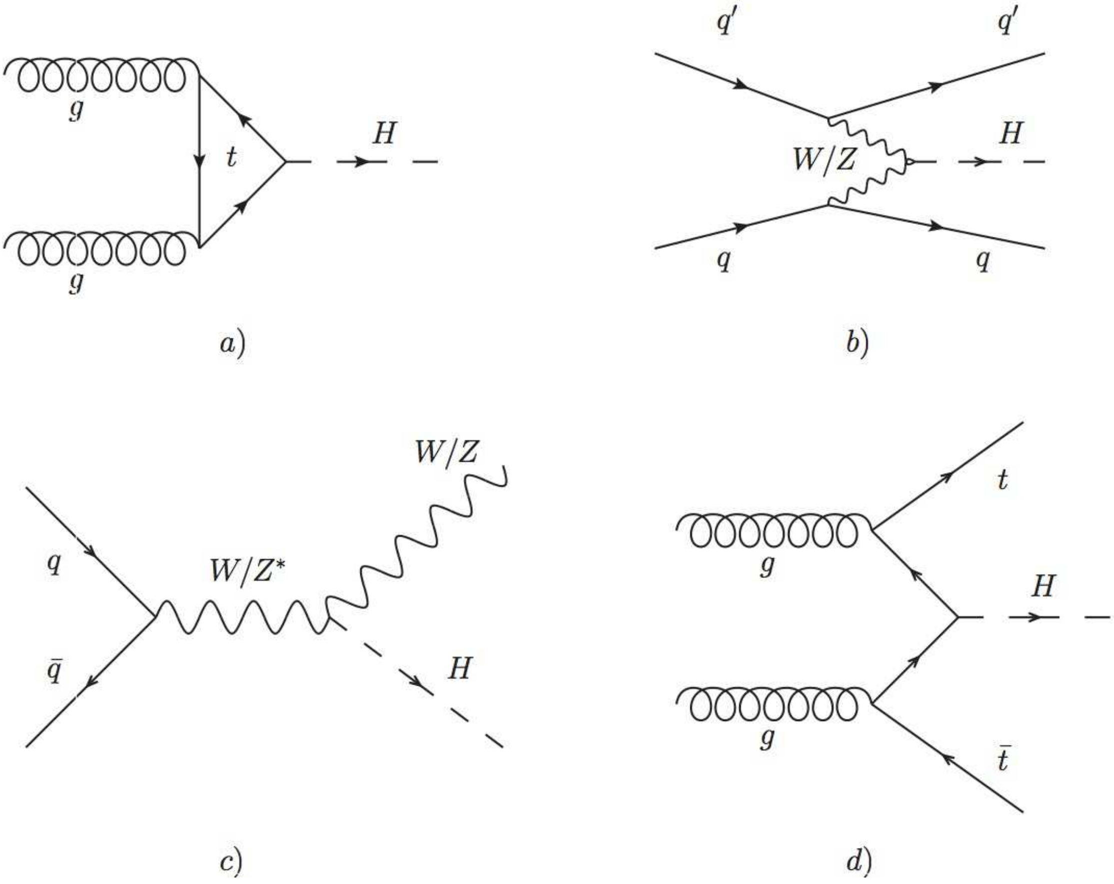

Figure 1.1: Feynman diagrams of leading Higgs production processes in proton-proton collisions: a) gluon-gluon fusion, b) vector boson fusion, c) Higgs-strahlung d) associated top production.

Table 1.1: Theoretical Higgs cross sections in pp collisions at centre-of-mass energy of 8 TeV for different production processes, assuming a Higgs mass of 125.0 GeV/c^2 [25].

Process	cross section (pb)
gluon-gluon fusion	19.27
vector boson fusion	1.578
Higgs-strahlung (WH)	0.7046
Higgs-strahlung (ZH)	0.4153
Higgs + $t\bar{t}$	0.1293

1.5 Experimental results on the Higgs boson

The ATLAS and CMS collaborations at the Large Hadron Collider observed, in proton-proton collisions, a Higgs boson compatible with the SM predictions with a mass of about 125 GeV/c^2 [4].

In the decay $H^0 \to \gamma\gamma$, the observed signal significance was found to be 5.2σ for ATLAS [26] and 5.7σ for CMS [27]. For $H^0 \to 4l$, the observed signal significance was found to be 8.1σ for ATLAS [5] and 6.8σ for CMS [6].

For H^0 produced associated with a W or Z boson, with H^0 decaying to a $b\bar{b}$ pair, the significance observed by ATLAS is 1.4σ [28], while for CMS it is 2.1σ [29], the combined significance being 2.6σ [30]. The upper limits on the cross section times the $H^0 \to b\bar{b}$ branching fraction ($\mathcal{B}$) are 1.2 times the SM prediction for ATLAS and 1.9 times the SM prediction for CMS, in both cases at 95% confidence level (CL). Previous searches for this channel have been performed by the CDF and D0 collaborations [31–36]. They observed an excess in data, corresponding to a combined global significance of 3.1σ for a mass of 125 GeV/c^2 [36]. Direct experimental limits on the inclusive production of the H^0 decaying to a $b\bar{b}$ pair do not exist yet.

In the search for $H^0 \to \tau^+\tau^-$, ATLAS published a signal significance of 4.5σ [37] and CMS found a signal significance of 3.2σ [38]. The LHCb collaboration has published a search for $H^0 \to \tau^+\tau^-$ with its 7 TeV dataset [39].

At the Tevatron the CDF and D0 collaborations searched for the Higgs produced via gluon-gluon fusion, WH^0, ZH^0, $t\bar{t}H^0$ and vector boson fusion decaying in the $H^0 \to b\bar{b}$, $H^0 \to W^+W^-$, $H^0 \to ZZ$, and $H^0 \to \tau^+\tau^-$ modes [40]. They observed a total combined significance of 3.0σ at a Higgs mass of 125 GeV/c^2.

The summary of the production cross sections measured by ATLAS and CMS, normalised to the SM expectations, for different Higgs boson final states and production mechanisms is shown in figure 1.2. The signal strengths measured by ATLAS and CMS for the combination of the Higgs decay modes, tagged for the different production processes are shown in figure 1.3.

ATLAS and CMS have found evidence that the SM Higgs boson couples to fermions. The upper and lower limits set by ATLAS and CMS on the coupling constants, as a

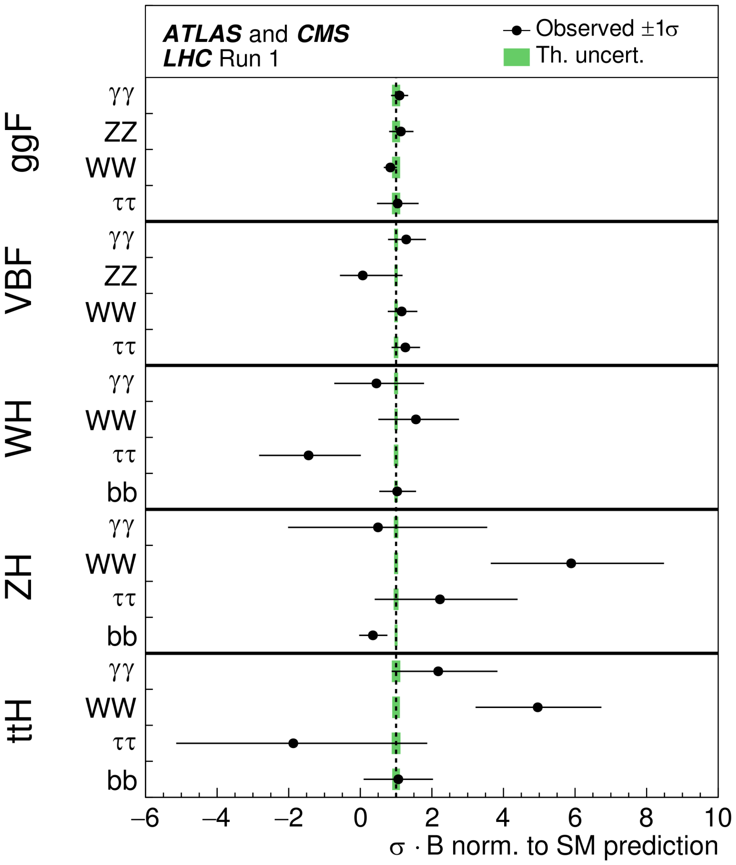

Figure 1.2: Best fit values of $\sigma_i \cdot B_f$ for each specific channel $i \to H \to f$, where σ_i is the production cross section and B_f the decay branching fraction, normalized to the SM, obtained from the combination of ATLAS and CMS. The error bars indicate the 1σ intervals [41].

function of the particle mass, are shown in figure 1.4.

ATLAS and CMS have measured the mass of the Higgs boson to be $m(H^0) = 125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV [42].

Moreover, no direct experimental limits exists for $H^0 \to c\bar{c}$, although indirect constraints exist exploiting the information from other searches [43]. Furthermore, the search for the decay $H^0 \to J/\psi\gamma$ also provides information about the coupling of the Higgs boson to the c quark. Both ATLAS [44] and CMS [45] have performed this search, providing an upper limit $\mathcal{B}(H^0 \to J/\psi\gamma) \lesssim 1.5 \times 10^{-3}$, which is about 500 times the SM prediction. Finally, from the bounds on the total branching fraction of unobserved Higgs boson decay modes, the $H^0 \to c\bar{c}$ branching ratio can not exceed ∼20% at 95% CL, or ∼50% if an additional new physics source of gluon fusion production is assumed [43, 46]. It has been suggested [47] that physics beyond the SM could enhance the $\mathcal{B}(H^0 \to c\bar{c})$ with respect to other decay modes.

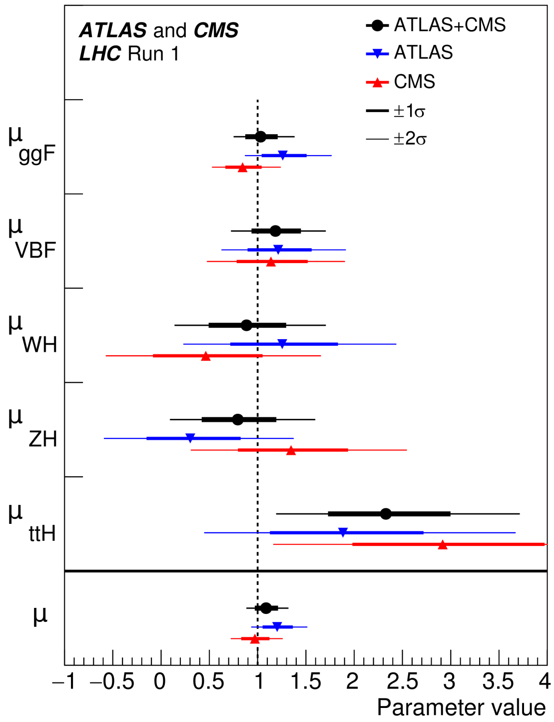

Figure 1.3: Best fit results for the production signal strengths for the combination of ATLAS and CMS data. The results for each experiment are also shown. The error bars indicate the 1σ (thick lines) and the 2σ (thin lines) intervals [41].

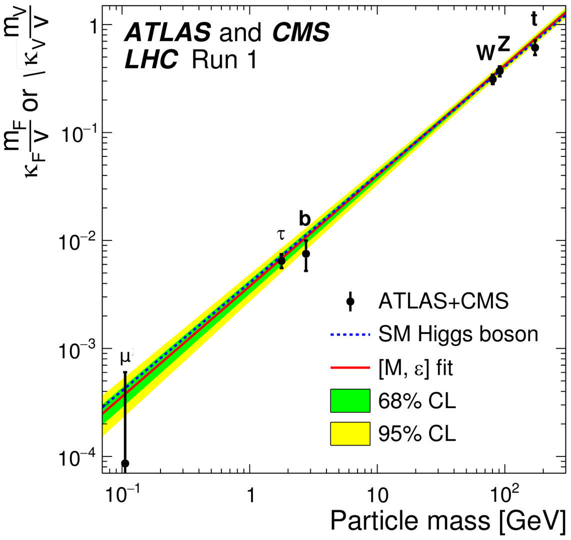

Figure 1.4: Best fit values of the coupling constants obtained from the combination of ATLAS and CMS data. The parameters k_i represent the deviation from the SM predictions. $[M, \epsilon]$ represents the phenomenological model used in the fit [41].

Chapter 2

LHCb detector description and performance

2.1 The Large Hadron Collider

The LHC [3] is a proton-proton circular collider built and operated by the European Organization of Nuclear Research (CERN), located near the Franco-Swiss border, in the vicinity of Geneve. The LHC tunnel lies 100 m underground and it is 26.7 km long. The two counter-rotating proton beams circulate in separated beam pipes, where hard vacuum has been obtained. The beam pipes intersects in the interaction points, where the experiments are placed. The protons path is bent by a series of 1232 dipolar superconducting magnets that are cooled by liquid helium to a temperature of 1.9 K, in order to achieve a peak dipole field of 8.33 T. The beams are focused and shaped by an additional sets of 3500 quadrupolar magnets that are likewise held at cryogenic temperatures during the operations. The protons acceleration is achieved by radiofrequency (RF) cavities, which are metallic chambers containing an oscillating electromagnetic field. The field in RF cavities is made to oscillate at a frequency of 400 MHz. The 16 RF cavities are housed in four cylindrical refrigerators and they are driven by high power klystrons.

LHC benefits from the CERN accelerators facility, which provide the initial acceleration stages of the proton beams. Protons are initially produced from a ionizing hydrogenum source and are accelerated up to 50 MeV by a Linac. Then they are injected into the circular Proton Synchrotron Booster (PSB) where they are accelerated to 1.4 GeV. The energy is successively raised to 28 GeV by the Proton Synchrotron (PS) and up to 450 GeV by the Super Proton Synchrotron (SPS). Finally they are injected in the LHC, which can accelerate them to the nominal energy of 7 TeV per beam.

The beams are not continuous, but they are structured in proton bunches, around 30 cm long and with a transverse dimension in the order of one mm, fully squeezed in 16 μm at the collision points. In the nominal configuration the proton beam is formed by 2808 bunches, each containing around 10^{11} protons, and spaced by 25 ns. The beam collision frequency is 40 MHz and the nominal peak luminosity is $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

The four main experiments are located at four interaction points. ATLAS (A Toroidal

LHC ApparatuS) [1] and CMS (Compact Muon Solenoid) [2] are the General Purpose Detectors. ALICE (A Lead Ion Collision Experiment) [48] is dedicated to the study of heavy ions collisions, to investigate the state of matter where quarks and gluons are freed. The last collision point is where the LHCb detector is housed.

In figure 2.1 the CERN accelerator system is schematized and the positions of the experiments are indicated.

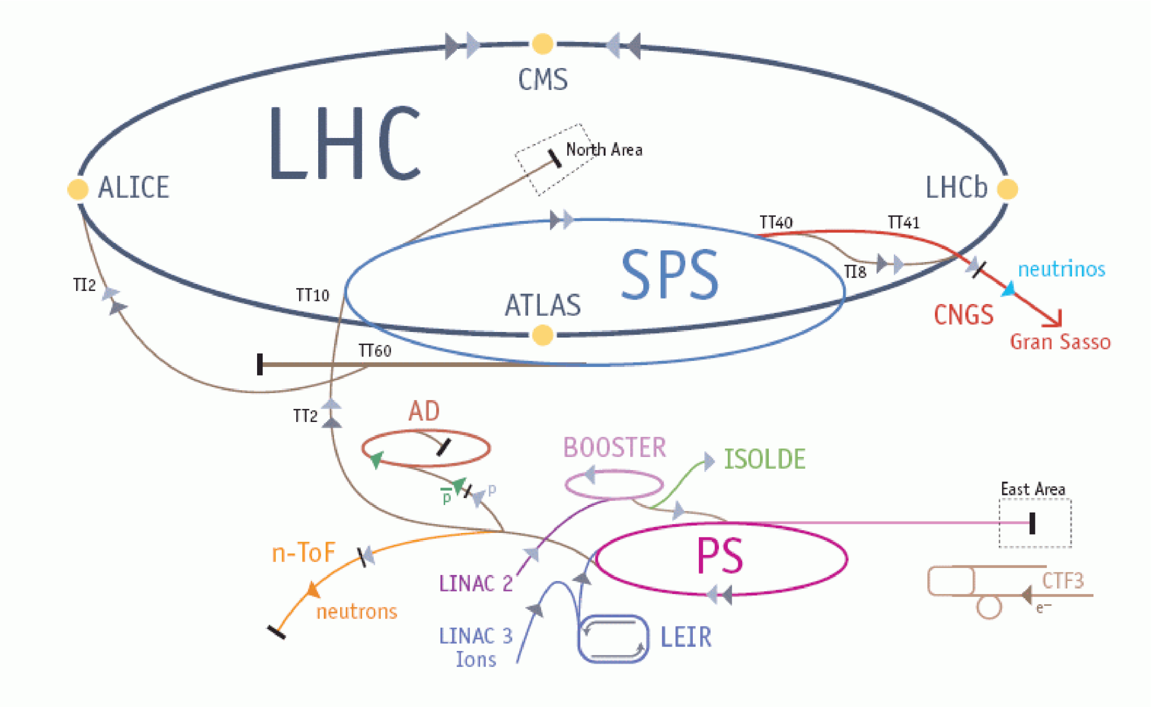

Figure 2.1: Schematization of LHC and CERN accelerators facility. The four interaction points, where the ATLAS, CMS, ALICE and LHCb are housed, are indicated.

2.2 Detector layout and operations

LHCb is a single-arm spectrometer with a forward angular coverage from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. It corresponds to a pseudorapidity[1] coverage between 1.8 and 4.9. The LHCb experiment is composed of several sub-detectors and its layout is presented in figure 2.2. In the standard LHCb coordinates system the z axis is parallel to the beam direction and the y axis (vertical direction) is parallel to the gravity acceleration. The x axis (horizontal direction) is orthogonal to the y and z axis. The LHCb sub-detectors are:

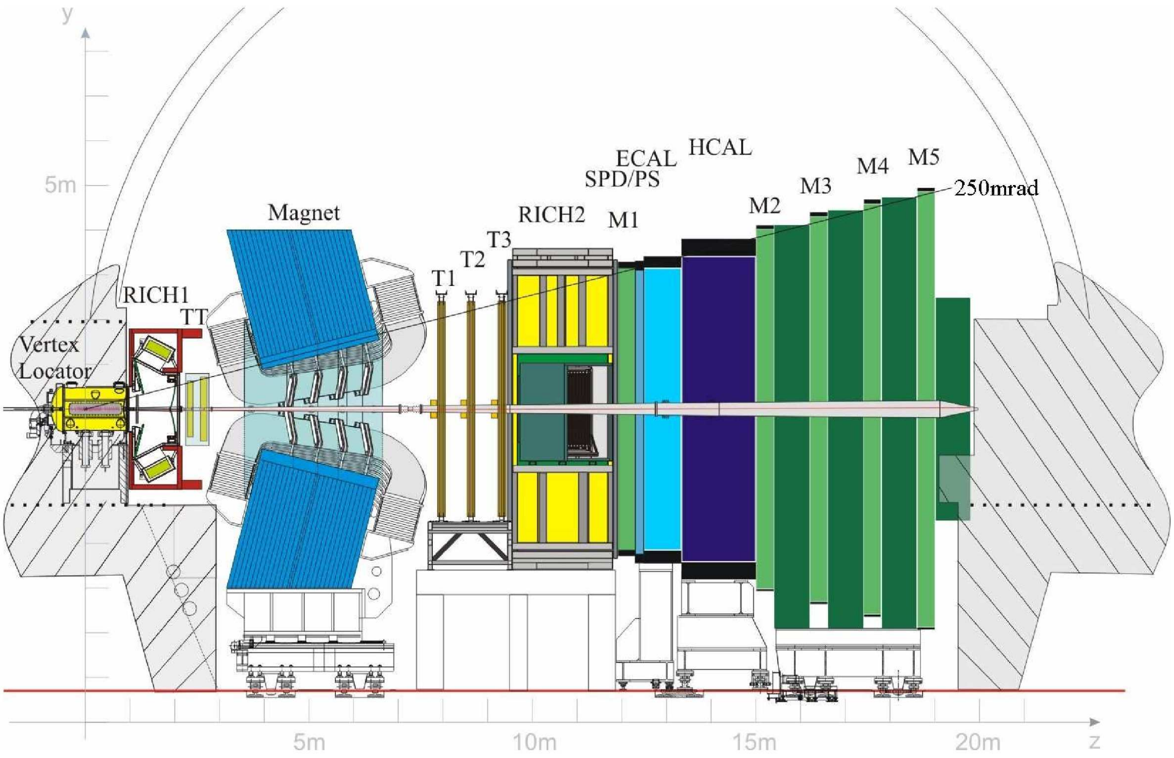

Figure 2.2: Layout of the LHCb detector during the Run I data taking. All the sub-systems are indicated. The z axis is parallel to the beam direction and the y axis is parallel to the gravity acceleration.

- a vertex locator system (VELO) located near the interaction point;
- a tracking system composed of four silicon microstrips stations, one called Trigger Tracker (TT), located upstream of the spectrometer magnet, and the other three (T1, T2 and T3) located downstream of the magnet;

---

[1] A particle pseudorapidity is defined as $\eta = -ln\left[tan\left(\frac{\theta}{2}\right)\right]$, where θ is the angle formed by the particle momentum and the beam axis.

- two rich imaging Cherenkov detectors (RICH1 and RICH2) used to achieve excellent π-K separation in the momentum range from 2 to 100 GeV/c;
- a calorimeter system composed of a Scintillator Pad Detector and Preshower (SPD/PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL);
- a muon detection system mainly formed of multi-wire proportional chambers (MWPC), while in the highest rate region triple-GEM detectors are used.

The optimal luminosity for the experiment is such that there are a maximum of 2.5 proton-proton interactions per bunch crossing. For this reason the luminosity provided by the LHC is reduced using a *luminosity levelling* technique, where the two beams do not collide head-on but they are tilted to obtain a larger interaction area. In this way the luminosity delivered to LHCb is constant to a level of $10^{32}\text{cm}^{-2}\text{s}^{-1}$. In the first three years of data taking (Run I) LHCb recorded proton-proton collisions at a centre-of-mass energy ($\sqrt{s}$) of 7 TeV in 2010-2011 and of 8 TeV in 2012. The total integrated luminosity ($\mathcal{L}_{int}$) was 0.037 fb^{-1}, 1.0 fb^{-1} and 2.0 fb^{-1} respectively in 2010, 2011 and 2012. The average number of interactions per bunch (μ) and the peak luminosity ($\mathcal{L}_{peak}$) are reported in table 2.1. Nowadays LHCb is currently taking data in the Run II campaign, started in 2015, at a centre-of-mass energy of 13 TeV. The data analysed in this thesis are those collected in the year 2012.

Table 2.1: Integrated luminosity, centre-of-mass energy, average number of interactions per bunch crossing and peak luminosity for the three years of data taking in the Run I campaign.

year	$\mathcal{L}_{int}[\text{fb}^{-1}]$	$\sqrt{s}[TeV]$	μ	$\mathcal{L}_{peak}$
2010	0.03	7	1-2.5	$1.6 \times 10^{32} cm^{-2}s^{-1}$
2011	1.0	7	1.5-2.5	$4.0 \times 10^{32} cm^{-2}s^{-1}$
2012	2.2	8	1.8	$4.0 \times 10^{32} cm^{-2}s^{-1}$

2.3 Vertex Locator

The VErtex LOcator (VELO) is the sub-system devoted to the measurement of tracks coordinates near the interaction point, necessary to identify secondary vertices associated to b and c hadrons decay. The VELO consists of two different types of silicon sensor, one for the measurement of the radial coordinate (r), from 8 mm to 42 mm, the other for the measurement of the azimuthal coordinate (ϕ), from 15 mrad to 390 mrad. This coordinates system has been chosen for a fast track reconstruction at trigger level. The two type of sensors differ for the geometry of the strips segmentation and they have a diameter of 84 mm and a thickness of 300 μm. The $r\phi$ geometry is illustrated in figure

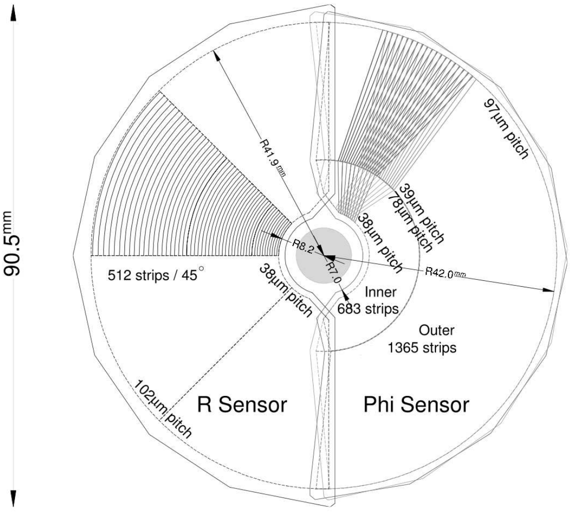

Figure 2.3: Illustration of the VELO sensors $r\phi$ geometry.

2.3. Inside the VELO there are 21 stations of these semicircular sensors, placed along a distance of 1 m, parallel to the beam axis. They are placed into a vessel were the vacuum is maintained, separated from the machine vacuum by alluminium sheets called *RF-foils*. The foils are very thin to reduce multiple scattering effects between the interaction point and the first hit of the track.

The 21 semicircular tracking stations surround the beam axis at a nominal working distance of 8 mm. Since the distance is smaller with respect to the beam width during the injection phase, the sensors are retractable. The VELO stays in the open configuration until the beam conditions are stable, therefore it is closed. The layout of the sensors in the close position is displaced in figure 2.4.

The VELO layout has been optimized to minimize the amount of material in the acceptance while providing a good geometrical coverage. The individual hit resolutions of the sensors have been measured during test beams, and the best raw resolution obtained is around 7 μm.

2.4 Magnet

A dipole magnet is used to measure the momentum of charged particles. A charged particle is bent when passing through a magnetic field, and from the measurement of the trajectory curvature its momentum can be determined. The magnet is designed to fit the detector geometry and the momentum measurements covers the forward acceptance of ±250 mrad vertically and of ±300 mrad horizontally. It is located between the TT station and the T1-T3 stations. The generated magnetic field is directed along the y axis, therefore the xz plane is the bending plane. The magnet is warm, not super-conductive, and it is composed

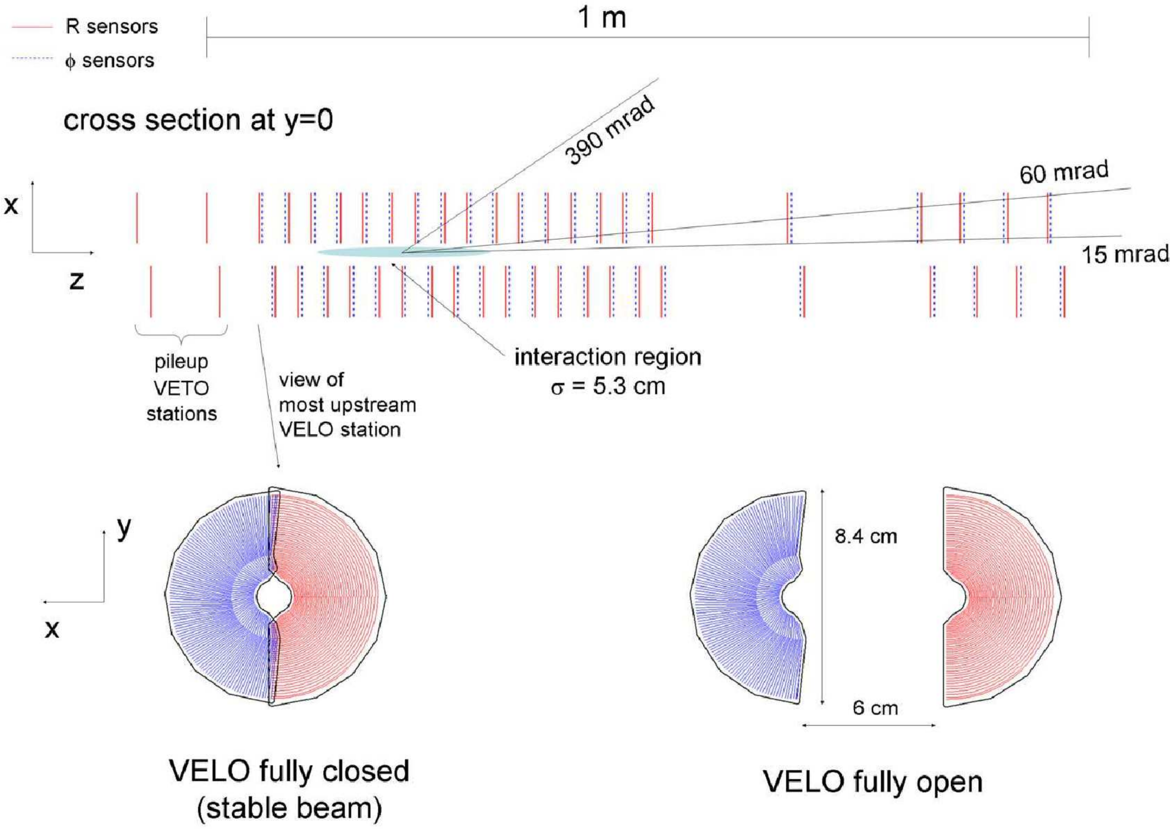

Figure 2.4: VELO layout in the fully closed position. The front face of the first modules is also illustrated in both closed and open positions.

by two identical alluminium coils of conical shape, placed mirror symmetrically to each other in the magnet yoke. A view of the magnet is given in figure 2.5. The integrated magnetic field is 4 Tm for tracks of 10 m length while the residual magnetic field in the RICHs is less the 2 Tm. The electric current that generates the magnetic field can be reverted, allowing measurements with two opposite field polarities.

2.5 Tracking System

The tracking system is composed by four planar stations, orthogonal to the beam axis: the Trigger Tracker (TT) and the T1, T2, T3 stations. The TT station and the inner region of T1, T2 and T3, called Inner Tracker (IT) use silicon microstrip sensors with a strip pitch of about 200 μm. Each one of the silicon tracker stations has four detection layers of three different type: the x-type layer has vertical strips, the u- and v-type stereo layers have strips rotated respectively to $-5°$ and $+5°$ with respect to the y axis. The arrangement scheme followed is x-u-v-x starting from the layer closest to the interaction point. The

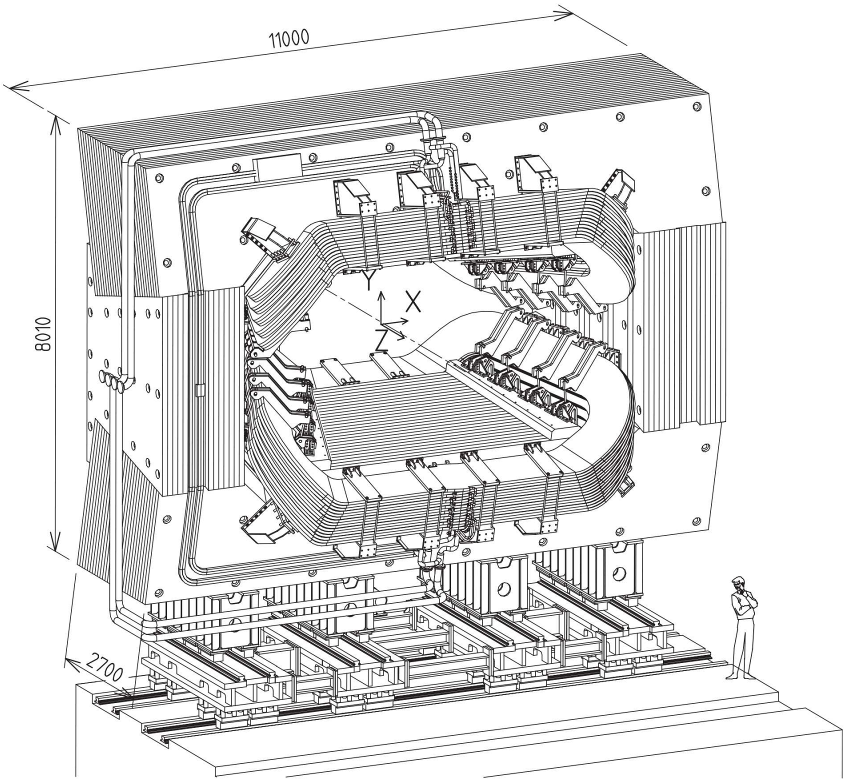

Figure 2.5: Representation of the LHCb dipolar magnet, units are in mm.

TT station is located downstream of the dipole magnet, 150 cm wide and 130 cm high, it covers the full acceptance of the experiment. The third layer (v) of the TT station is represented in figure 2.6. Given the proximity to the magnet, the TT station is affected by a residual magnetic field. The IT located in the centre of T stations is 120 cm wide and 40 cm high and its layout is shown in figure 2.7. The T stations are placed upstream of the magnet and the acceptance region covered by the IT is approximately 1.3% of the LHCb acceptance. In this region the particle flux is high and the silicon dectors must have radiation protection properties. The single hit spatial resolution has been determined to be around 50 μm for both the TT and the IT.

The external region of T stations is called Outer Tracker (OT) and it is a drift-time detector. The OT is designed as an array of straw-tube modules. Each module is formed by two staggered layers of drift-tubes with an inner diameter of 4.9 mm. The tubes are filled with counting gas, a mixture of Argon (70 %), CO_2 (28.5%) and O_2 (1.5 %). These are chosen to guarantee a sufficient drift distance resolution of about 200 μm. The OT covers an acceptance of 300 mrad (250 mrad) in the horizontal (vertical) plane.

A schematization of the tracking system with the OT, IT and TT detectors in evidence

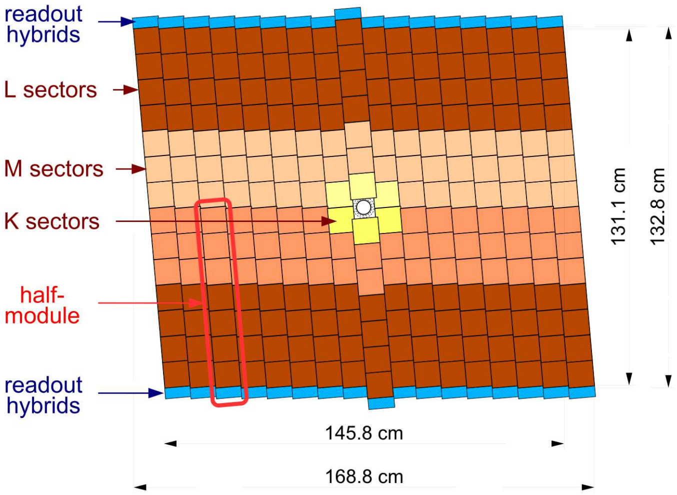

Figure 2.6: Layout of the third layer (v) of the TT station. The readout sectors are indicated with different colours.

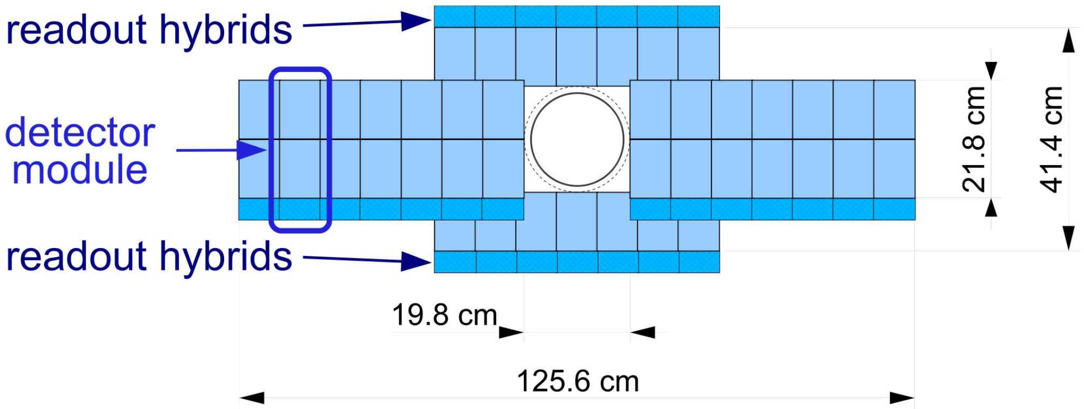

Figure 2.7: Layout of a x of one IT station.

is presented in figure 2.8.

2.6 Tracking performance

The tracking stations, together with the VELO, are used to measure the trajectory of charged particles, allowing the determination of the momentum from their curvature. Pattern recognition algorithms are used to reconstruct the track trajectory starting from the hits coordinates in all the tracking sub-detectors. In order to find the best estimate of the tracks parameters a Kalman fit [49] is performed. The tracks are classified in four

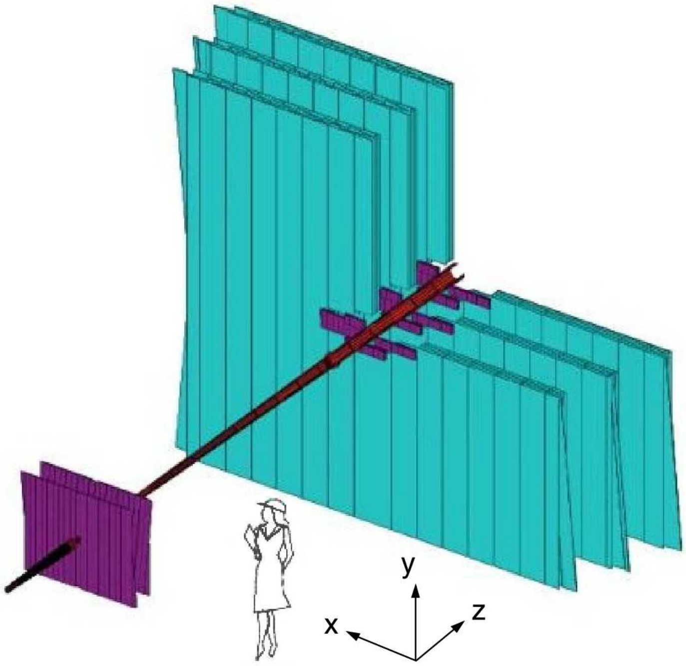

Figure 2.8: Layout of OT straw tubes (light blue) and silicon detectors (purple) in the LHCb tracking system.

categories:

- *long* tracks, which have hits in the VELO and in all the T-stations;
- *downstream* tracks, which have hits in the VELO and in the TT station but not in the T1, T2 and T3 stations;
- *upstream* tracks, which have hits in all the T-stations but not in the VELO;
- VELO tracks, which have hits in the VELO only.

The momentum resolution of long tracks has been measured using a data samples of $J/\psi \to \mu^+\mu^-$ decays, collected with a trigger configuration that selects couples of high energy muons (dimuons) [50]. The resolution as a function of the muon track momentum is shown in figure 2.9. It is of about 0.5% for momentum particles below 20 GeV/c and 0.8% for momentum particles around 100 GeV/c. The invariant mass resolution has been studied using six resonances observed in the dimuon data sample: J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ mesons and Z boson [50]. The invariant mass resolution (σ_m/m where σ_m is the peak width of the resonance with mass m) as a function of the invariant mass is shown in figure 2.10.

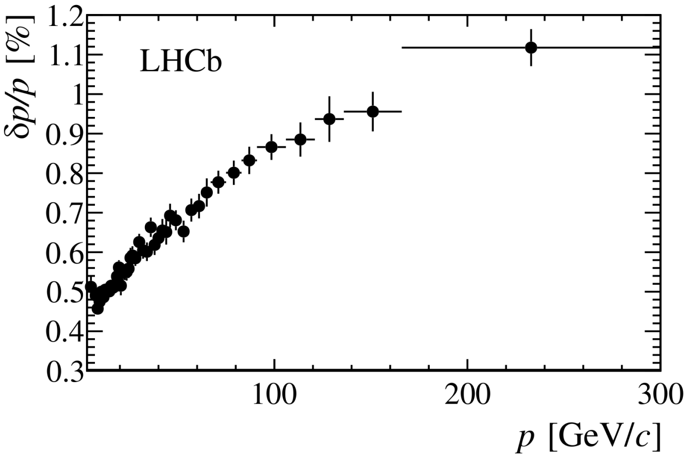

Figure 2.9: Relative momentum resolution as a function of the particle momentum, measured in data using muon tracks from J/ψ decays [50].

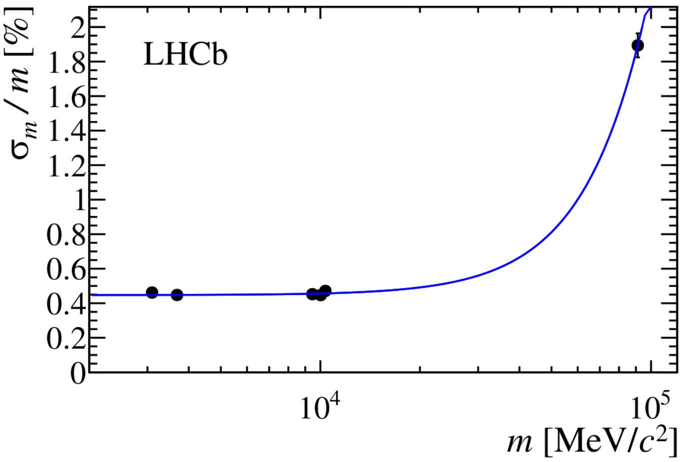

Figure 2.10: Relative mass resolution as a function of the dimuon resonance mass measured in data [50].

The primary vertex (PV), defined as the pp interaction point where particles are generated, is reconstructed from the detected tracks. First tracks are clustered in seeds, using the algorithm described in [51]. Then for each seed the PV position is calculated by minimizing:

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \frac{d_{0i}^2}{\sigma^2 d_{0i}},$$

where d_0 is the impact parameter of the track and σ_{d_0} its error. If one or more tracks have $d_0/\sigma_{d_0} > 4$ then the one with the highest d_0/σ_{d_0} is removed from the cluster, and a new PV position is calculated by minimizing χ^2_{PV}. If the cluster has less then 6 tracks then it is discarded. The procedure stops if there are no more tracks to discard.

The PV reconstruction efficiency and resolution have been obtained in simulation [51] and it depends from the tracks multiplicity. The average efficiency goes from 90.0% to 97.5%, with a probability to reconstruct a false PV of about 1%. The resolution is of about 8 μm, 10 μm and 50 μm for the x, y and z coordinates respectively.

2.7 Ring Imaging Cherenkov detectors

Since the particle identification (PID) is a fundamental requirement for LHCb, it is important to separate pions from kaons. This is achieved by two Ring Imaging Cherenkov detectors, RICH1 and RICH2, which primary role is the identification of charged hadrons (π, K, p). In addition the RICH system can contribute to the identification of charged leptons (e, μ) complementing the information of the calorimeters and of the muon system.

If a charged particle passing through a RICH radiator has a velocity larger then the speed of light in the same medium, it emits a cone of Cherenkov light, with axis parallel to the particle speed. The cone angle depends on the particle velocity, therefore its measurement, combined with the momentum measurement from the tracking system, can be used to determine the particle mass. The particle velocity (v) can be computed from the following formula:

$$cos(\theta_c) = \frac{c/n}{v},$$

where θ_c is the cone angle, n is the radiator refraction index and c is the speed of light.

Due to the difficulty of covering the full momentum range with only one radiator type, two RICH detectors are used instead of one. The RICH1 is located upstream of the magnet, between the VELO and the TT station, and it has a wide acceptance covering almost the full LHCb acceptance, from ±25 mrad to ±300 mrad along the horizontal direction and from ±25 mrad to ±250 mrad along the vertical direction. The RICH1 contains a solid aerogel radiator and a fluorobutane (C_4F_10) gas radiator, covering the low momentum charged particles from 1 to 60 GeV/c. Spherical and plane mirrors reflect the Cherenkov photons, directing them to photon detectors (Hybrid Photo Detectors, HPDs) where they are converted into electrons. On the focal planes the Cherenkov light forms rings which radius determines the Cherenkov angle. The HPDs can detect Cherenkov photons in the 200-600 nm wavelenght range and they are surrounded by external iron shields to permit operations in the residual magnet field. A schematic view of the RICH1 detector is presented in figure 2.11.

The RICH2 is located downstream of the magnet, between the last tracking station (T3) and the first muon station (M1). It has a limited angular coverage from ±15 mrad to ±120 mrad along the horizontal direction and from ±15 mrad to ±100 mrad along the vertical direction, but it covers the region where high momentum particles are produced.

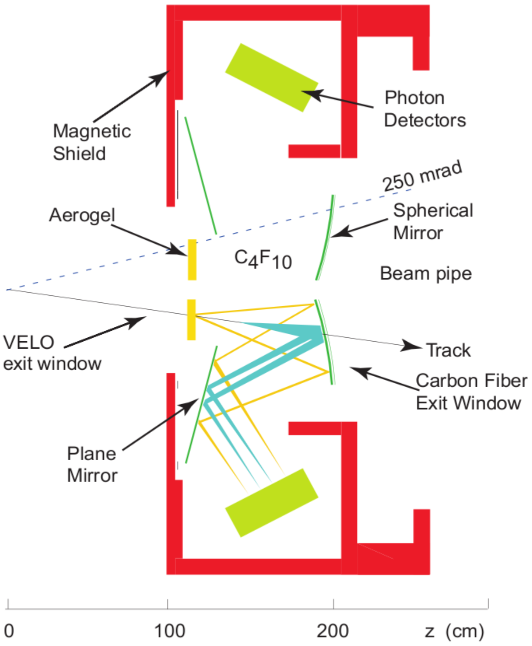

Figure 2.11: Schematic view of the RICH1 detector.

The RICH2 is filled with a CF_4 gas radiator which provides particle identification from 15 GeV/c to 100 GeV/c. As for the RICH1, in the RICH2 the Cherenkov light is collected and detected by a system of mirrors and HPDs. A schematic view of the RICH2 detector is presented in figure 2.12. The Cherenkov angle as a function of the particle momentum, for different particles and RICH radiators, is shown in figure 2.13.

2.8 Calorimeters

The calorimeters system provides several functions:

- it selects hadrons, electrons and photons with significant transverse energies at the first level trigger (L0), making a decision 4 μs after the interaction;
- it provides the identification of electrons, photons and hadrons;
- it measures the energy of neutral particles, such as photons, π^0 and neutral hadrons, which is a fundamental feature for jet reconstruction.

The calorimeters system is composed by a scintillating pad detector (SPD), a preshower detector (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) located in this order downstream of the RICH2 and M1 muon station, before the M2-M5 stations.

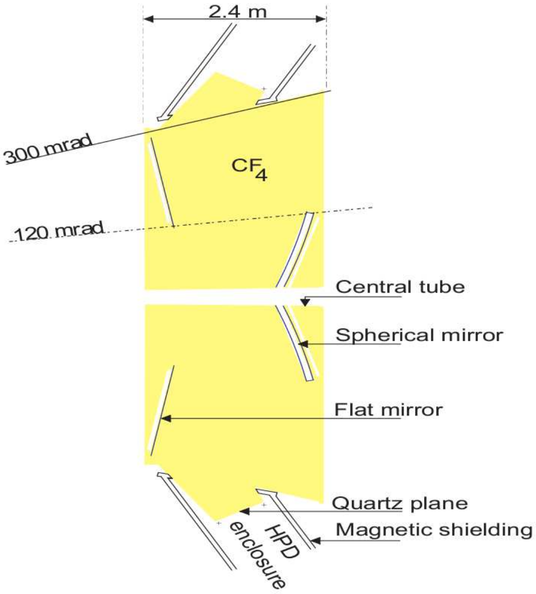

Figure 2.12: Schematic view of the RICH2 detector.

The SPD/PS detectors are devoted to the electrons and photons identification. The covered angular acceptance is between ±30 mrad and ±300 mrad horizontally and between ±30 mrad and ±250 mrad vertically. The SPD/PS system consist in a 15 mm lead converter sandwiched into two almost identical planes of rectangular scintillator pads of high granularity. The scintillator plane closest to the interaction point is called SPD and the other is called PS. The scintillation light is collected by wavelenght-shifting (WLS) fibres which are read by photomultiplier tubes (PMT). In the SPD the photons do not produce any signal before triggering a shower, while the electron does. On the other hand photons which pass through the lead after the SPD generate showers in the PS. For this reason the SPD/PS system can be used to separate electrons from photons. Moreover hadrons do not produce showers, allowing the PS to distinguish them from electrons. At trigger level a cut on the maximum number of hits in the SPD ($nSPD$) is applied. This requirement is called Global Event Cut and it is used to reject events with multiple interactions.

The ECAL is placed 12.5 m from the interaction point and its acceptance is ±300 mrad horizontally and ±250 mrad vertically, limited in the inner region around the beam pipe at ±25 mrad. A ECAL cell is built alternating layers of 2 mm thick lead tiles and 4 mm scintillator tiles. In depth 66 layers form a 42 cm stack, corresponding to 25 radiation lengths. As for the SPD/PS the scintillation light is transmitted by WLS fibres and read by PMTs. Photons and electrons deposit their energy in the ECAL, where they are absorbed. The segmentation of the ECAL cells achieved a one-to-one projective correspondence with the SPD/PS pads. The granularity depends on the distance from the beam axis to

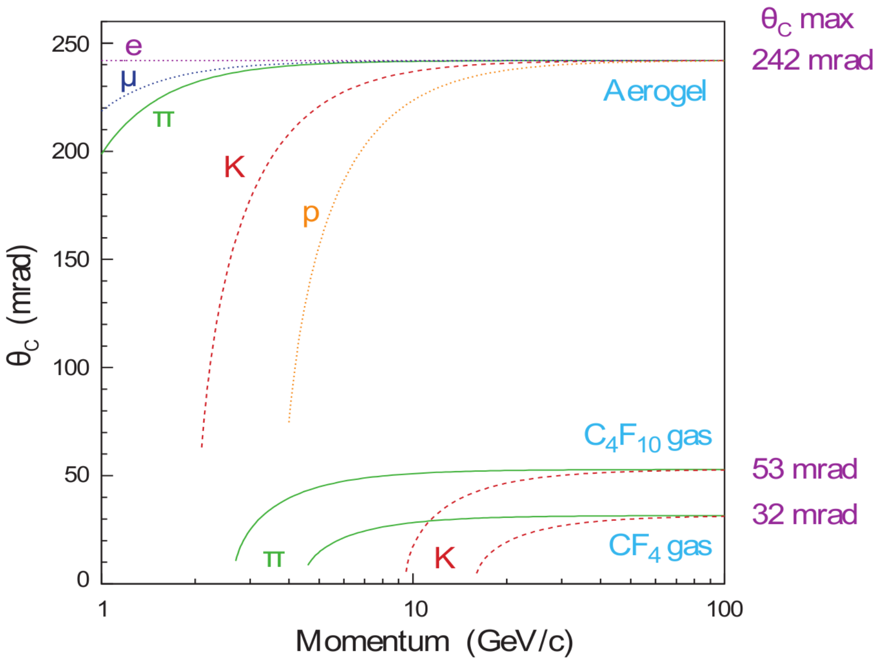

Figure 2.13: Cherenkov angle as a function of the particle momentum, for different particles and radiators.

guarantee an optimal detector occupancy, as shown in figure 2.14. The ECAL energy resolution is given by

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%,$$

where the E is the energy measured in GeV and $\oplus$ indicates the sum in quadrature.

The HCAL is made of iron and scintillating tiles, as absorber and active material respectively. In the HCAL neutral and charged hadrons deposit their remaining amount of energy. It is placed 13.3 m from the interaction point covering the same ECAL angular acceptance. The orientation of the scintillating tiles run parallel to the beam axis, and they are interspersed with 1 cm of iron, reaching in total 5.6 nuclear interaction lengths of hadrons in iron. Again the scintillation light is transmitted by WLS fibres and read by PMTs. The HCAL granularity is different from ECAL and it is shown in figure 2.14. The HCAL energy resolution is given by

$$\frac{\sigma_E}{E} = \frac{69\%}{\sqrt{E}} \oplus 9\%,$$

where the E is the energy measured in GeV.

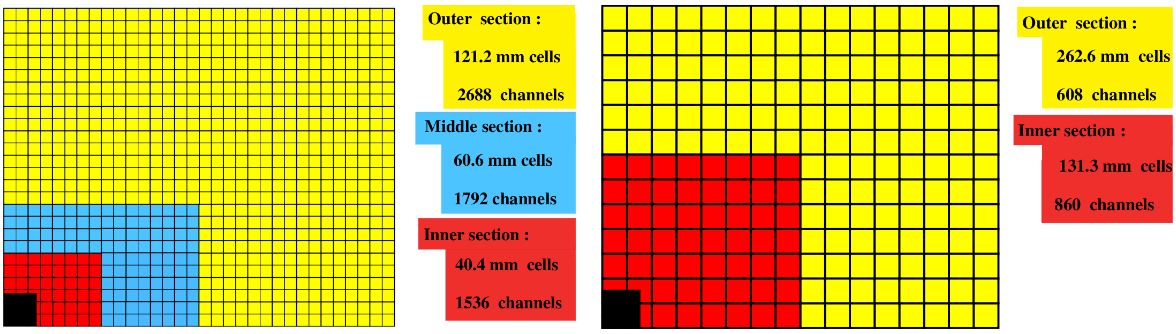

Figure 2.14: Segmentation of the SPD/PS and ECAL (left) and the HCAL (right). A quarter of the front plate is represented. The cell dimensions on the left are for the ECAL only.

2.9 Particle identification performance

Using the information from the RICH detectors a likelihood function is calculated for each particle mass hypothesis. The variable used for the particle identification is the difference of logarithmic likelihood $DLL_{rich}(A-B) = ln\mathcal{L}_A - ln\mathcal{L}_B$, where $\mathcal{L}_A$ ($\mathcal{L}_B$) is the likelihood for mass hypothesis A (B). As general criteria the particle is identified as A if $DLL_{rich}(A-B)$ is measured to be above a certain threshold. The particle identification performances are measured with two parameters:

- the identification efficiency for A is defined as the number of A particles correctly identified to the total number of detected A;
- the B misidentification rate is defined as the number of B particles wrongly identified as A to the total number of detected B.

Identification efficiencies and misidentification rates have been measured in data as a function of particles momentum, using genuine K, π and p tracks from $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$ samples collected using a trigger configuration that selects high energy hadrons [50]. In figure 2.15 the kaon identification efficiencies and the pion misidentification rates are shown for two different $DLL_{rich}(K-\pi)$ requirements. The average kaon efficiency for $DLL_{rich}(K-\pi) > 0$ is of about 95% and the pion misidentification rate is of about 10%.

In analogy with the RICH, a $DLL_{calo}(A-B)$ is defined using calorimeters observables, as the difference of logarithmic likelihood for different mass hypotheses. The electron identification efficiency is 91.9% with $DLL_{calo}(e-h) > 2$, where h indicates a hadron, with a hadron misidentification rate of 4.5%. Including the information from the RICH detectors the electron efficiency raises up to 97% while the misidentification rate becomes lower than 2%.

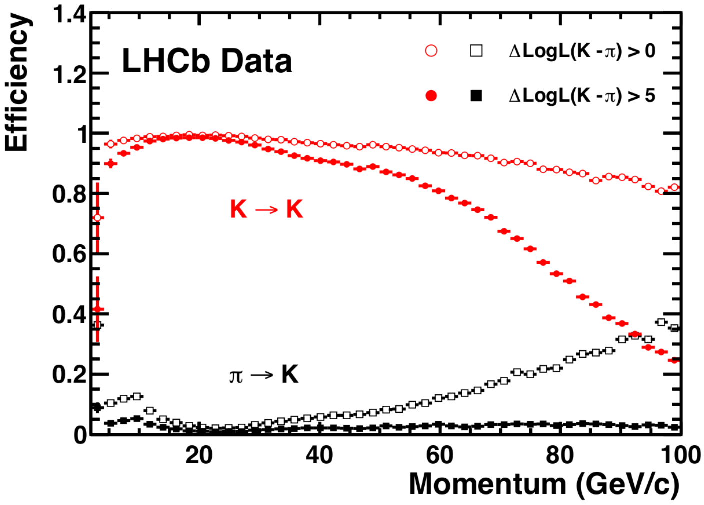

Figure 2.15: Kaon identification efficiencies and pion misidentification rates for two different $DLL(K-\pi)$ requirements, as a function of particle momentum [50].

2.10 Muon system

The muon system is the outermost detector from the interaction point, and it is devoted to the muon identification. It also provides fast information on the muons transverse momentum for the trigger. The system is composed of five stations (M1-M5) of rectangular shape, place along the beam axis, as shown in figure 2.16. The muon system angular acceptance is between ±20 mrad and ±306 mrad horizontally and between ±16 mrad and ±258 mrad vertically. The M1 station is located after the RICH2 in front of the calorimeters and it is used to improve the transverse momentum measurement in the trigger. The M2-M5 stations are placed downstream of the calorimeters and they are interleaved with iron absorbers 80 cm thick to select penetrating muons. The total absorbers thickness is approximately 20 interactions lengths, therefore the minimum momentum for a muon to cross the five stations is 6 GeV/c. The geometry and segmentation of the five stations is projective, all dimensions scale with the distance from the interaction point. The stations provide space point measurements of the tracks. They are divided into *logical pads* that defines the x and y spatial resolutions.

The detection technique has been designed to cope with the intense flux of particles and with the trigger time constraints. Multi-wire proportional chambers (MWPC) are used for all regions with the exception of the inner region of station M1 where triple-GEM [52] detectors are used. Each station from M2 to M5 contains 276 chambers for a total of 1368 MWPC. The chambers layout in one of this station is shown in figure 2.17. The chambers are filled with a gas mixture of $Ar/CO_2/CF_4$ in a 40:55:5 proportion. A total number of $3 \cdot 10^6$ wires of gold-plated tungsten with 2 mm spacing is placed inside the chambers, achieving a ionization gain of about 10^5 at a voltage of 2600-2700 V. The time resolution

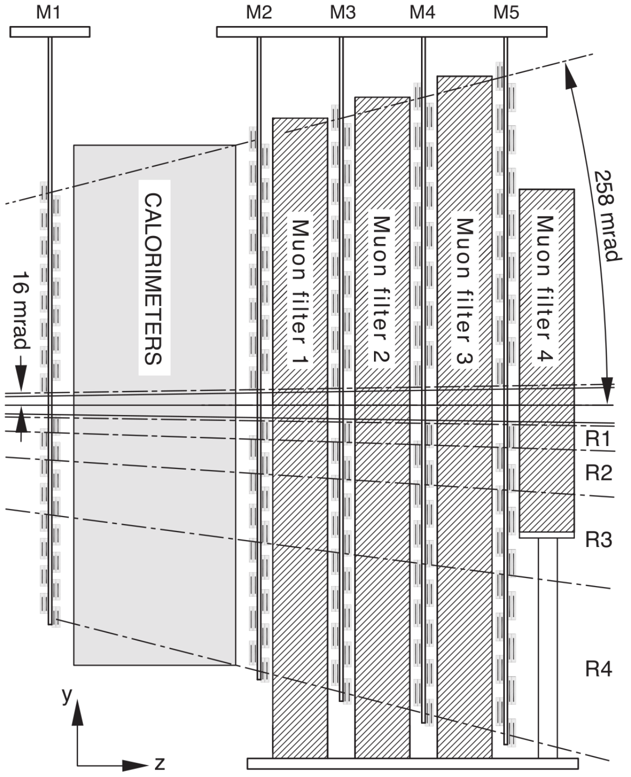

Figure 2.16: Side view of the muon system.

achieaved is around 5 ns. Each of the GEM detector in the inner region of M1 consists of three gas electron multiplier foils sandwiched between anode and cathode plates. The fast Ar/CO_2/CF_4(45:15:40) gas mixture allows a time resolution better than 3 ns. The GEMs are used instead of MWPC to sustain the high rate of 500 kHz/cm^2 charged particles in this region.

A particle is identified as a muon if it has hits in all the five stations placed in a straight line. The average muon identification efficiency is around 98% while the pion and kaon misidentification is below 1%. The muon p_T is evaluated at trigger level using the hits in the first two stations.

2.11 Trigger

LHCb operates at an average luminosity of $2\times10^{32}cm^{-2}s^{-1}$, much lower than the maximum design luminosity of the LHC, and with a bunch crossing rate of 40 MHz. During the 2012 data taking the storage system was able to write and save events at a rate of 5 kHz, corresponding to 0.3 GB/s. The rate reduction from 40 MHz to 5 kHz is performed by a trigger system, designed to select only a small fraction of potentially interesting events.

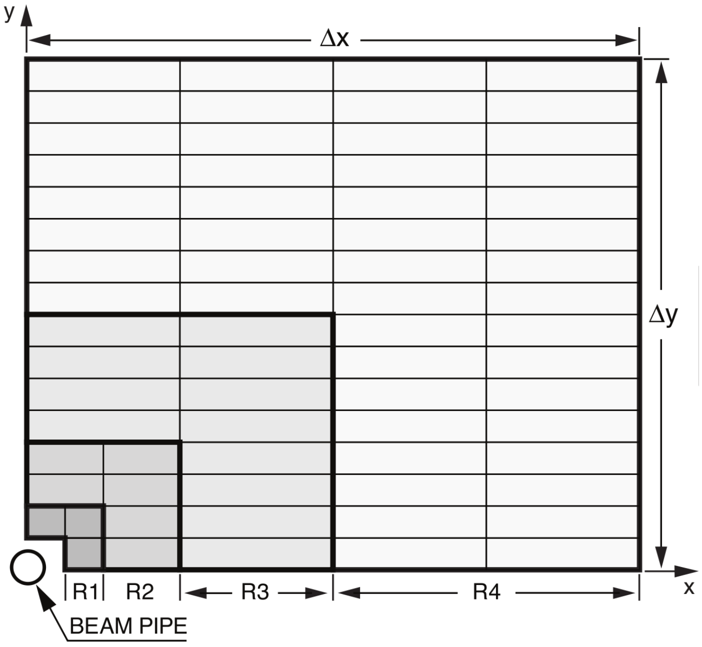

Figure 2.17: Front view of a quadrant of a muon station. Each rectangle represents a muon chamber.

The trigger makes its decision in two stages: a hardware trigger stage, called Level 0 (L0), which is synchronous with the bunch crossing frequency, and a software trigger stage called High Level Trigger (HLT) which is executed asynchronously on a processors farm. The 2012 trigger schema is presented in figure 2.18. The trigger stages are described in the next Sections.

2.11.1 Level 0 trigger

Events with b hadron decays are characterized by particles with high transverse momentum (p_T) or high transverse energy (E_T). The Level 0 trigger uses the calorimeters and the muon system detector to spot these signatures. The calorimeters and the muon system are connected to the Level 0 Decision Unit (L0-DU) that decides if the event is accepted, reducing the rate from 40 MHz to 1 MHz. To reject events with multiple interactions, which need long time to be processed, L0 applies a cut on the number of the hits in the SPD detector (Global Event Cut, GEC). Then the event is accepted if one of these conditions is satisfied:

- *L0-Muon*: a track identified as a muon with a p_T above a given threshold is found;
- *L0-Dimuon*: two track identified as muons with a $p_T^1 \cdot p_T^2$ above a given threshold are found;

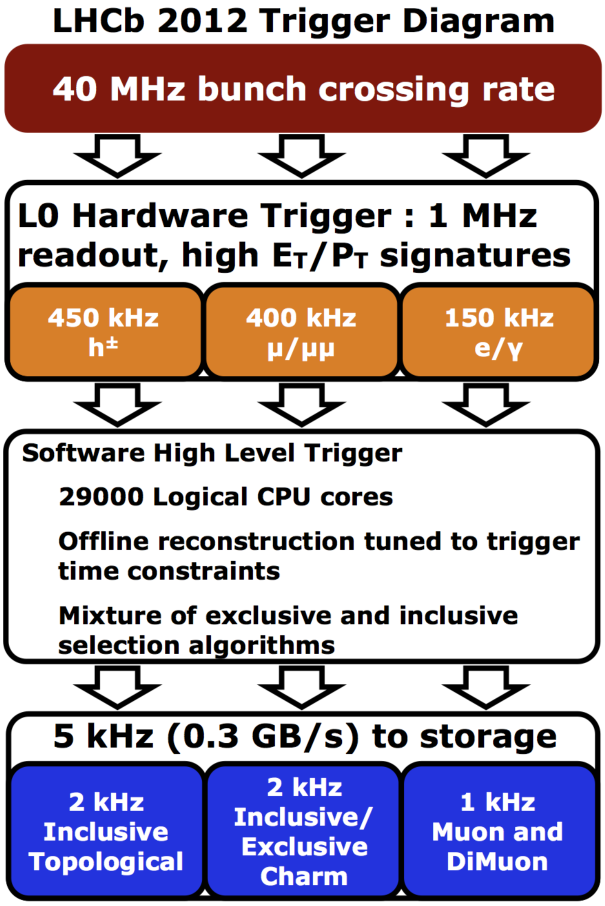

Figure 2.18: Schema of the LHCb trigger.

- *L0-Hadron*: a HCAL cluster with a E_T above a given threshold is found;
- *L0-Photon*: a ECAL cluster with a E_T above a given threshold is found;
- *L0-Electron*: a ECAL cluster with a E_T above a given threshold anticipated by hits in the PS and at least 1 hit in the SPD is found.

The L0 thresholds have been decided to maximize the trigger efficiencies of benchmark decays. The trigger thresholds used in the 2012 data taking are reported in table 2.2.

2.11.2 High Level Trigger

Events accepted by the L0 trigger are processed by the software High Level Trigger which is divided in two stages, the High Level Trigger 1 (HLT1) and the High level Trigger 2 (HLT2). It runs on a large computing cluster called Event Filter Farm and it reduces the event rate from 1 MHz to 5 kHz. The HLT1 performs a partial reconstruction of the L0

Table 2.2: Level 0 trigger thresholds used during the 2012 data taking.

line	p_T [GeV/c]	$p_T^1 \cdot p_T^2$ [GeV/c^2]	E_T [GeV]	$nSPD$
L0-Muon	> 1.76	-	-	< 600
L0-DiMuon	-	$> (1.6)^2$	-	< 900
L0-Hadron	-	-	> 3.7	< 600
L0-Photon	-	-	> 3	< 600
L0-Electron	-	-	> 3	< 600

candidates using information from the VELO and from the T1-T3 stations. Requirements are applied depending on the L0 decision. As general criteria tracks with high momentum and high displacement from the interaction point are selected.

The HLT2 performs a full event reconstruction and selects events using inclusive and exclusive algorithms. Examples of HLT2 algorithms are:

- inclusive topological algorithms: they reconstruct b hadrons decays looking at 2-,3- or 4-body vertices and applying a selection on those;
- exclusive/inclusive charm algorithms: they reconstruct and select D mesons from specific decay modes (exclusive algorithms) or from the combinations of all charged hadrons in the event (inclusive algorithms);
- muon and dimuon algorithms: they select high p_T muons or dimuons with requirements depending on the physics goals.

2.12 LHCb softwares

The LHCb experiment uses several software packages to reconstruct events, apply the software trigger, analyse data and generate simulated events. The softwares are based on ROOT [53] and on the Gaudi framework [54]. The main packages are listed below.

- GAUSS [55]: it is used to produce Monte Carlo simulations. GAUSS makes use of different generator softwares like PYTHIA [56], POWHEG [57] or ALPGEN [58] to generate pp interactions and uses EvtGen [59] to simulate B mesons decay; the interaction of particles with the detector and beam pipe materials is simulated with GEANT4 [60, 61].
- BOOLE [62]: it manages the last part of the simulation providing sub-detectors responses using GEANT4. It also reproduces the digitization of the front end electronics. At this stage all the instrumental effects, like the readout electronic noise and inefficiencies, are included.

- MOORE [62]: in the online system it applies the HLT algorithms to events accepted by the L0 hardware trigger. It is also used offline to simulate the software trigger stage.
- BRUNEL [62]: it performes the full event reconstruction starting from raw data, e.g. using track reconstruction and particle identification algorithms. It is used for both simulated and real data.
- DaVinci [62]: it reconstructs the decay chains and the jets combining together different particles and applying requirements on the observables reconstructed by BRUNEL. It is used for both simulated and real data.

Chapter 3

Jets reconstruction and flavour tagging

3.1 Introduction

LHCb is a detector initially designed for B hadron physics, but thanks to its unique phase space coverage and to the large amount of collected data, it is worthwhile to analyse processes with jet signature. In Quantum Chromodynamics (QCD) jets are produced via the fragmentation of quarks and gluons. In detectors they manifest themselves as collimated spray of charged particles in trackers or as concentrated clusters of energy in calorimeters. The LHCb environment is characterized by highly boosted jets and, in case of events with multiple jets, resolve and separate one from each other may be difficult.

In this Chapter the jet reconstruction algorithm used and tuned by LHCb during the Run I data taking is explained and discussed in details. It has a crucial role in the analyses and measurements presented in this thesis, and it is important to understand its properties and limitations, as well as the systematics uncertainties that can arise from the reconstruction procedure. A discussion on the jets heavy flavour identification algorithms used at LHCb is also reported, together with the evaluation of their performance, obtained using data driven techniques.

3.2 Jets reconstruction algorithm

3.2.1 Overview

The LHCb electromagnetic calorimeter described in Section 2.8 has an energy resolution of $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$ (where E is the energy measured in GeV) and it is designed for the measurements of photons and π^0. The hadronic calorimeter has an energy resolution of $\frac{\sigma_E}{E} = \frac{69\pm5\%}{\sqrt{E}} \oplus 9 \pm 2\%$. It is mainly used for triggering on high p_T hadrons and it does not allow precise measurements of hadrons energy. On the other hand LHCb has an excellent tracking system, with an efficiency of about 97% on charged tracks with a momentum greater than 5 GeV/c and a resolution of about 0.5%. For these reasons it is necessary to involve both the tracking system and the calorimeters in the jet reconstruction, resulting in a hybrid algorithm significantly different from the fully calorimetric reconstruction used by other experiments. The tracking system is used to select charged particles while the calorimeters select mainly the neutral particles coming from jets: this is coherent with the LHCb event reconstruction model used for B hadron physics. The jet reconstruction algorithm consists of several stages:

1. Particle Flow: tracks from the tracking system and calorimeter clusters are selected as input particles if they satisfy specific criteria;
2. anti-k_t: particles selected by the Particle Flow are clustered into jets;
3. E-recombination scheme: jet four-momentum is computed as the sum of the jet constituents four-momenta;
4. Jet Energy Correction: the jet four-momentum is multiplied by a correction factor that depends on the jet kinematic.

In the next section each step of the jet reconstruction algorithm is described in details.

3.2.2 Particle Flow

The Particle Flow algorithm applies selection criteria to tracks and calorimeters clusters before they enter as inputs the jet clustering algorithm. The Particle Flow steps are the following:

1. requirements are applied to tracks, which are selected as "charged particles";
2. requirements are applied to calorimeter clusters not associated to tracks. The clusters selected in this way are called "isolated neutral particles";
3. requirements are applied to calorimeter clusters associated to at least one track. The estimated energy released by the tracks in the calorimeters is subtracted from the associated cluster energy. The resulting object is called "not-isolated neutral particle".

In the following the treatment of charged particles, isolated neutral particles and not-isolated neutral particles is explained.

Charged particles

Different selection requirements are applied to tracks depending on the category they belong (*long*, *downstream*, *upstream*, VELO tracks, see Section 2.3). This is done to ensure the quality of those entering the jet clustering step. The tracks selection is performed applying requirements to the following observables:

- the track transverse momentum (p_T);
- the χ^2 obtained from the track Kalman fit;
- P_{ghost}, defined as the probability for a track to be an artifact of the pattern recognition, not associated to a real particle. It is computed as an output of a neural network [63] that has in inputs the result of the track fit and the track kinematic;
- The momentum resolution, $\sigma(q/p)/(q/p)$, where q is the particle charge and p the particle momentum.

The list of tracks requirements is displayed in table 3.2 for the four track categories.

Table 3.1: List of requirements applied to tracks by the Particle Flow algorithm. The requirements depend on the track category.

	long	*downstream*	*upstream*	VELO
p_T [MeV/c]	-	-	> 100	-
χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	< 0.5	-	-	-
$\sigma(q/p)(q/p)$	> 10	> 10	> 2	-

VELO tracks do not contain information on the particle momentum, but they provide information on the the Primary Vertex associated to the jet, therefore they are included in the jet clustering. A particle type and mass is also assigned to each track depending on the particle identification decision. The particle type categories are $p/\bar{p}$, $\pi^{\pm}$, $\mu^{\pm}$, $e^{\pm}$ and $K^{\pm}$.

Isolated neutral particles

In this step calorimeters clusters not associated to tracks are stored as inputs for the jets clustering.

Neutral particles detected by the ECAL are mainly photons and π^0 decayed into two photons. Their identification relies on the shape of the calorimeter clusters, therefore a likelihood for the photon or π^0 hypotheses is computed. The π^0 are divided into two categories depending on how they are detected: *merged* π^0, when the two photons are almost collinear and produce a single cluster, and *resolved* π^0, where the two photons are detected as two separate clusters. Requirements are applied to select ECAL clusters isolated from the tracks and to ensure their quality before entering the jet clustering step. The following observables are considered:

- the likelihood for the photon hypothesis (PhotonID);
- the cluster transverse energy E_T;

- a χ^2 for each track-cluster combination ($\chi^2_{track-cluster}$): it evaluates how much the cluster is likely to be originated by the particle associated to the track.

The list of the requirements applied in the ECAL clusters selection is reported in table 3.2 It can be noticed that different conditions are applied to photon clusters associated to T tracks, which are defined as tracks that have hits only in T1-T2-T3 stations.

Table 3.2: List of requirements applied to ECAL clusters identified as photons or π^0.

	merged π^0	resolved π^0	photons
E_T [MeV/c]	-	-	> 200
PhotonID	-	> -4	> -1 (> -2 with T track)
PhotonID for 1 γ	-	> -2	-
$\chi^2_{track-cluster}$	> 25	> 25	> 25 (> 16 with T track)

The selection of isolated HCAL clusters is performed by applying different $\chi^2_{track-cluster}$ requirements for different cluster energy thresholds: a $\chi^2_{track-cluster}$ greater than 25 is required for HCAL clusters with energies below 10 GeV while a $\chi^2_{track-cluster}$ greater than 15 is required for energies above 10 GeV. No particle identification requirements are applied on the HCAL clusters selection.

Non-isolated neutral particles

The selection of calorimeter clusters not isolated from tracks works as follow:

1. ECAL clusters with $\chi^2_{track-cluster}$ below 25 and HCAL clusters with $\chi^2_{track-cluster}$ below 25(16) for energies below(above) 10 GeV are selected;
2. ECAL and HCAL clusters are grouped in the way that clusters in different groups do not share the same associated tracks;
3. the expected energy released in the calorimeters by charged particles pointing to a clusters group is evaluated using an empirical parametrization of E/p as a function of p, where E is the cluster energy and p the track momentum;
4. if the total expected energy of the cluster group (E^{exp}) is larger than 1.8 times the measured energy (E^m) then the clusters group is discarded;
5. if $E^m > 1.8\ E^{exp}$ then E^{exp} is subtracted from E^m;
6. the remaining energy is selected as non-isolated neutral particle if its E_T is greater than 2 GeV and it is used as input in the jet clustering.

The parametrization of E/p as a function of p is obtained through a procedure called E/p calibration. The calibration has been performed on a data sample of pp collisions at 7 TeV, collected using a minimum bias trigger configuration. From this sample isolated tracks matched to a calorimeter cluster have been selected. Requirements are applied to remove the background from minimum ionizing particles, like muons, that release a small amount of their energy in calorimeters:

- only one PV in the event;
- *long* or *downstream* track;
- No other tracks within $\Delta R < 0.5$ from the selected track, with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where η is the pseudorapidity and ϕ the azimuthal angle;
- track p_T greater than 50 MeV/c;
- $\chi^2_{track-cluster} < 1$ for one ECAL or HCAL cluster;
- no other track with $\chi^2_{track-cluster} < 100$ associated to the same cluster;
- cluster transverse energy greater than 200 MeV.

Track-cluster objects are divided in different categories with the following features (track-cluster objects can belong to one or more of these categories):

- the track is associated to a cluster in ECAL but not in HCAL;
- the track is associated to a cluster in HCAL but not in ECAL;
- the track is associated to a cluster in ECAL, independently of the presence of a HCAL cluster;
- the track is associated to a cluster in HCAL, independently of the presence of a ECAL cluster;
- the track is associated to a cluster in HCAL;
- both a ECAL and a HCAL cluster are associated to the tracks.

A separate E/p calibration is performed for each of these categories. Moreover the calibration is done separately for tracks identified as hadrons or electrons. In the calibration procedure the mean value of E/p is computed in different intervals of p as the mode of the E/p distribution. Then the E/p dependence from p is fitted with empirical functions. For the first five categories listed above, where only one cluster from ECAL or HCAL is considered, the function used is:

$$\frac{E}{p}(p) = a_1 e^{-a_2 p} + a_3.$$

For the last category E is computed as the sum of ECAL and HCAL energies and the function used is:

$$\frac{E}{p}(p) = (a_1 + a_2 p + a_3 p^2)e^{-a_4 p} + a_5.$$

Few examples of this parametrization are reported in figure 3.1, compared with the E/p vs p obtained using a Monte Carlo sample of pp collision at 7 TeV with a minimum bias trigger configuration.

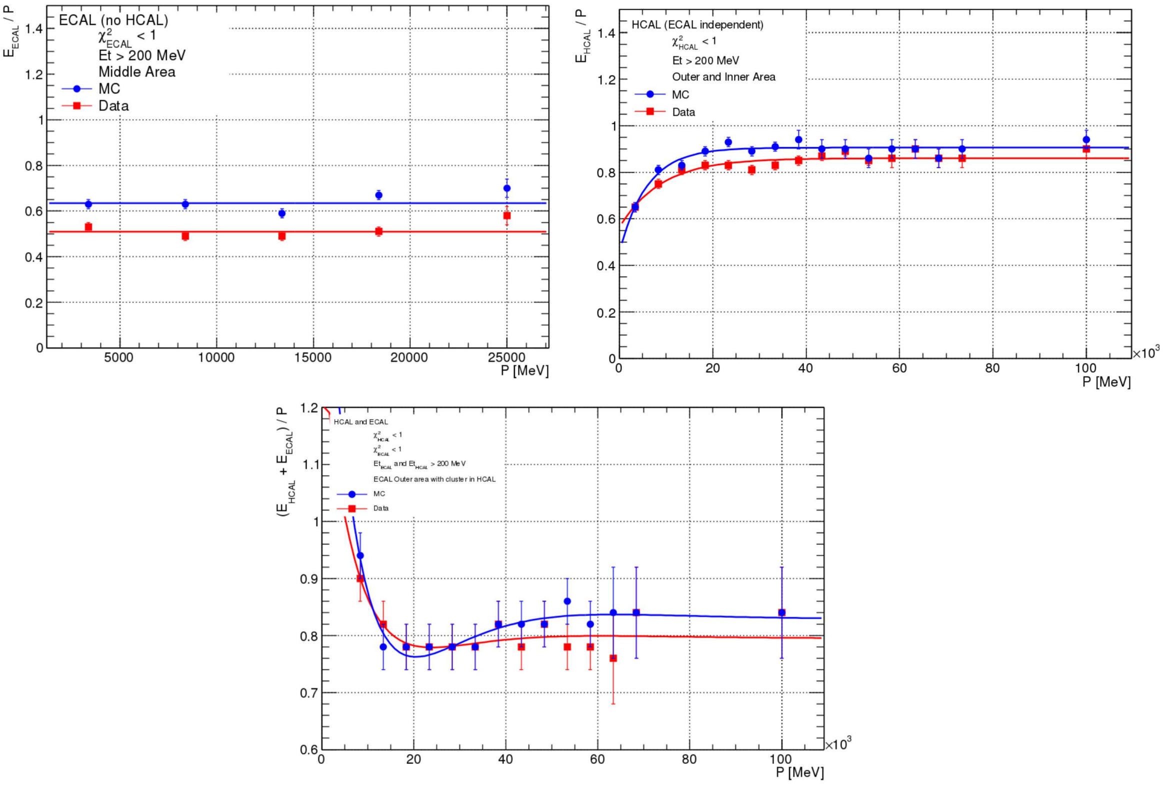

Figure 3.1: E/p calibration for hadrons. The plots show the calibration for different categories that are, from left to right and top to bottom: the category where a ECAL cluster is associated to a track but bot a HCAL cluster; the category where a HCAL cluster is associated to a track, independently from ECAL; the category where both a ECAL and a HCAL clusters are associated to a track.

3.2.3 Jet clustering with anti-k_t

Charged particles, isolated and not-isolated neutral particles selected by the Particle Flow are used as inputs in the jet clustering algorithm. First tracks are associated to primary vertices (PVs) as explained in Section 2.6. The clustering algorithm is applied to all the selected tracks associated to the same PV and to all the selected calorimeters clusters.

If more than PV is found in the event a clustering procedure for each PV is performed. Since calorimeter clusters are not associated to a PV they are used as inputs in all the procedures.

Several jet finding algorithm exist: the one used by LHCb is the "anti-k_t" algorithm [64]. The anti-k_t algorithm has some notable features that are not provided by many of the most used jet algorithms (e.g. cone algorithms). The most important are the infrared and collinear safeties. A jet algorithm is infrared safe if the soft gluon emissions does not change the jet definition while it is collinear safe if the jet definition is insensitive to the collinear gluon emissions. This features are fundamental because the gluon emission probability gives infinity if integrated in the soft and collinear phase space, leading to an "unstable" jet definition. Moreover, the anti-k_t algorithm is different from all the other known infrared and collinear safe algorithms for not producing irregularities due to soft emissions in the boundaries of final jets.

The anti-k_t algorithm works as follow:

1. a list with all input particles is created;
2. the distance d_{ij} is computed for each combination of two particles i and j, defined as

   $$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2},$$

   where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, the rapidity and the azimuthal angle of particle i. R is the radius parameter that has to be chosen, it will be discussed later. In the anti-k_t p is equal to -1. For $p = 1$ the k_t algorithm is obtained while for $p = 0$ the Cambridge/Aachen algorithm is obtained.
3. the distance between each particle i and the beam is calculated with

   $$d_{iB} = k_{ti}^{2p};$$
4. the smallest distance between the d_{iB} and d_{ij} is found;
5. if the smallest distance is d_{ij} then the two particles i and j are combined in a single particle by summing their four-momenta. The two particles are substituted with the combined particle in the list and the algorithm re-starts from step 2;
6. if the smallest distance is d_{iB} then the particle i is removed from the list and it is called jet;
7. if no more particles are in the list the algorithm ends, otherwise it restarts from step 2.

One consequence of the anti-k_t distance definition is that soft particles will tend to cluster with a hard particle before clustering among themselves. If a hard particles has no hard neighbours within a $2R$ distance in the (y,ϕ) plane it will accumulate all soft particles within a circle of radius R resulting in a conical jet. If a second hard particle is present in the $[R,2R]$ distance range, then two jets are formed, and is it not possible for them to be perfectly conical.

The choice of the parameter R is crucial in the jet reconstruction. It should be large enough to catch all the QCD perturbative radiation, but not too large to avoid an excessive contamination from the underlying event. Its choice should be dependent from the experimental environment, from the detector properties and from the specific channels under study. At LHCb studies has been performed to determine the best R, trying to maximize figures of merits, like the jet energy resolution or the jet identification efficiency. The optimal radius have been found to be between $R = 0.5$ and $R = 0.7$. In figure 3.2 the invariant mass distribution of reconstructed b-jets from the Z decay, obtained using a Monte Carlo sample at a centre-of-mass energy of 8 TeV, for $R = 0.5$ jets and $R = 0.7$ jets is shown. It can be seen that in the $R = 0.7$ case the mass is closest to the nominal Z mass but the peak resolution is similar with both radius choices. As final choice for this thesis $R = 0.5$ is used.

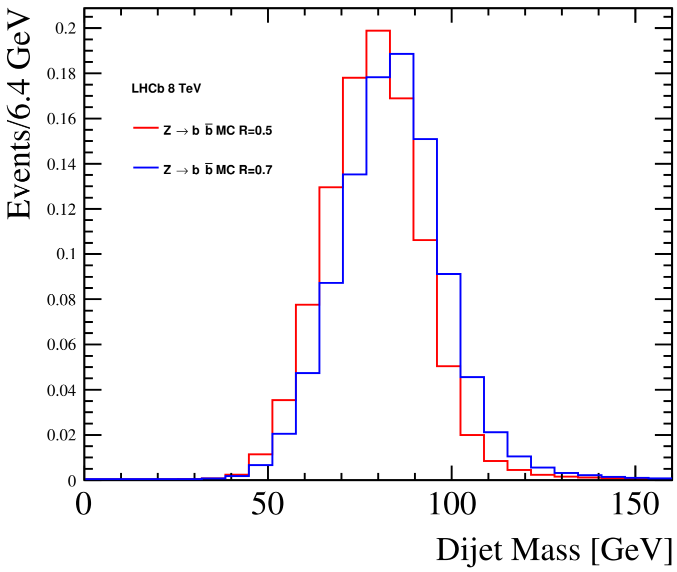

Figure 3.2: Invariant mass distribution of reconstructed b-jets from the Z decay, obtained using a Monte Carlo sample at a centre-of-mass energy of 8 TeV, for $R = 0.5$ jets and $R = 0.7$ jets.

3.2.4 Jet four-momentum and MC validation

After the clustering step, the jet four-momentum (p_{jet}) is determined using the "E-recombination scheme", as the sum of the constituents particles momenta $p_i = (E_i, \vec{p}_i)$:

$$p_{jet} = (E_{jet}, \vec{p}_{jet})$$

where

$$E_{jet} = \sum_i E_i$$

and

$$\vec{p}_{jet} = \sum_i \vec{p}_i.$$

At this stage jets reconstruction can be validated using Monte Carlo (MC) samples. In simulated events MC jets (jet$_{MC}$) are defined as jets clustered by anti-k_t but using as inputs all the stable MC particles ($\tau > 10^{-8}$ s) instead of the reconstructed particles. MC particles are characterized by having the true kinematical values, before passing through the detector resolution. Invisible particles, like neutrinos, are removed from the list, in order to not bias the evaluation of the reconstructible energy. The association between a reconstructed jet (jet$_{reco}$) and a jet$_{MC}$ is done by selecting the jet$_{MC}$ with a distance from the jet$_{reco}$ in the (η, ϕ) plane below 0.4 ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$). If more then one jet$_{MC}$ is found in this way, then the closest in distance is selected. The particle composition of the jet$_{MC}$ according to the MC truth is reported in figure 3.3 as a function of jet p_T and η, obtained from a minimum bias simulated sample at centre-of-mass-energy of 7 TeV. It can be seen that around 60% of jets constituents are charged particles, detected by the tracking system, around 30% are π^0 or photons detected by the ECAL and 10% are neutral hadrons detected by HCAL. The ratio of the energy of jet$_{reco}$ to the energy of jet$_{MC}$, obtained from the minimum bias MC, is shown separately for charged particles (figure 3.4), for neutral particles (figure 3.5) and for isolated neutral particles (figure 3.6) as a function of the jet$_{MC}$ p_T, η and ϕ. Figures 3.5 and 3.6 show that, without including the not-isolated neutral particles, a certain percentage of energy is lost.

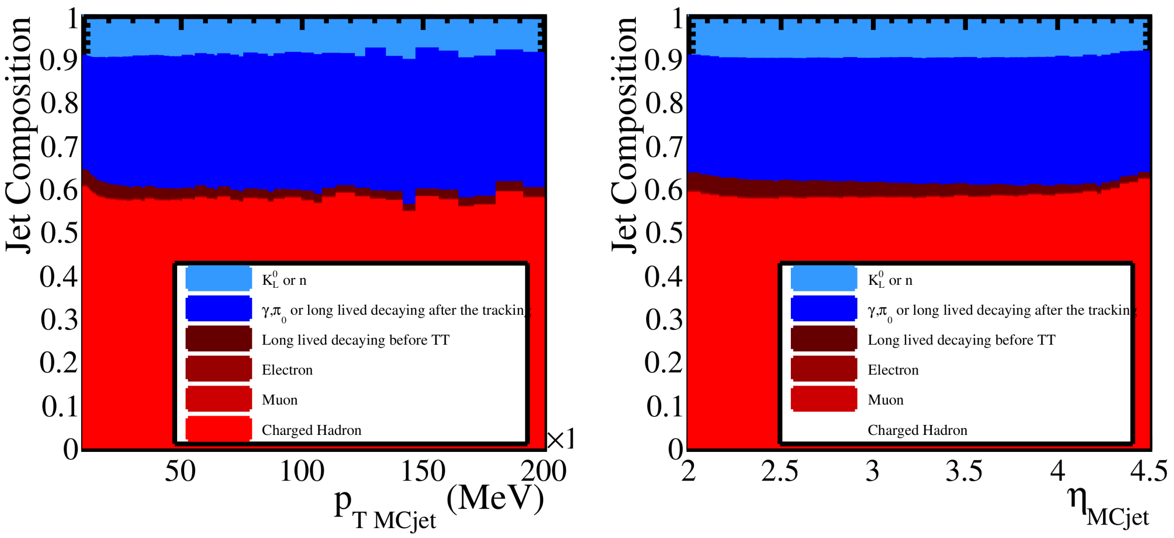

Figure 3.3: Composition of MC jets according to the MC truth as a function of jet p_T and η.

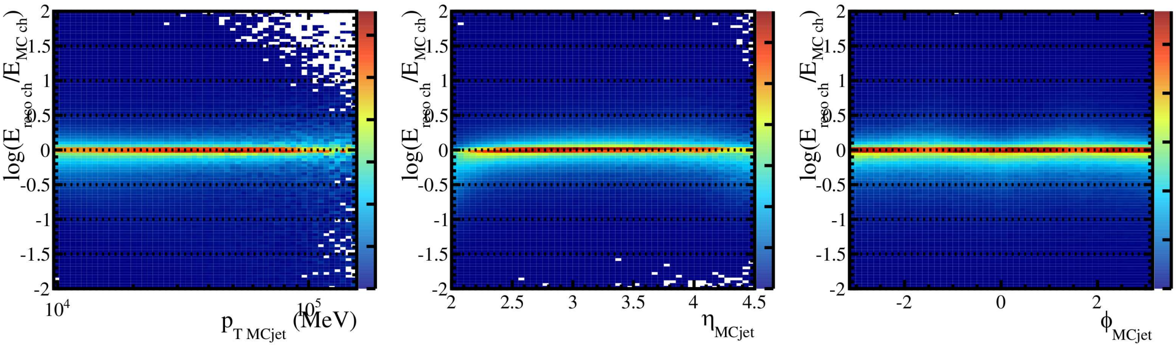

Figure 3.4: Jet/MC jet ratio of the energy from charged particle, as a function of p_T, η and ϕ of the MC jet.

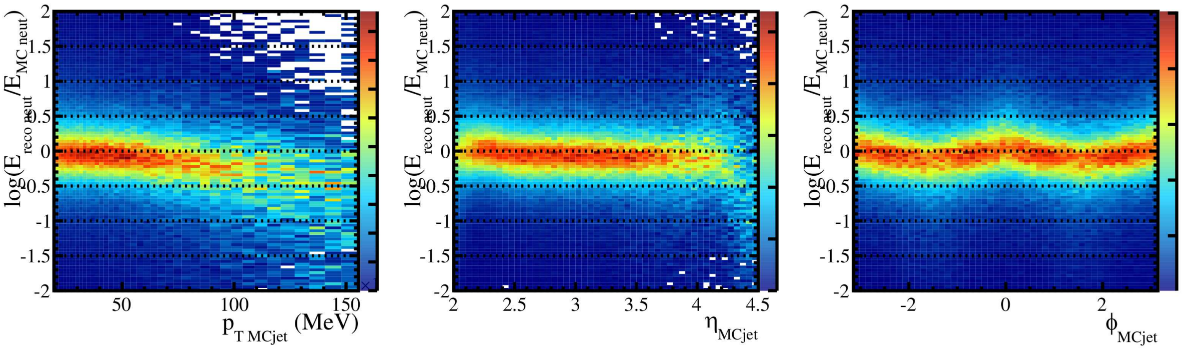

Figure 3.5: Jet/MC jet ratio of the energy from neutral particles, as a function of p_T, η and ϕ of the MC jet.

3.2.5 Jet Energy Correction

In MC events the energy of reconstructed jets, $E(\text{jet}_{reco})$, differs from the energy of associated MC jets, $E(\text{jet}_{MC})$. To correct the jet energy to its true value, a multiplicative correction factor k_{MC} is evaluated in simulation:

$$E(jet_{MC}) = k_{MC}E(jet_{reco}).$$

The jet_{MC} direction is found to be the same as the jet_{reco}, therefore the same correction factor is applied to each component of the jet four-momentum vector. k_{MC} takes into account the effects of the pile-up, the noise and the non-uniformity of the detector.

k_{MC} is evaluated in simulated events of b, c, light quarks and gluons jets at a centre-of-mass energy of 7 TeV and 8 TeV. It is found to be non-uniform with respect to the jets η, ϕ and the fraction of charged particles in the jet (cpf). A parametrization of k_{MC} as a function of p_T has been obtained with fits in different regions of jet η, ϕ, cpf and number

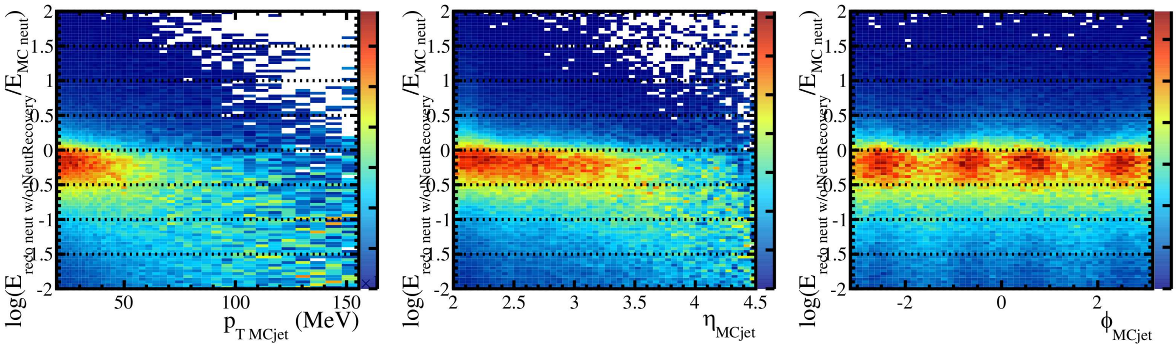

Figure 3.6: Jet/MC jet ratio of the energy from isolated neutral particles, as a function of p_T, η and ϕ of the MC jet.

of PV (nPV), using a cubic function model. The k_{MC} values used in the fit have been calculated as the median of k_{MC} distributions in p_T intervals.

k_{MC} may depend on the jet flavour. The relative systematic uncertainty on the flavour dependence is calculated as $max\left(\frac{k^i_{MC}-k^j_{MC}}{k^i_{MC}+k^j_{MC}}\right)$, where k^i_{MC} is the correction calculated using jets of flavour i. The uncertainty obtained is 2.7% (2.6%) for jets with $R = 0.5$ ($R = 0.7$). Other systematic uncertainties are due to fake tracks (1.2% probability per track to be a fake), to the track p_T resolution (of about 1 % per track) and to the calorimeter clusters energy resolution.

Figure 3.7 shows the mean k_{MC} in different intervals of η, ϕ, cpf and nPV.

An additional correction should be applied to take account of possible differences between the jet energy in MC and in real data. This correction factor is called Jet Energy Scale and it is measured in this thesis in Chapter 4.

3.2.6 Jet identification efficiencies

Requirements are applied to reject jets originated from noise and high energy isolated leptons:

- number of tracks pointing to the PV (nPVtrks) ≥ 2;
- maximum fraction of transverse momentum carried by a single PF particle (mpf) < 0.8;
- maximum transverse momentum carried by a track (mpt) $> 1.2\,\text{GeV}$;
- fraction of charged particles on the jet (cpf) > 0.1.

The jet identification efficiency is defined as the number of reconstructed jet to the number of true jets. It has been measured using $Z \rightarrow \mu\mu$+jet MC events at a centre-of-mass energy of 7 TeV [65]. To obtain the real jet efficiency, MC events are weighted

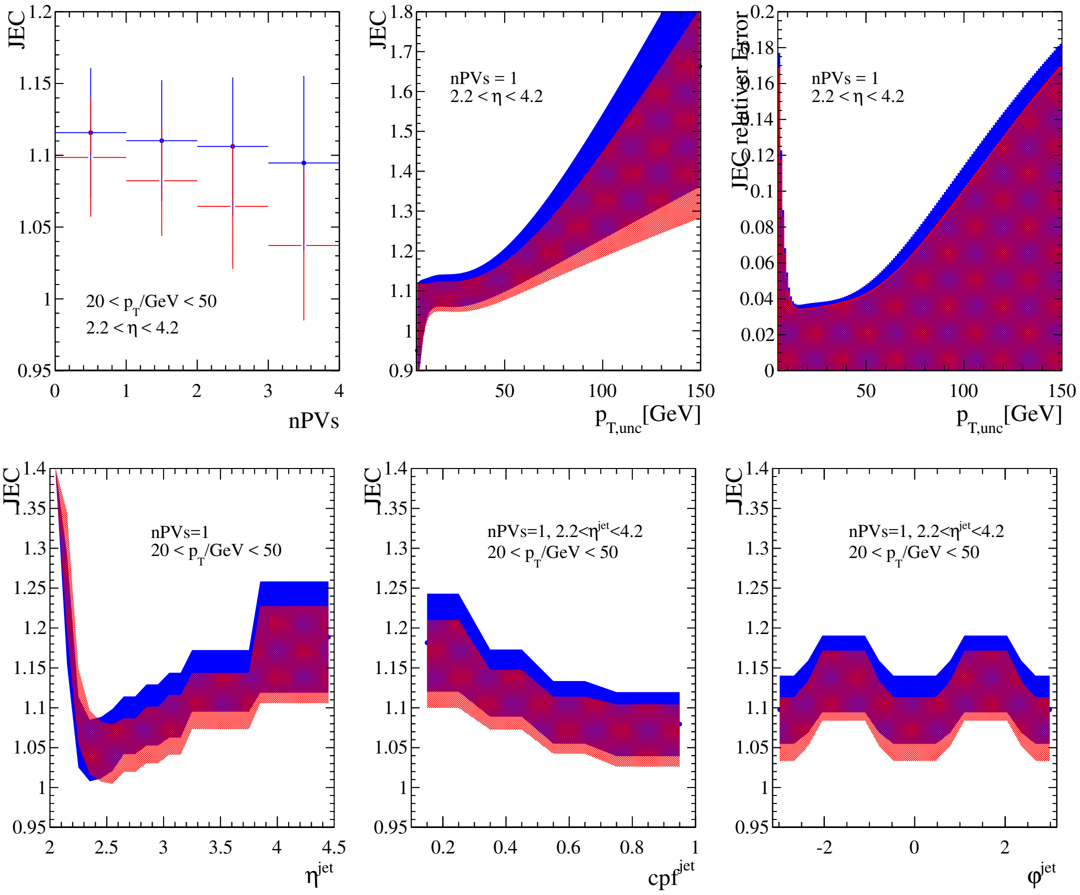

Figure 3.7: k_{MC} in different intervals of η, ϕ, cpf and nPV for $R = 0.5$ and Right $R = 0.7$.

such that the reconstructed jet multiplicity distribution in $Z \to \mu\mu$+jet MC matches the distribution in data. The weights are determined as a function of the jet p_T and their uncertainty depends on the statistics of the $Z \to \mu\mu$+jet data sample. Figure 3.8 shows the jet identification efficiency as a function of the p_T of the jet.

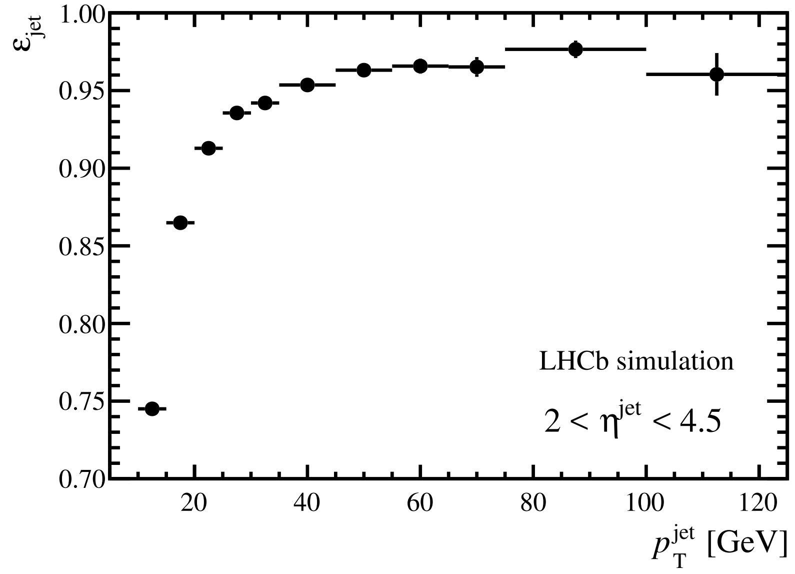

Figure 3.8: Jet identification efficiency as a function of the p_T of the jet.

3.3 b tagging algorithms

In the study of processes with heavy quarks in the final state (b or c) it is fundamental to identify the flavour of the quark that generates a jet. This is achieved by using heavy flavour tagging algorithms. The algorithms analyse the jet properties to take a decision on the jet flavour. There are different methods for flavour identification:

- Secondary Vertex tagging: D and B hadrons fly for a distance d before decaying, with $<d>=\beta\gamma c\tau$, where τ is their mean lifetime, c is the speed of light, $\beta=\frac{v}{c}$, v is the particle velocity and $\gamma=\frac{1}{\sqrt{1-\beta^2}}$. Therefore a Secondary Vertex (SV) detached from the Primary Vertex is present in B and D events, formed by meson decay products. If a SV is found inside a jet then it can be tagged as generated from a heavy flavour quark. SV related observables can be studied to separate b jets from c jets;

- jet-charge tagging: an effective jet charge can be defined as the weighted sum of the charges of jet components:
$$Q_{jet}=\frac{\sum_i p_{Ti}^{\alpha}\cdot q_i}{\sum_i p_{Ti}^{\alpha}},$$
where p_{Ti} and q_i are respectively the transverse momentum and the charge of particles in the jet, α is a parameter to be tuned. Q_{jet} can be used as estimator of the generating quark charge;

- lepton-charge tagging: D and B hadrons may decay semileptonically, generating high energy leptons. Requirements may be applied to search for this lepton inside a jet: if such lepton is found then the jet is tagged as generated from a heavy flavour quark.

The tagging algorithms used in this thesis are described in the following sections. The first one, called “SV tagging”, is used in the offline event reconstruction, it separate light partons from heavy quarks jets and b-jets from c-jets. The second algorithm, called “Topological tagging”, is used in the online event selection, and can be required at the HLT2 trigger level. The performance of the two algorithms have been measured using data driven techniques and they are compared with those obtained from the MC simulation.

3.3.1 SV tagging algorithm

The SV tagging algorithm [66] works as described next.

1. Tracks are selected with the following requirements:
    - *long* tracks;
    - $p>5$ GeV/c and $p_T>0.5$ GeV/c;
    - the χ^2/ndof associated to the track fit must be less than 3;

  - χ^2_{IP}, defined as the variation of the χ^2 obtained from the PV fit when the track is removed from the fit result, must be greater than 16;
  - $P_{ghost} < 0.3$.

2. Selected tracks are used to build all possible 2-body SVs in the 3-dimensional space. Fits are performed to determine the SV position. The two tracks associated to a SV are combined to form a particle which flight direction is defined as the vector that points the SV from the PV. Its four-momentum is defined as the sum of tracks four-momenta, assuming the π mass.

3. The 2-body particles have to fulfill the following requirements:
   - the distance of closest approach (DOCA) between the two tracks must be less than 0.2 mm;
   - the χ^2 associated to the SV fit must be less then 10;
   - the invariant mass must be greater than 400 MeV/c^2 and less then the B^0 mass (5279.4 MeV/c^2).

   The lower mass requirement removes strange-hadrons decays, while the upper mass requirement rarely remove B vertices since the mass of π is assigned to all tracks.

4. A 2-body particle is considered inside a jet if $\Delta R < 0.5$, with $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ where $\Delta\eta = \eta_{jet} - \eta_{SV}$ and $\Delta\phi = \phi_{jet} - \phi_{SV}$. At this stage the 2-body particles in the same jet are associated with all other 2-body particles that share at least one track. In this way n-body particles that do not have tracks in common are created. The procedure is applied for every jet in the event, using all tracks with $\Delta R < 0.5$ from the jet axis, even if these have been included in a SV associated to a different jet. The resulting n-body particles are called tag$_{SV}$. The tag$_{SV}$ position is the weighted average of the 2-body SV positions, where the weights are the χ^2 from the 2-body vertex fit. The tag$_{SV}$ flight direction is the vector that points its position from the PV and the four-momentum is the sum of tracks four-momenta, assuming the π mass.

5. To further remove the light jet contamination, selection criteria are applied to tag$_{SV}$:
   - $p_T > 2$ GeV/c;
   - z-position $< 200\ mm$;
   - flight distance/$p < 1.5\ mm$/(GeV/c);
   - the flight distance χ^2, defined as the χ^2 obtained from the PV fit if the tag$_{SV}$ tracks are added to the fit result, must be above 5σ;
   - if tag$_{SV}$ is formed by only two tracks and a mass compatible with the K_S it is rejected;
   - the tag$_{SV}$ must have at most one track with $\Delta R > 0.5$ from the jet axis;

- the tag$_{SV}$ corrected mass is defined as

  $$M_{cor} = \sqrt{M^2 + p^2 \sin^2\theta} + p\sin\theta,$$

  where M and p are respectively the invariant mass and the momentum of the tag$_{SV}$, θ is the angle between the tag$_{SV}$ momentum and flight direction. M_{cor} must be greater than 600 MeV/c^2 to remove any remaining kaon or hyperon. The formula above is obtained by correcting the invariant mass for the missing momentum:

  $$M_{cor} = \sqrt{M^2 + |p_T^{miss}|^2} + |p_T^{miss}|,$$

  where p_T^{miss} is the transverse component of the missing momentum with respect to the particle flight direction, therefore $|p_T^{miss}| = p\sin\theta$.

6. If more than one tag$_{SV}$ inside a jet satisfies the requirements, then the one with greater p_T is chosen. A jet is identified to be generated by a heavy flavour quark ("SV tagged") if at least one tag$_{SV}$ is found with this procedure.

To further remove light jet contamination and to distinguish b jets from c jets a Multivariate Algorithms (MVA) is used. Two boosted decision trees (BDTs) [67–69] are employed: one for the heavy/light jets separation (BDT$_{bc|udsg}$) and the other for the b/c jets separation (BDT$_{b|c}$). Simulated samples of heavy quark/light quark jets and b/c jets generated with Pythia 8 are used as signal/background samples for the BDTs training. The observables in input to the BDTs are those related to tag_{SV} that provide the highest discrimination power between the different flavours. Their distributions are shown in figure 3.9 and they are:

- the tag$_{SV}$ mass M;
- the tag$_{SV}$ corrected mass M_{cor};
- the transverse flight distance of the 2-body particle closest to the PV within those that form tag$_{SV}$;
- the fraction of the jet p_T carried by tag$_{SV}$, $p_T(SV)/p_T(jet)$;
- the number of tracks that form tag$_{SV}$;
- the number of tracks that form tag$_{SV}$ with $\Delta R < 0.5$ from the jet axis;
- the total charge of tracks in tag$_{SV}$;
- the tag$_{SV}$ flight distance χ^2;
- the sum of χ^2_{IP} for all the tracks in tag$_{SV}$.

The two-dimensional BDT outputs distribution are shown in figure 3.10 for the b, c and light quarks MC samples generated with Pythia 8 for 8 TeV collisions. Cuts on these two observables can be applied to enrich the jets samples of a specific hadron flavour.

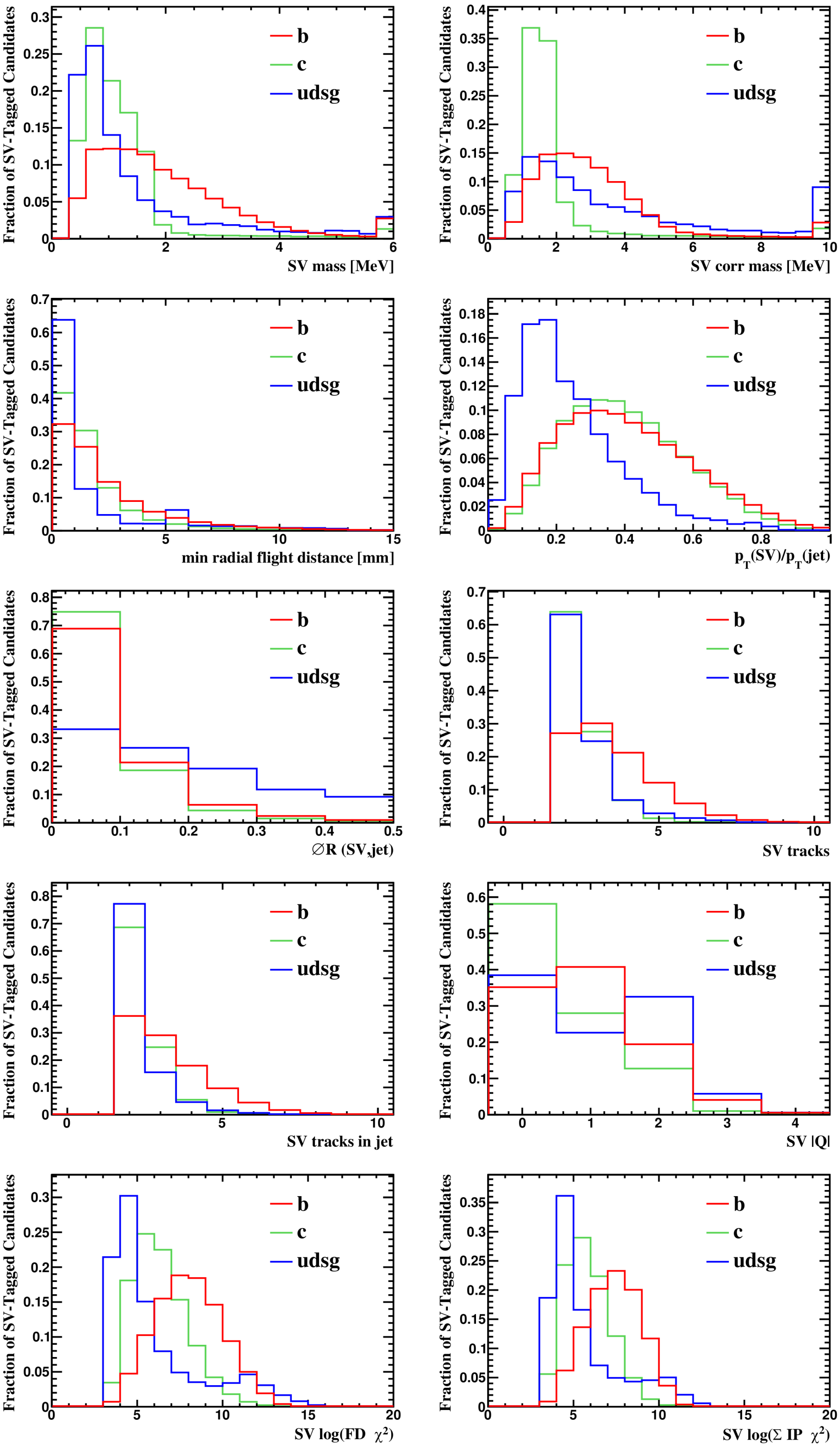

Figure 3.9: Observables, related to tag$_{SV}$, in input to the SV algorithm BDTs. They provide discrimination between the different flavours [66].

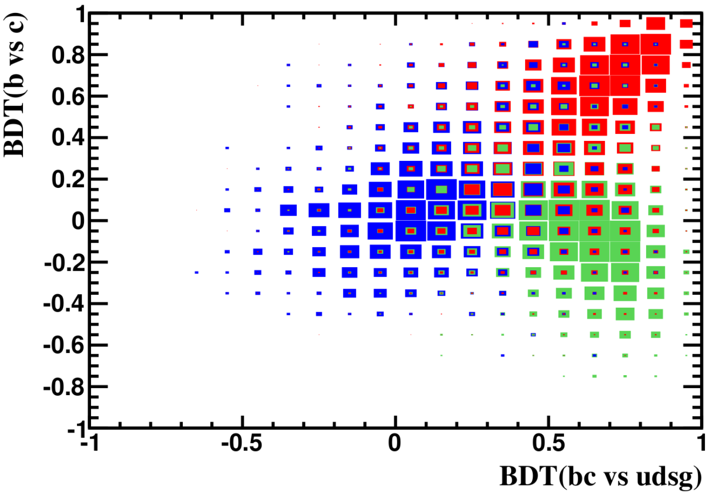

Figure 3.10: 2-dimensional $\text{BDT}_{bc|udsg}$ and $\text{BDT}_{b|c}$ distribution for b, c and light jet, obtained from simulation [66].

3.3.2 Online b-jets selection

The Run I data taking had no dedicated jet trigger, therefore the HLT2 inclusive topological trigger, introduced in Section 2.11, has been used to select a data sample enriched with b jets. The HLT2 topological algorithm has been designed to run on the online LHCb farm, so many features have been introduced to optimize the time performances. In the following the HLT2 inclusive topological algorithm is described in details. Only events that pass one of the L0 and one of the HLT1 trigger lines are processed by the HLT2 topological line. Selection requirements are applied to reconstructed tracks to ensure their quality and to limit their number for good time performances. The requirements are:

- $p_T > 500$ MeV/c and $p > 5$ GeV/c;
- the χ^2 obtained from the track fit is required to be less than 5, to reduce the fake tracks contamination;
- a χ^2_{IP} greater than 16 is required, to reduce the background of not-prompt tracks coming from the PV.

Successively 2,3,4-body particles are built as follow:

1. input tracks are combined together to obtain all possible 2-body particles;
2. if the DOCA between the two tracks is greater than 0.15 mm then the 2-body particle is rejected;
3. 3-body particles are build by combining the 2-body particles with all the other single tracks;
4. if the DOCA between the 2-body particle and the track is greater than 0.15 mm then the 3-body particle is rejected;
5. 4-body particles are build by combining the 3-body particles with all the other single tracks;
6. if the DOCA between the 3-body particle and the track is greater than 0.15 mm then the 4-body particle is rejected.

The n-body particle four-momentum is defined as the sum of tracks four-momenta, assuming the K mass. The n-body particle position is obtained with a vertex fit to the tracks, while the flight direction is defined as the vector that points its position from the PV. 2,3,4-body particles built in this way must satisfy selection criteria, to remove the background from light jets and D mesons:

- the corrected mass must be in the [4,7] GeV/c^2 range;
- the p_T of the hardest track in the n-body particle must be greater than 1.5 GeV/c;

- the sum of tracks p_T must be greater than 4 GeV/c, 4.25 GeV/c and 4.5 GeV/c respectively for 2-body, 3-body and 4-body particles;
- the sum of tracks χ^2_{IP} must be greater than 100, 150 and 200 respectively for 2-body, 3-body and 4-body particles;
- at least one track must have a track fit χ^2 less then 3;
- the flight distance χ^2 from the PV must be greater than 64;
- the direction of flight is required to be downstream from the PV;
- the mass of the (n-1)-body particle, used in the n-body particle combination, must be greater than 2.5 GeV/c^2 or alternatively the (n-1)-body particle must have a χ^2_{IP} greater than 16.

To further remove the background a Multivariate Algorithm is used. A BDT has been trained to separate b hadrons from the background. The following n-body particle related observables are used as input to the BDT:

- the n-body particle mass;
- the n-body particle corrected mass;
- the sum of tracks p_T;
- the maximum DOCA between the tracks;
- the χ^2_{IP} of the n-body particle, defined as the variation of the PV fit χ^2 when the tracks belonging to the n-body particles are removed from the fit.

The BDT is trained using as signal a b hadrons simulated sample, generated with Pythia 6 for pp collisions at 7 TeV, and as background a data sample obtained with the minimum bias trigger configuration in the 2010 data taking. Input variables to the BDT are discretized in order to ensure stability during the data taking. This technique is called Bonsai BDT (BBDT) [70]. If at least one n-body particle in the event has a BDT output (BDT$_{topo}$) greater then a fixed threshold, then the inclusive topological trigger decision is positive and the event is stored. If the n-body particles contains a muon track then a looser BDT requirement with respect to the nominal one is applied (*loose* topological trigger). The n-body particles that satisfy the trigger requirements are label with tag$_{trig}$. A jet is "topological tagged" if all the tracks of a tag$_{trig}$ belong to the jet: it is identified to be generated from a b quark, since the topological algorithm strongly suppress c jets.

3.3.3 Tagging performance

Two parameters are representative of the tagging algorithms performance:

- the $b(c)$ tagging efficiency, defined as the number of tagged $b(c)$ jets to the total number of reconstructed $b(c)$ jets;
- the light jets misidentification defined as the number of tagged light jets (generated from u, d, s or g) to the total number of reconstructed light jets.

The SV tagging and the topological tagging performance have been evaluated on simulation [66]. For this study b, c and light partons samples generated with PYTHIA 8 for pp collisions at 8 TeV have been used. The tagging efficiencies and the misidentification probability have been determined as a function of the reconstructed jet p_T and pseudo-rapidity (η) and they are shown in figure 3.11. The following conclusions are obtained:

- for jets with p_T greater than 20 GeV the mean SV-tagging efficiency on b jets is of about 60% and of about 25% for c jets;
- for jets with p_T greater than 20 GeV the mean trigger-tagging efficiency on b jets is of about 30% and less than 5% for c jets;
- for jets with p_T less than 20 GeV the $b(c)$ tagging efficiencies are significantly lower;
- for $2.2 < \eta < 4.2$ the tagging efficiencies are almost constant;
- the light parton misidentification probability is of about 0.3% for SV-tagging and of about 0.1% for topological trigger tagging.

The low trigger-tagging efficiency is due to the fact that the trigger algorithm is optimized for B meson measurements, where a strong c background suppression is required at the cost of a lower selection efficiency. This limitation has been overcame in the Run II, where algorithms optimized for the selection of heavy flavour jets have been implemented in the trigger system, as explained in Section 6.2.

The performance of jet tagging algorithms have been measured on data [66], to check if differences with simulation exist. Data events that contain a fully reconstructed b or c hadron or a high-p_T muon, which are referred as "event-tag", have been used. A test jet associated to the same PV of the event-tag is required, with a $|\Delta\phi|$ with respect to the event-tag greater than 2.5 to reduce the possibility of contamination from the event-tag. Only jets with p_T in the [10,100] GeV/c range are selected, since there are no large data sample to measure the efficiencies of jets with p_T greater than 100. The following samples with different type of event-tags have been considered:

- B+jet: an enriched sample of b-jets, collected with the topological trigger in pp collisions at a centre-of-mass-energy of 8 TeV;

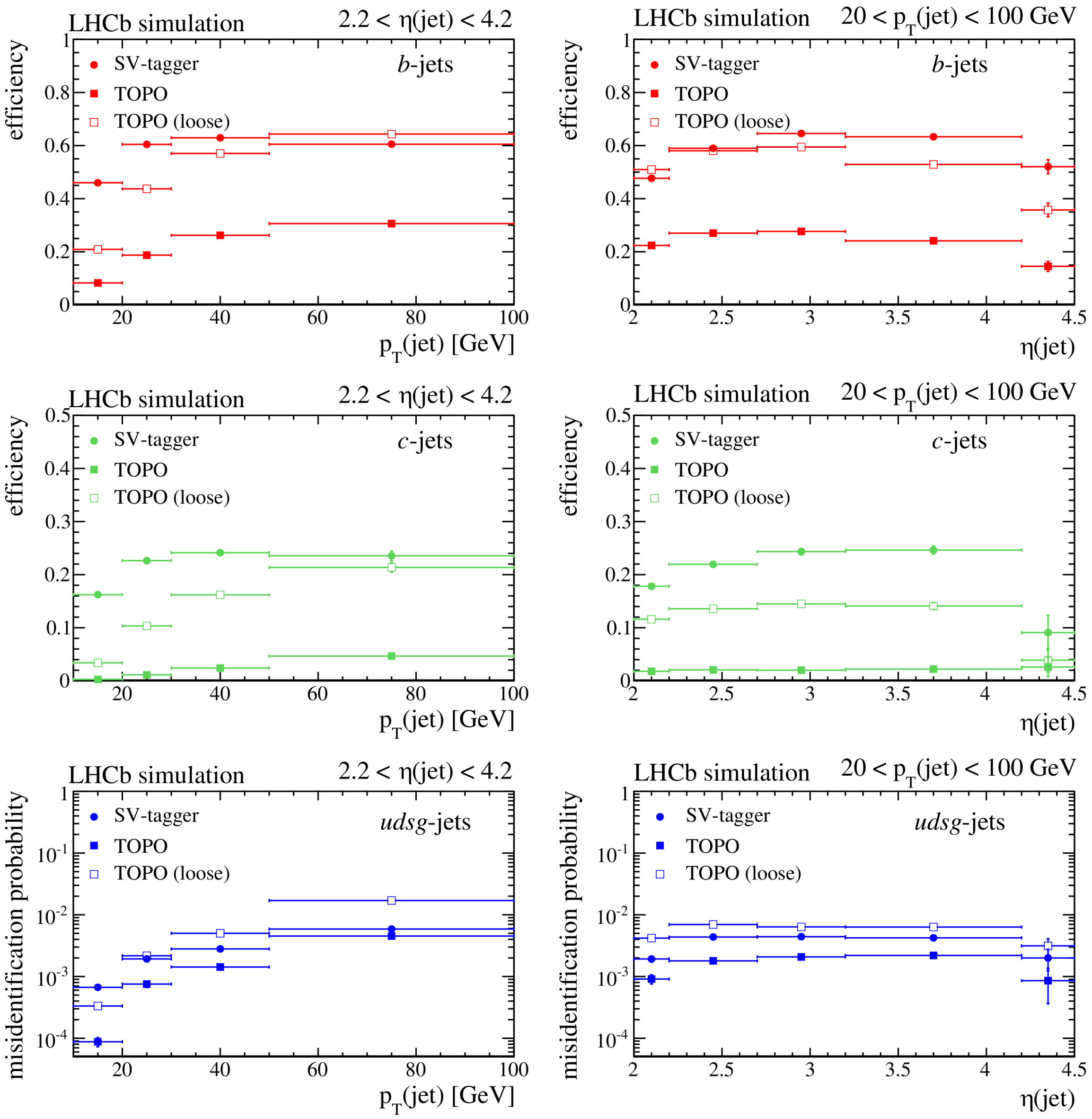

Figure 3.11: SV, topological and *loose* topological tagging efficiencies on b jets (first row) and c jets (middle row) and misidentification probability of light parton jets (last row) in different p_T and η intervals. These results have been obtained from simulation [66].

- D+jet: an enriched sample of b and c-jets, due to the $b \rightarrow c$ transition in the B hadrons decays, collected with the charm trigger in pp collisions at a centre-of-mass energy of 8 TeV;

- μ+jet: a sample where a high p_T displaced muon is selected to enrich the content of

b and c-jets, collected with the high p_T muon trigger in pp collisions at a centre-of-mass-energy of 8 TeV;

- W+jet: where the W boson is identified through a high p_T prompt muon, collected with the high p_T muon trigger in pp collisions at a centre-of-mass-energy of 8 TeV; this sample is composed by 95% of light parton jets and it is used to measure the light jets misidentification.

To obtain the b and c SV tagging efficiencies the following method is used:

1. templates of $\mathrm{BDT}_{b|c}$, $\mathrm{BDT}_{bc|udsg}$ distributions for b, c and light jets are obtained using MC events;
2. a combined fit is performed to $\mathrm{BDT}_{b|c}$ and $\mathrm{BDT}_{bc|udsg}$ distributions of B+jet, D+jet and μ+jet events where the test jet is tagged. In the fit the yield of tagged b, c and light jets events is measured (N_b^{tag}, N_c^{tag} and N_ℓ^{tag} respectively). The $\mathrm{BDT}_{b|c}$ and $\mathrm{BDT}_{bc|udsg}$ templates described in point 1 are used in the fit. As example, in figure 3.12, the result of the fit to the $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ distributions of the B+jet sample is shown;

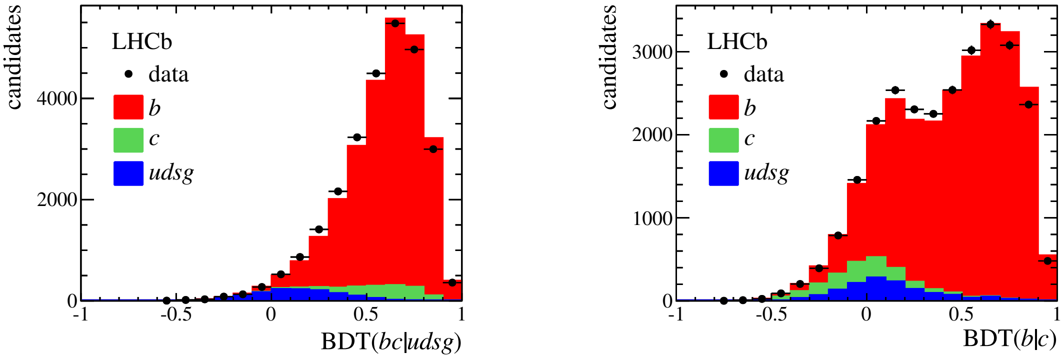

Figure 3.12: Fits to the $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ distributions of the B+jet sample [66].

3. templates of $\chi^2_{IP(max-pt)}$ distributions for b, c and light jets are built using MC events. $\chi^2_{IP(max-pt)}$ is defined as the χ^2_{IP} of the track with the highest p_T in the jet;
4. a combined fit is performed to the $\chi^2_{IP(max-pt)}$ distributions of B+jet, D+jet and μ+jet events where the SV tagging is not applied to the test jet. The templates described in point 3 are used. In the fit the yield of total b, c and light jets events is measured (N_b^{tot}, N_c^{tot} and N_ℓ^{tot} respectively);
5. the $b(c)$ tagging efficiency is calculated as

$$\epsilon_{b(c)} = \frac{N^{tag}_{b(c)}}{N^{tot}_{b(c)}}.$$

The topological tagging efficiencies are obtained with the same method, but fitting the BDT$_{topo}$ distributions in B+jet, D+jet and μ+jet instead of the BDT$_{b|c}$ and BDT$_{bc|udsg}$ distributions. The efficiency measurements have been performed in different intervals of p_T and η. Figures 3.13 and 3.14 show the measured SV tagging efficiencies and the topological tagging efficiencies, where they are presented as the ratio between the efficiencies measured in data and those obtained in simulation. The uncertainties bars reported take into account of statistical and systematic uncertainties. Systematic uncertainties are mainly associated to the BDT templates in the fits, to the mismodeling $\chi^2_{IP(max-pt)}$ and to the gluon splitting, that can create $b\bar{b}$ and $c\bar{c}$ couples with high probability to be tagged. The measured efficiencies and the data/MC scale factors for the SV tagger in different p_T intervals are reported in table 3.3.

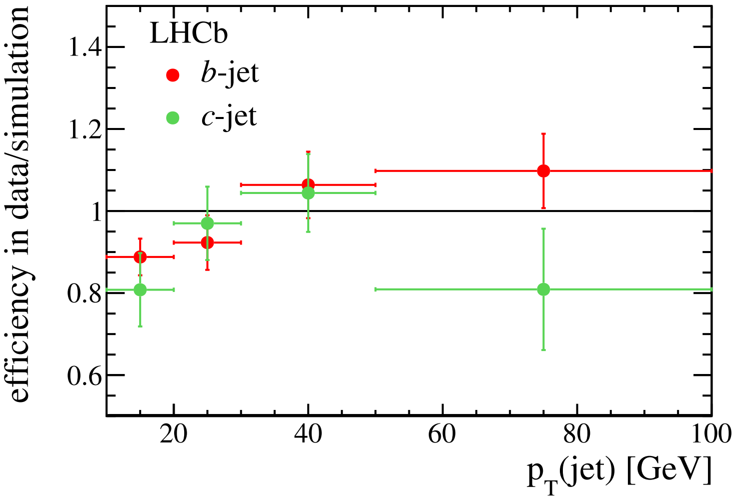

Figure 3.13: Ratio between the SV tagging efficiencies measured in data and those obtained from simulation, for b and c jets [66].

Table 3.3: SV tagging algorithm efficiencies measured on data.

		ϵ(data)/ϵ(simulation)			ϵ(data) (%)	
jet p_{T} (GeV)	jet η	b jets	c jets	(b, c) jets	b jets	c jets
10–20	2.2–4.2	0.89 ± 0.04	0.81 ± 0.09	0.91 ± 0.04	38 ± 2	14 ± 1
20–30	2.2–4.2	0.92 ± 0.07	0.97 ± 0.09	0.97 ± 0.04	61 ± 3	23 ± 1
30–50	2.2–4.2	1.06 ± 0.08	1.04 ± 0.09	0.97 ± 0.04	65 ± 3	25 ± 1
50–100	2.2–4.2	1.10 ± 0.09	0.81 ± 0.15	1.05 ± 0.06	70 ± 4	28 ± 4
20–100	2–2.2	1.00 ± 0.07	1.12 ± 0.10	1.05 ± 0.03	56 ± 2	20 ± 1

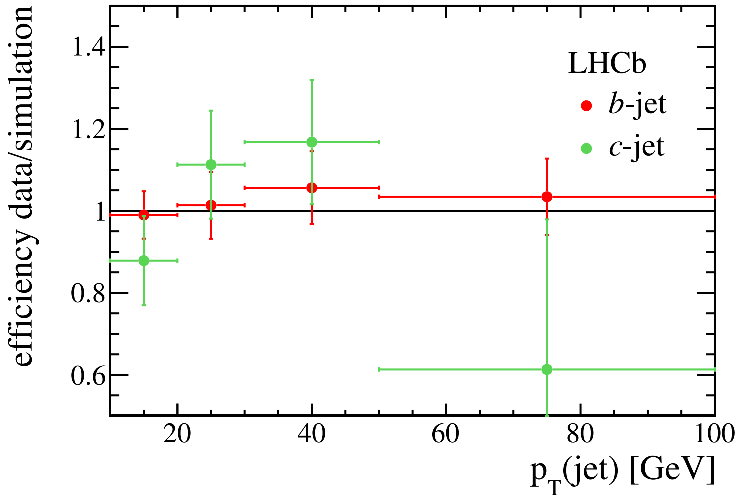

Figure 3.14: Ratio between the topological tagging efficiencies measured in data and those obtained from simulation, for b and c jets [66].

The light jets misidentification (ϵ_ℓ) has been obtained applying a similar method to the W+jet data sample. As before N_b^{tag}, N_c^{tag}, N_ℓ^{tag} are obtained with a fit to the BDTs distribution of tagged jets and N_b^{tot}, N_c^{tot}, N_ℓ^{tag} are obtained with a fit to $\chi^2_{IP(max-pt)}$ distribution of events where the jet tagging is not applied. As example the result of the fit to the $\chi^2_{IP(max-pt)}$ distribution in the W+jet sample is presented in figure 3.15. The light

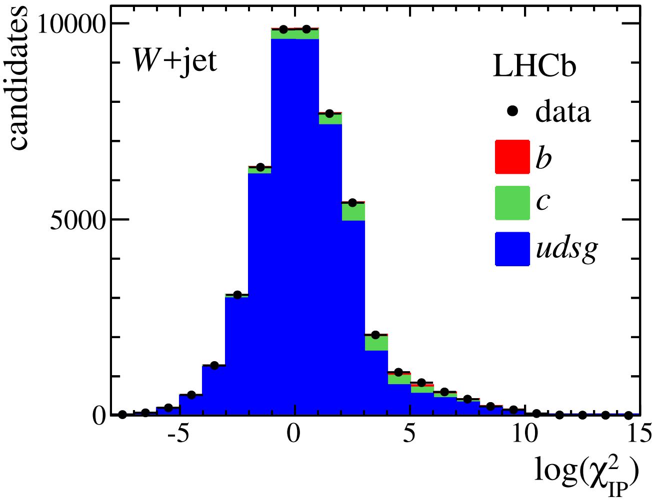

Figure 3.15: Fits to the $\chi^2_{IP(max-pt)}$ before applying any tagging requirements, in the W+jet sample [66].

jets misidentification is calculated as

$$\epsilon_\ell = \frac{N_\ell^{tag}}{N_\ell^{tot}}.$$

The measured misidentification probability is consistent with simulation for both SV and topological taggers as shown respectively in figures 3.16 and 3.17, where the data/MC ratios are presented.

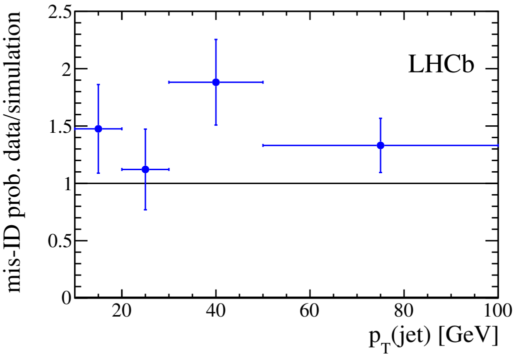

Figure 3.16: Ratio between the SV tagging light jet misidentification probability measured in data and those obtained from simulation, for b and c jets [66].

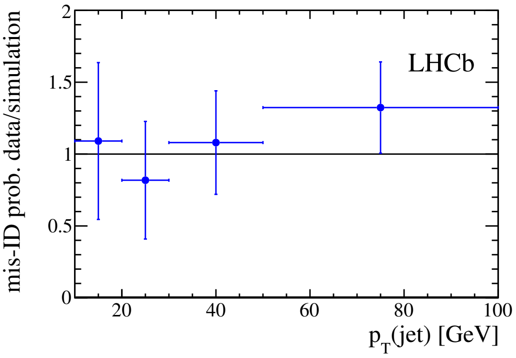

Figure 3.17: Ratio between the topological trigger light jet misidentification probability measured in data and those obtained from simulation, for b and c jets [66].

Chapter 4

Measurement of the $Z \rightarrow b\bar{b}$ cross section and determination of the Jet Energy Scale

4.1 Introduction

The observation of the $Z \rightarrow b\bar{b}$ is considered challenging at a hadron collider, due to the high $b\bar{b}$ QCD background with respect to the signal yield. Nevertheless its identification is important for b-jets reconstruction validation, and to measure the Jet Energy Scale (definition in Section 3.2.5). $Z \rightarrow b\bar{b}$ can be considered a standard "candle" for any $b\bar{b}$ resonance search since its mass and width are well known [21]. Many techniques have been developed to reconstruct $Z \rightarrow b\bar{b}$ by the CDF [71] and the ATLAS [72] collaborations.

The analysis proceeds through the following steps:

1. a b-dijet selection is applied to the dataset collected in 2012, requiring two reconstructed jets identified as originating from a b quarks;

2. a recoil jet selection is applied. The recoil jet can be used to define discriminating observables to separate the $Z \rightarrow b\bar{b}$ from the QCD background;

3. a multivariate technique (MVA) is employed to build a discriminator uncorrelated with the dijet invariant mass;

4. the MVA output is used to define a control region with low $Z \rightarrow b\bar{b}$ contribution where the QCD background mass model can be probed;

5. a fit to the dijet invariant mass is performed to extract the $Z \rightarrow b\bar{b}$ cross section and Jet Energy Scale measurements.

4.2 Datasets selection

4.2.1 Data and MC samples

Data used to search for $Z \rightarrow b\bar{b}$ have been collected by LHCb during the Run I, corresponding to a centre-of-mass-energy of 8 TeV. During the Run I LHCb did not have a dedicated jets trigger. Therefore events are required to satisfy the inclusive topological trigger described in Section 3.3.2. The total integrated luminosity of this dataset is 2.00 ± 0.02 fb^{-1}. The luminosity calibration has been obtained with two different and independent direct methods, the "van der Meer scan" and the "beam-gas imaging" described in [73]. For proton-proton interactions at 8 TeV a relative precision of the luminosity calibration of 1.47% is obtained using the van der Meer scan, and of 1.43% using the beam gas imaging, for a combined precision of 1.12%.

A $Z \rightarrow b\bar{b}$ MC sample is used to study the signal properties an to determine the selection efficiency. The simulation have been performed with the LHCb framework GAUSS [55]. Events have been generated with PYTHIA 6 [56] requiring the two b quarks from the Z decay inside the LHCb acceptance, $2 < \eta < 5$. About 10 millions of $Z \rightarrow b\bar{b}$ events have been generated in this way. The high statistics $Z \rightarrow b\bar{b}$ signal sample has been compared with a smaller sample of $Z \rightarrow b\bar{b}$ MC events generated with PYTHIA 8, verifying that no significative difference exist.

The main backgrounds are:

- $Z \rightarrow c\bar{c}$, estimated via MC using PYTHIA 8 as generator, requiring two c in the LHCb acceptance;
- $W \rightarrow q\bar{q}'$, estimated via MC using PYTHIA 8 as generator, requiring two light quarks in the LHCb acceptance;
- $t\bar{t}$, estimated via MC using PYTHIA 8 as generator, requiring two b quarks from top decays in the LHCb acceptance;
- QCD background: it is evaluated with data driven techniques but a MC sample is used to perform cross-checks. The sample is a mixture of QCD processes, generated with PYTHIA 8: an inclusive $b\bar{b}$ process with a transferred momentum between the two partons in the hard interaction ($\hat{p}_T$) greater than 120 GeV and an inclusive $b\bar{b}$ process where one lepton (muon or electron) from the b decay is in the LHCb acceptance.

4.2.2 Data selection

In the $Z \rightarrow b\bar{b}$ measurement jets are reconstructed using the algorithm described in Section 3.2 with radius parameter $R = 0.5$. A pre-selection is applied to the events selected by the inclusive topological trigger described in Section 3.3.2:

- at least one reconstructed jet tagged by the Topological tagging algorithm (tag_{trig}, Section 3.3.2) is required; a jets with this feature is defined Triggered On Signal (TOS) and labeled as jet_1;
- at least one second reconstructed jet, different from jet_1, tagged by the SV tagging algorithm (tag_{SV}, Section 3.3.1) is required; this jet is labeled as jet_2;
- jet_1 and jet_2 must be associated to the same PV;
- $p_T(\text{jet}_1)$ and $p_T(\text{jet}_2) > 17$ GeV[1];
- $|\Delta\phi_{12}| > 2.5$;

The two selected jets, jet_1 and jet_2, form the b-dijet candidate. If more than one b-dijet candidate is found in the same event only one is selected with the following criteria: One b-dijet candidate per event is selected with the following procedure:

1. if more than one jet in the event has a tag_{trig} than the one with the highest p_T is selected as the first b-jet candidate (jet_1);
2. if more than one jet in the event has a tag_{SV} than the one with the highest p_T is chosen as the second b-jet candidate (jet_2).

After the multiple candidates removal additional requirements are applied:

- $2.2 < \eta(\text{jet}_{1(2)}) < 4.2$, to ensure that the entire jets cones are inside the fully instrumented LHCb acceptance;
- $p_T(\text{jet}_1)$ and $p_T(\text{jet}_2) > 20$ GeV, to reduce the QCD background at low mass but preserving most signal events.

The p_T distributions of the two b-jets candidates after the b-dijet selection is different in data and signal MC as shown in figures 4.1 and 4.2. The $|\Delta\phi_{12}|$ distribution is shown in figure 4.3.

After the b-dijet selection, a recoil jet is selected in order to discriminate the $Z \to b\bar{b}$ signal from the QCD background. Recoil jet techniques have been also used in the CDF [71] and ATLAS [72] $Z \to b\bar{b}$ analyses. The recoil jet is defined as followed: for each jet associated to the same PV of jet_1 and jet_2, different from those, the sum of three jets momenta, jet_1, jet_2 and the third jet, is computed; the jet that minimizes the p_T is chosen as recoil jet and called jet_3. The p_T distribution of the sum of jet_1, jet_2 and jet_3 momenta is shown in figure 4.4 for data and $Z \to b\bar{b}$ MC. Additional requirements are:

- $2.2 < \eta(\text{jet}_3) < 4.2$;
- $p_T > 10$ GeV, to reject underlying event jets that are not well reproduced by MC.

---

[1] In this Chapter natural units where $\hbar = c = 1$ are used.

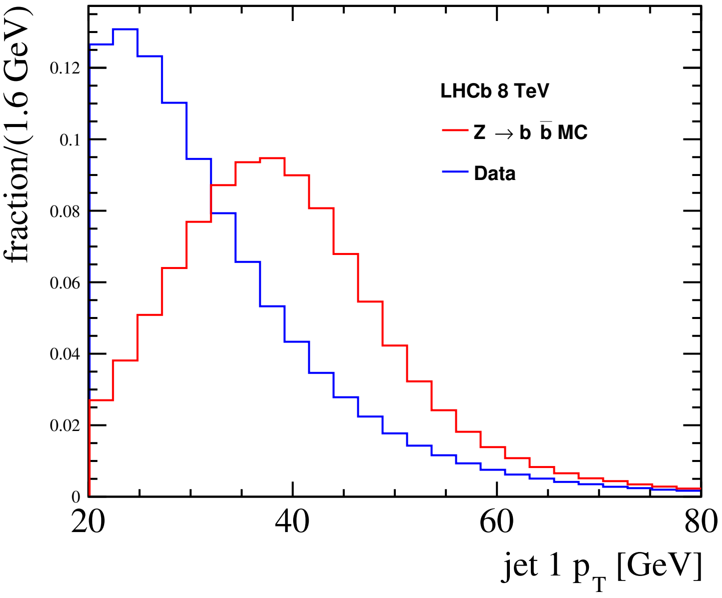

Figure 4.1: Distribution of the jet$_1$ transverse momentum for data and $Z \to b\bar{b}$ MC events, after the b-dijet selection. The distributions are normalized to unit area.

Finally events in the b-dijet invariant mass (m_{jj}) window [45,165] GeV are selected, to avoid edge effects in the fit to the dijet invariant mass described in Section 4.4. The cumulative efficiencies of the selection cuts on the $Z \to b\bar{b}$ signal are presented in table 4.1. They are computed as

$$\epsilon = \frac{N_{sel}}{N_{gen}},$$

where N_{sel} is the number of MC events that pass the corresponding cut and all the previous requirements in the list, N_{gen} is the number of generated MC events. The total number of selected data events is 616267.

4.2.3 Yields prediction

The predictions of the $Z \to b\bar{b}$ signal and backgrounds yields after the full selection have been obtained using the formula

$$N^{exp} = \mathcal{L} \cdot \sigma \cdot \mathcal{A} \cdot \epsilon,$$

where:

- $\mathcal{L}$ is the integrated luminosity;
- σ is the theoretical cross section. The theoretical cross sections for relevant processes are reported in table 4.2. The $Z \to b\bar{b}$ and $Z \to c\bar{c}$ cross sections are obtained

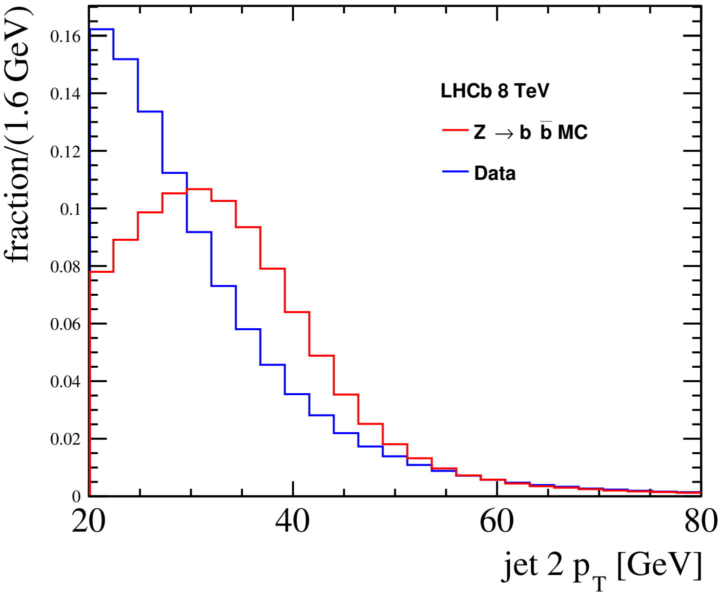

Figure 4.2: Distribution of the jet$_2$ transverse momentum for data and $Z \to b\bar{b}$ MC events, after the b-dijet selection. The distributions are normalized to unit area.

Table 4.1: Cumulative $Z \to b\bar{b}$ MC efficiencies of the selection cuts. Each efficiency is computed as the number of MC events that pass the corresponding cut and all the previous requirements divided to the number of generated events. Corrections for data/MC differences are not included.

Requirement	efficiency
pre-selection and multiple candidates removal	8.7%
$2.2 < \eta(jet_1) < 4.2$	8.1%
$p_T(jet_1) > 20$ GeV	8.0%
$2.2 < \eta(jet_2) < 4.2$	7.9%
$p_T(jet_2) > 20$ GeV	7.6%
n. jets from PV > 2	3.3%
$2.2 < \eta(jet_3) < 4.2$	2.8%
$p_T(jet_3) > 10$ GeV	1.1%
$45 < m_{jj} < 165$ GeV	0.99%

scaling the theoretical $Z \to \mu\mu$ cross section at NNLO computed in [74] for the corresponding measured branching ratios taken from [21]. In the same way the $W \to q\bar{q}'$ cross section is obtained scaling the theoretical $W \to \mu\nu$ cross section at NNLO computed in [74] for the corresponding branching ratio [21]. The theoretical $t\bar{t}$ cross section at NNLO is taken from [75].

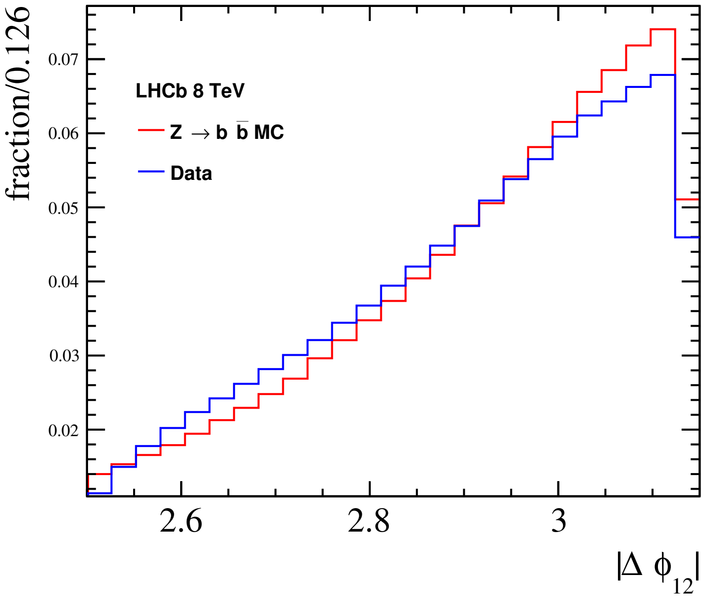

Figure 4.3: Distribution of $|\Delta\phi_{12}|$ for data and $Z \to b\bar{b}$ MC events, after the b-dijet selection. The distributions are normalized to unit area.

Table 4.2: Theoretical cross sections and errors for the studied processes.

Process	Cross Section (nb)	Uncertainty
$Z \to b\bar{b}$	5.076	3.5%
$Z \to c\bar{c}$	4.039	3.9%
$W \to q\bar{q}'$	76.24	2.7%
$t\bar{t}$	0.2477	6.4%

- A is the LHCb acceptance factor calculated using PYTHIA, requiring the decay products in the LHCb acceptance ($2 < \eta < 5$).
- ϵ is the selection efficiency, obtained as the number of selected MC events to the number of generated events. In this computation weights for data/MC differences for b-tagging and jets identification efficiencies are applied to each event. These weights are obtained as explained in Chapter 3 and they depends on the jet p_T.

The estimated efficiencies and the predicted yields are reported in table 4.3.

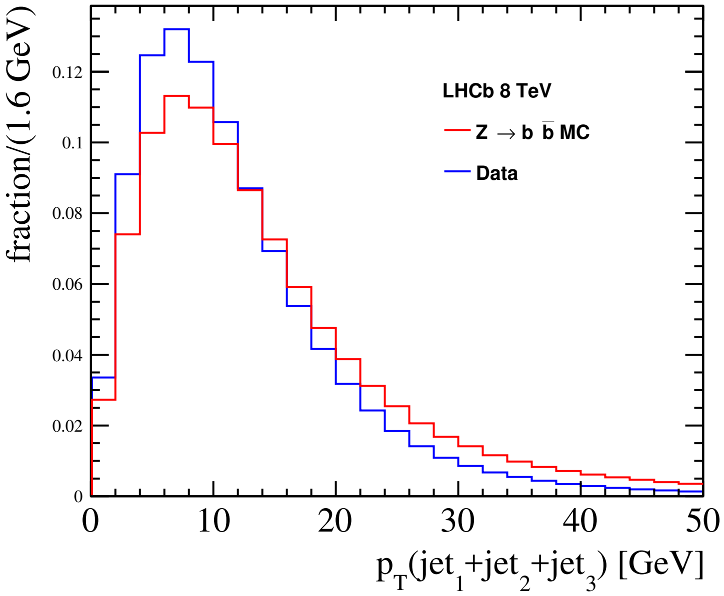

Figure 4.4: p_T distribution of the sum of jet$_1$, jet$_2$ and jet$_3$ momenta after the recoil jet selection, before applying η and p_T cuts on jet$_3$. Distributions for data and $Z \rightarrow b\bar{b}$ MC are shown. The distributions are normalized to unit area.

Table 4.3: Selection efficiencies, acceptance and yield predictions for the studied processes. The selected data yield is also reported.

Process	Acceptance	$\sigma \cdot A$	Efficiency	Yield
$Z \rightarrow b\bar{b}$	0.161	817 pb	1.04%	17020
$Z \rightarrow c\bar{c}$	0.213	860 pb	0.0340%	584
$W \rightarrow q\bar{q}'$	0.110	8386 pb	0.00138%	231
$t\bar{t}$	0.109	27.0 pb	0.0497%	51
data				616267

4.3 Signal selection

The $Z \to b\bar{b}$ signal has to be extracted from QCD events. This background is constituted by b jets and misidentified c and light jets. The observables used to this purpose have to be as much independent as possible from the dijet mass in order not to bias the QCD distribution and use it in resonances search. We divide the physical quantities into three categories:

- observables related to the dijet system;
- observables related to the recoil jet;
- global event observables.

The first category is connected to the kinematic and the topology of the $Z \to b\bar{b}$ decay. The second is related to the Z production mechanism since the recoil jet can be produced directly in the hard interaction, as shown in the Feynman diagram in figure 4.5. In other cases it is possible that a third jet is originated from a gluon emitted by one of the b quark in the Z decay, and it is misidentified as a recoil jet. The third category is related to the shape of the entire event. In order to compare signal to background, data is considered representative of the QCD background. In fact the fraction of expected $Z \to b\bar{b}$ events in the selected data sample is less than 2% and less then 5% in the dijet invariant mass region from 65 to 95 GeV. These fractions are computed as the number of expected Z events, obtained as explained in Section 4.2.3, to the number of selected data events.

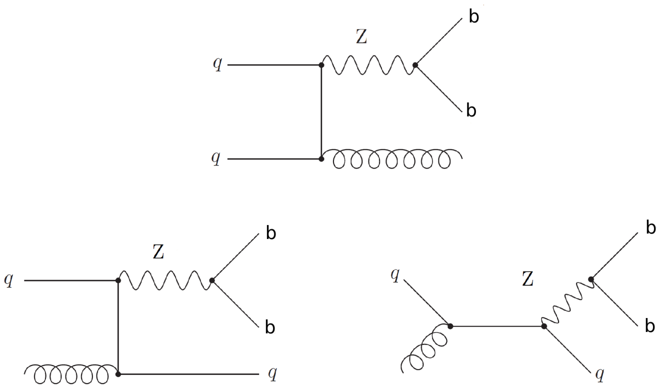

Figure 4.5: Leading order Feynman diagrams for $Z(\to b\bar{b}) + jet$ production.

In this Section the observables distributions in data are compared to those of $Z \to b\bar{b}$ MC, for events in the dijet invariant mass region from 65 to 95 GeV. Figure 4.6 shows the dijet invariant mass (m_{jj}) distribution of $Z \to b\bar{b}$ MC in comparison with data. An estimation of the signal mean and width has been obtained with a gaussian fit to the

$Z \to b\bar{b}$ invariant mass distribution: the signal peak has a mean of about 80 GeV and a resolution of about 12 GeV, below the nominal Z mass of 91.2 GeV [21], as explained in section 3.2.3. As can be seen in the comparison with data, this invariant mass resolution makes the signal extraction from the background challenging.

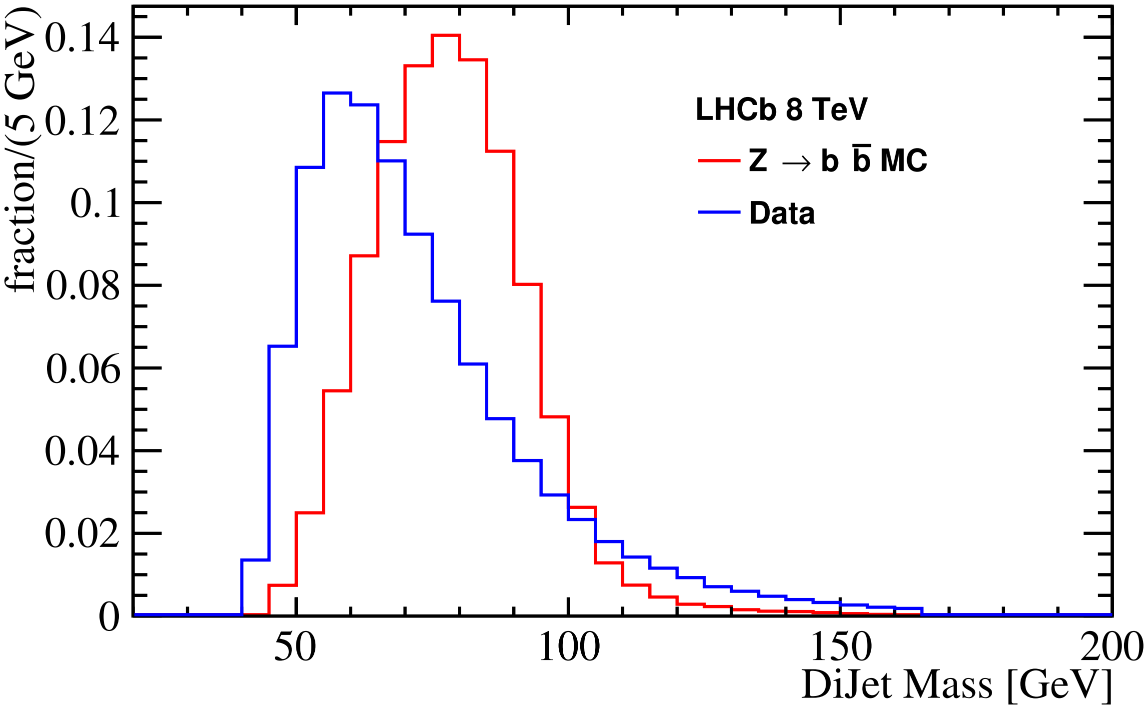

Figure 4.6: Distribution of the dijet invariant mass for data and $Z \to b\bar{b}$ MC events. The full selection is applied with the exception of the $45 < m_{jj} < 165$ GeV requirement. The distributions are normalized to unit area.

4.3.1 Dijet related observables

The p_T of the dijet system is shown in figure 4.7, for data and $Z \to b\bar{b}$ MC. The mean of the distributions is significantly higher in the signal with respect to the data. Since the dijet p_T is strongly correlated with the dijet mass it is not used to separate the signal from the background. We consider the absolute difference between the two jets pseudo-rapidity ($|\Delta\eta_{12}|$) and the p_T asymmetry defined as:

$$A_{12} = \frac{|p_T(jet_1) - p_T(jet_2)|}{p_T(jet_1) + p_T(jet_2)}.$$

These observables have a low correlation with the Jet Energy Scale: $|\Delta\eta_{12}|$ depends only on the jets direction and A_{12} is a ratio of momenta therefore the JES cancel out. $|\Delta\eta_{12}|$

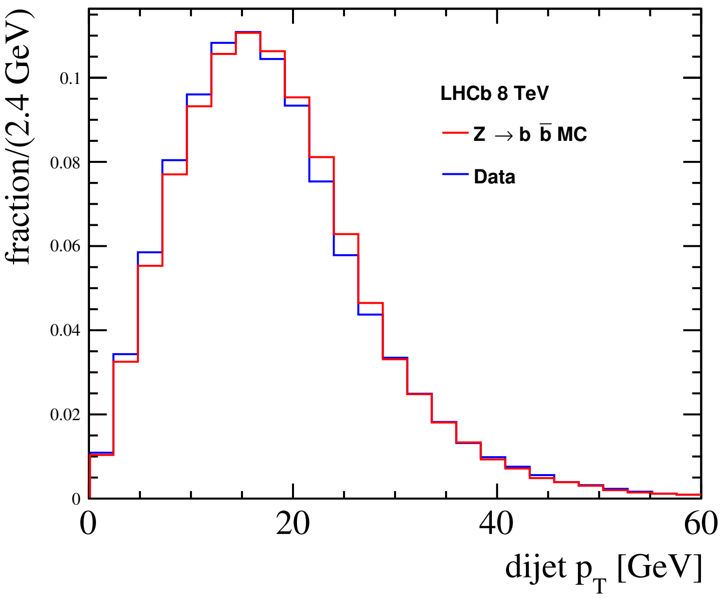

Figure 4.7: Distribution of the dijet p_T for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

and A_{12} for data and MC are presented in figures 4.8 and 4.9. These variables have a good discriminating power.

4.3.2 Recoil jet related observables

In principle the observables related to the recoil jet are uncorrelated to the dijet mass, since its production is independent from the Z decay. Exceptions are the cases where a third jet, originated from a gluon emitted by one of the b quarks in the Z decay, is misidentified as the recoil jet, resulting in a loss of dijet invariant mass. Figure 4.10 shows the p_T distribution of the recoil jet for $Z \to b\bar{b}$ and data, demonstrating a good agreement with data softer than MC.

In figure 4.11 the angle between the recoil jet and the b-dijet direction in the xy plane ($|\Delta\phi_{bb3}|$) shows that the recoil jet tends to be back-to-back with the dijet in this plane.

Another useful observable is the angle between the recoil jet momentum and the b-dijet momentum in the 3-dimensional space. $|\Delta\Theta_{bb3}|$ is defined as the angle between the b-dijet momentum direction in the laboratory frame and the recoil jet momentum in the b-dijet rest frame. As can be seen in figure 4.12 this variable has a good discriminating power.

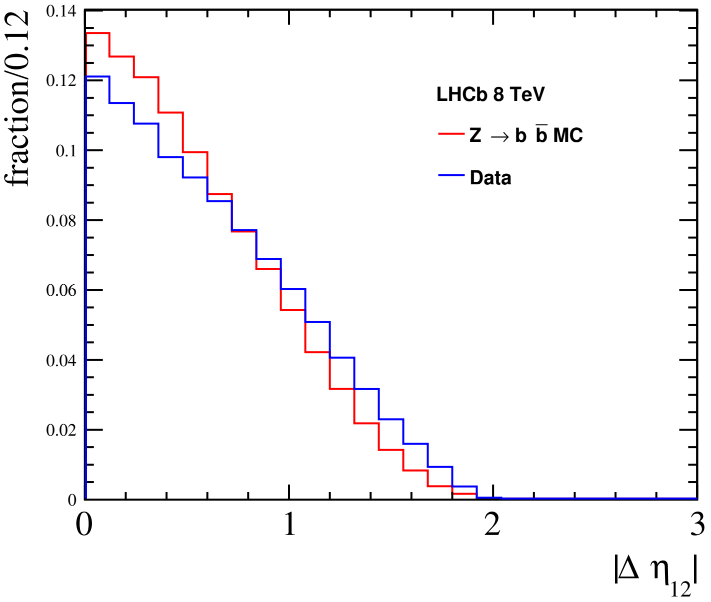

Figure 4.8: Distribution of $|\Delta\eta_{12}|$ for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

4.3.3 Global event observables

Global event observables are related to the event shapes, measuring the geometrical properties of the energy flow in QCD events [76]. These can help to discriminate the signal from the QCD background. Two global event observables are studied: the thrust (T) and the sphericity (S). The thrust is defined as

$$T = max_{\vec{n}} \left(\frac{\sum_i \vec{p_i} \cdot \vec{n}}{\sum_i |\vec{p_i}|} \right),$$

where the sum runs on the particles that fulfill the Particle Flow requirements, $\vec{p_i}$ are particles momenta and $\vec{n}$ is an element of the versors space. The thrust measures if the event has evolved in a preferred direction: in forward events, such those at LHCb, T is close to 1. The sphericity is computed starting from the sphericity tensor, defined as:

$$S^{\alpha\beta} = \frac{\sum_i p_i^{\alpha} p_i^{\beta}}{\sum_i |\vec{p_i}|^2},$$

where the sum runs on the particles that fulfill the Particle Flow requirements and $p_i^{\alpha(\beta)}$ is the $\alpha(\beta)$ component of the particle momentum vector. $S^{\alpha\beta}$ is a 3×3 symmetric tensor: if its eigenvalues are λ_1, λ_2 and λ_3, where $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then the sphericity is defined as:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3).$$

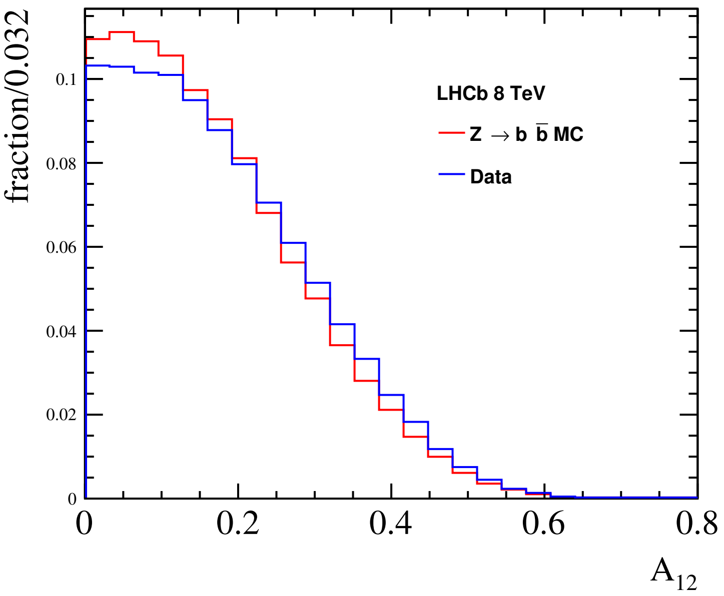

Figure 4.9: Distribution of A_{12} for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

S measures the fraction of particles momenta spherically distributed: S is equal to 1 for perfectly spherical events while it is close to 0 for 2-body decays. The distributions of T and S in $Z \to b\bar{b}$ MC and data are shown in figures 4.13 and 4.14.

In table 4.4 the observables described in the latter are summarized and the Pearson's correlation coefficients between them and the dijet invariant mass are reported, for signal (r_s) and background (r_b). The Pearson's coefficients are computed in the [45,160] GeV dijet invariant mass range. In the table the separation power of each variable is also shown, quantified by $max(\epsilon_s/\epsilon_b)$, where ϵ_s and ϵ_b are respectively the signal and background selection efficiencies for the cut on the analysed observable that makes ϵ_s/ϵ_b maximum. The efficiencies are computed using events in the [65,95] GeV range.

4.3.4 MC validation

The MC used to study the $Z \to b\bar{b}$ is validated using the $Z \to \mu\mu$+jet process. The $Z \to \mu\mu$ + jet data has a high purity of Z events, almost 99%, as shown in [77], therefore it can be compared directly to MC. This process can be used as benchmark for $Z \to b\bar{b}$+jet since they have the same production mechanism. Differences can arise when the recoil jet in $Z \to b\bar{b}$ is generated from gluons emitted by b quarks. The kinematic of the Z decay is also similar, with differences due to the smaller muons mass with respect to the b quarks.

A similar kinematical selection to that in [77] is applied to select the $Z \to \mu\mu$+jet sample in 8 TeV data, with some differences introduced to match the $Z \to b\bar{b}$ phase space

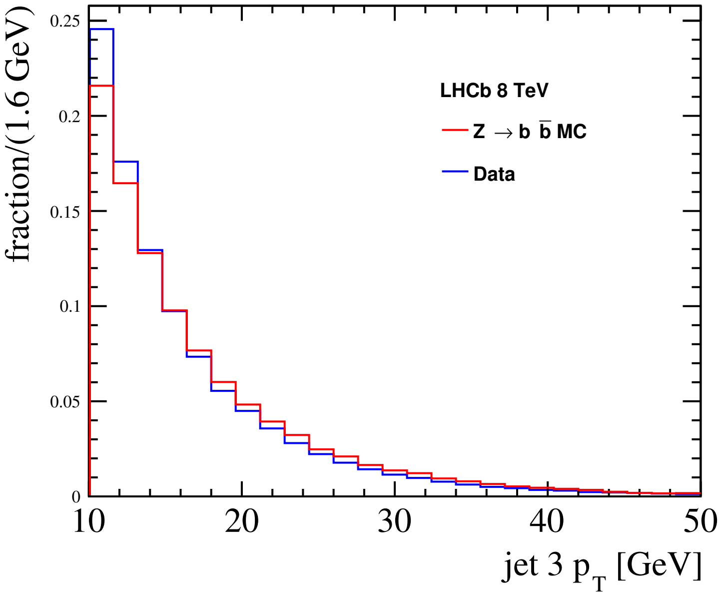

Figure 4.10: Distribution of the recoil jet p_T for data and $Z \rightarrow b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

Table 4.4: Summary of the observables presented in this Section. The Pearson's correlation coefficient between them and the dijet invariant mass are reported, for signal (r_s) and background (r_b). The separation power of each variable is also shown, quantified by $max(\epsilon_s/\epsilon_b)$, where ϵ_s and ϵ_b are respectively the signal and background selection efficiencies for the cut on the analysed observable that makes ϵ_s/ϵ_b maximum.

Observables	r_s	r_b	$max(\epsilon_s/\epsilon_b)$
$p_T(jet_1 + jet_2)$	0.362	0.549	1.02
$\|\Delta\eta_{12}\|$	0.0733	0.167	1.10
$\|A_{12}\|$	0.0877	0.284	1.05
$p_T(jet_3)$	0.124	0.201	1.10
$\|\Delta\phi_{bb3}\|$	0.0266	0.0782	1.12
$\|\Delta\Theta_{bb3}\|$	0.00729	0.0203	1.19
T	-0.0604	-0.0766	1.22
S	0.0456	0.131	1.14

considered in this analysis. The requirements are the following:

- two opposite charged tracks identified as μ (μ_1 and μ_2);

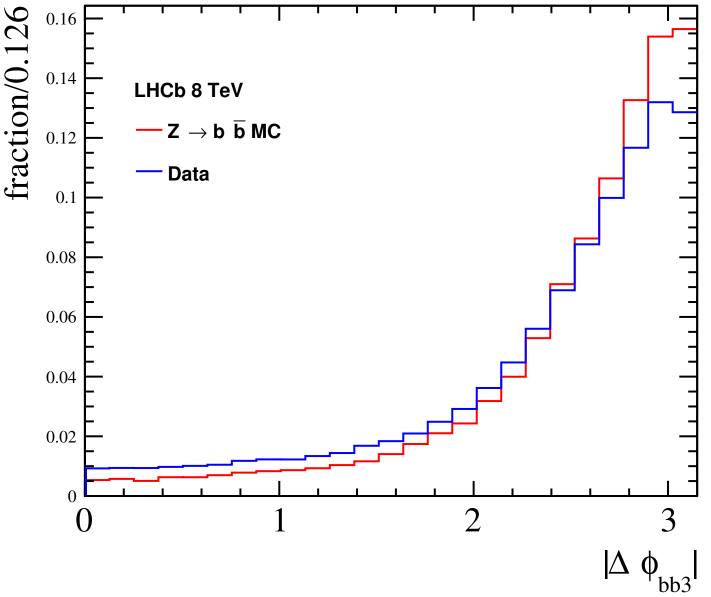

Figure 4.11: Distribution of $|\Delta\phi_{bb3}|$ for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

- μ_1 and μ_2 associated to the same PV;
- $p_T(\mu_{1,2}) > 20$ GeV;
- $2.2 < \eta(\mu_{1,2}) < 4.2$;
- $|\Delta\phi_{12}| > 2.5$;
- at least one jet in the event, that not contains the selected muons but it is associated to their PV;
- $p_T(\text{jet}) > 10$ GeV;
- $2.2 < \eta(\text{jet}) < 4.2$.

The $Z \to \mu\mu$ + jet MC sample used in the cross-checks has been generated with PYTHIA 8.

The number of selected $Z \to \mu\mu$+jet data events for an integrated luminosity of 928 pb^{-1} is 8517. The expected number of $Z \to \mu\mu$+jet calculated using the MC selection efficiency and the cross section provided by PYTHIA 8 is 10976 ± 963, where the error is due to the uncertainty on the cross section. The prediction is compatible with the observed yield within 2σ.

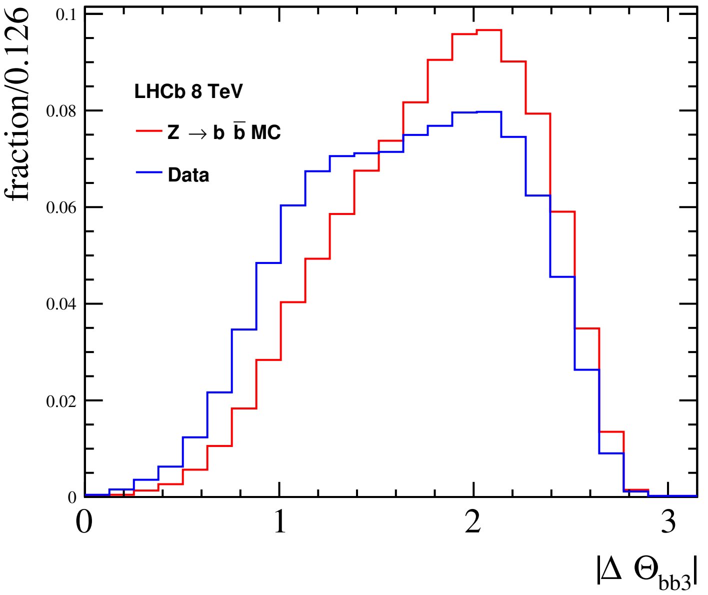

Figure 4.12: Distribution of $|\Delta\Theta_{bb3}|$ for data and $Z \rightarrow b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

The dijet related observables $|\Delta\eta_{12}|$ and $|A_{12}|$, calculated using muons instead of b jets, are shown in figures 4.15 and 4.16. From these distributions it is evident that $|\Delta\eta_{12}|$ is well reproduced in MC but A_{12} has differences with data. For this reason only $|\Delta\eta_{12}|$ will be considered for the MVA training.

The recoil jet observables p_T(jet), $|\Delta\phi_{\mu\mu3}|$ and $|\Delta\Theta_{\mu\mu3}|$ are shown in figures 4.17, 4.18 and 4.19 respectively, for $Z \rightarrow \mu\mu$ + jet data and MC. These distributions show that data and MC are in good agreement.

Finally global event observables T and S are compared between $Z \rightarrow \mu\mu$ + jet data and MC in figures 4.20 and 4.21 respectively. Even in this case the data is in agreement with MC.

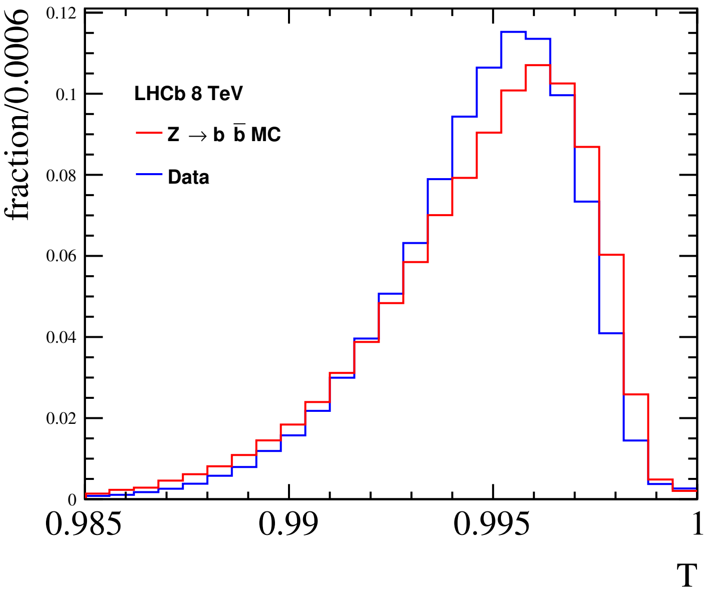

Figure 4.13: Distribution of T for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

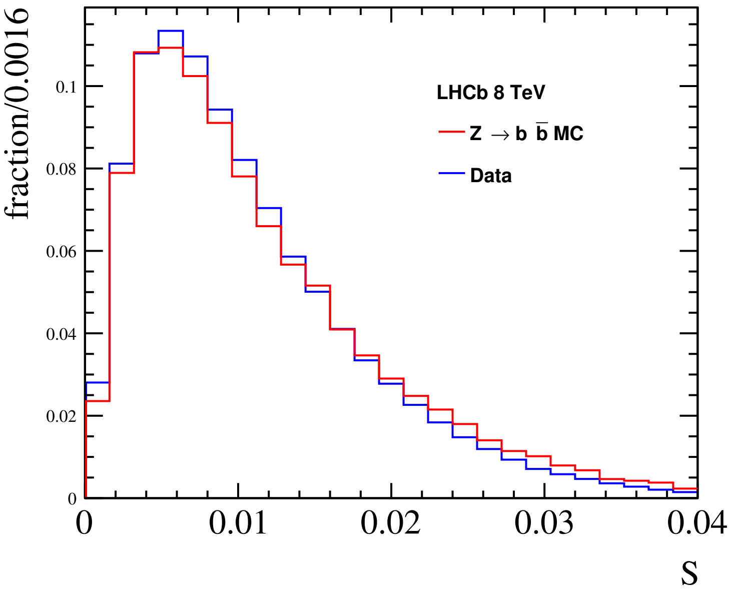

Figure 4.14: Distribution of S for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

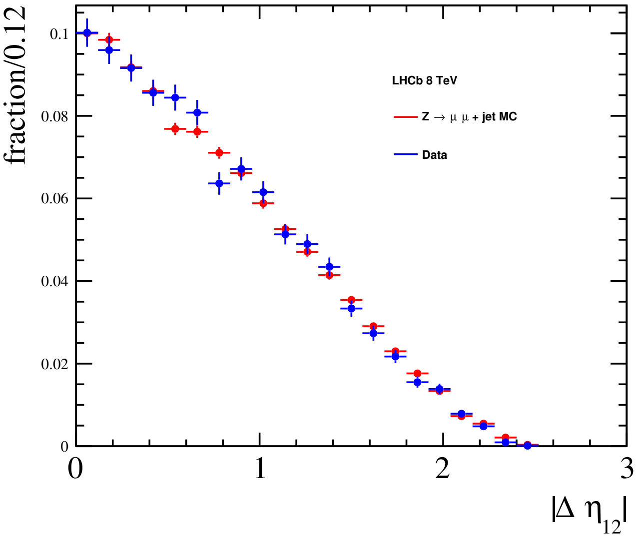

Figure 4.15: Distribution of $|\Delta\eta_{12}|$ for the $Z \rightarrow \mu\mu$ + jet data and MC, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is 0.16, the Kolmogorov-Smirnov probability is 0.60.

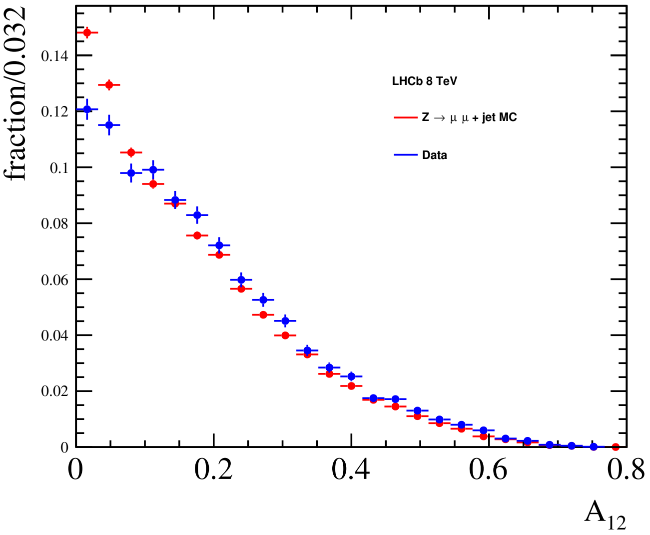

Figure 4.16: Distribution of A_{12} for the $Z \rightarrow \mu\mu$ + jet data and MC, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is $6.1 \cdot 10^{-10}$, the Kolmogorov-Smirnov probability is $1.0 \cdot 10^{-15}$.

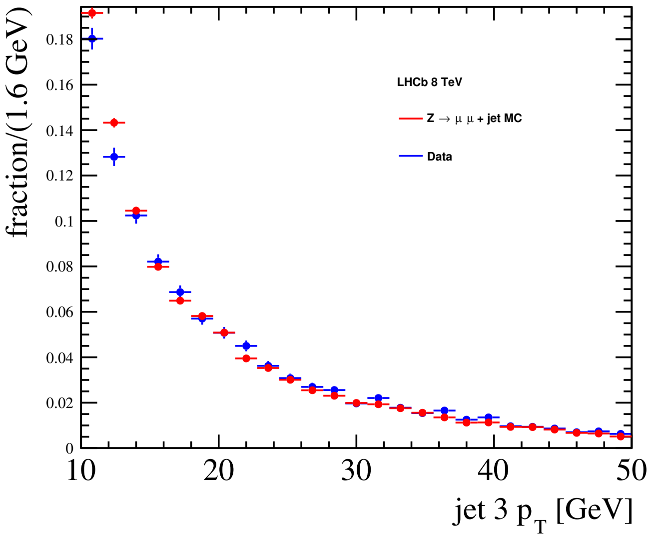

Figure 4.17: Distribution of the recoil jet p_T for the $Z \to \mu\mu$ + jet data and simulation samples, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is 0.041, the Kolmogorov-Smirnov probability is $6.1 \cdot 10^{-5}$.

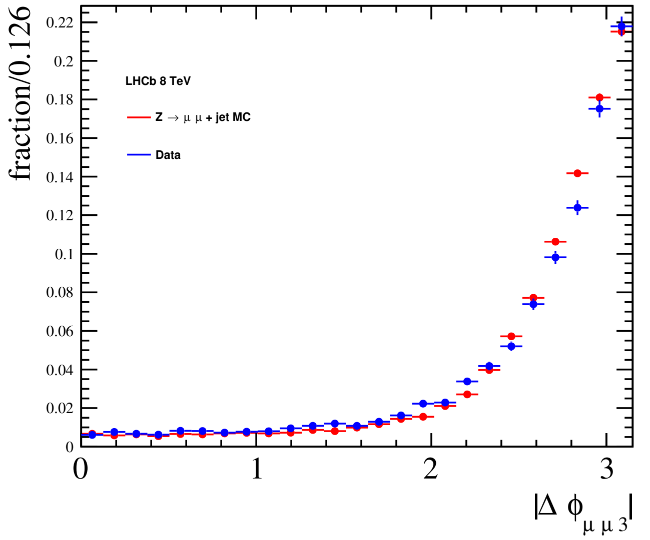

Figure 4.18: Distribution of $|\Delta\phi_{\mu\mu 3}|$ for the $Z \to \mu\mu$ + jet data and simulation samples, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is $5.6 \cdot 10^{-10}$, the Kolmogorov-Smirnov probability is $8.2 \cdot 10^{-9}$.

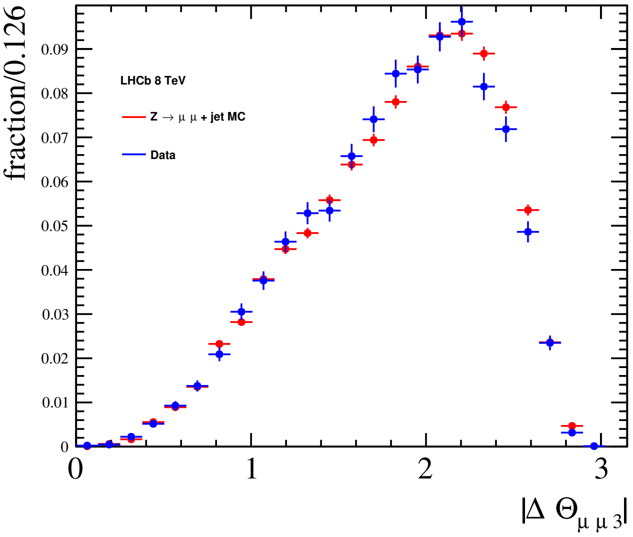

Figure 4.19: Distribution of $|\Delta\Theta_{\mu\mu3}|$ for the $Z \to \mu\mu$ + jet data and simulation samples, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is 0.065, the Kolmogorov-Smirnov probability is 0.015.

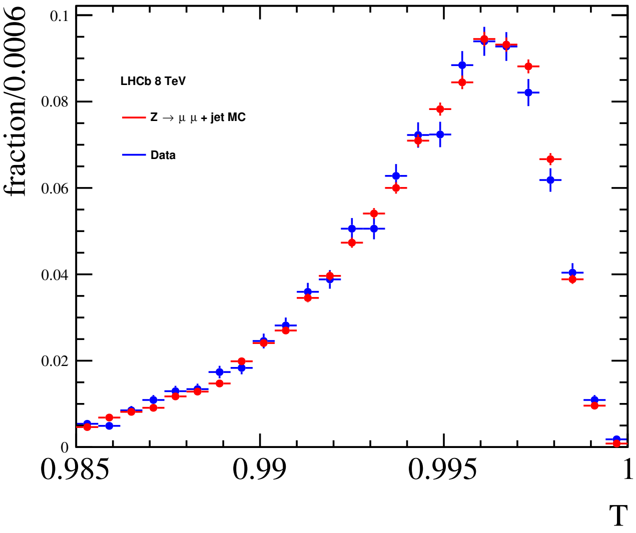

Figure 4.20: Distribution of T for the $Z \to \mu\mu$ + jet data and simulation samples, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is 0.081, the Kolmogorov-Smirnov probability is 0.52.

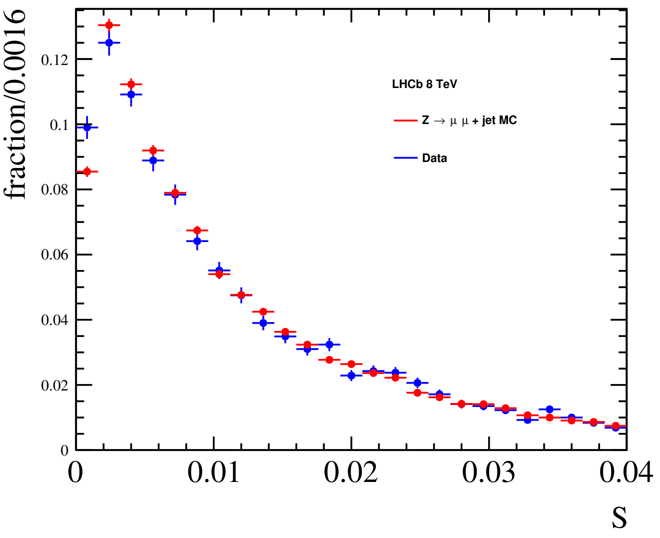

Figure 4.21: Distribution of S for the $Z \rightarrow \mu\mu$ + jet data and simulation samples, after the full selection. The distributions are normalized to unit area and uncertainties are statistical. The χ^2 test probability for compatibility of the two distributions is 0.030, the Kolmogorov-Smirnov probability is 0.19.

4.3.5 Uniform Gradient Boost algorithm

A multivariate classifier (MVA) is used to separate the signal $Z \rightarrow b\bar{b}$ from the QCD background. A Boosted Decision Trees (BDT) classifier is used for this purpose [67–69]. The structure and the weights of the BDT defined in [67–69] are determined with an iterative procedure called training, performed using a signal sample and a background sample. The classifier has to be uncorrelated with the dijet invariant mass to have an unbiased selection, therefore the uniform Gradient Boost method (*uGB*) [78] is applied to the BDT. In order to explain the *uGB* method several quantities are defined:

- $\vec{x}$: the input observables;
- $\vec{y}$: the observables we want uncorrelated from the BDT output, different from $\vec{x}$;
- γ_i: it is equal to +1 if the i-th event in the training is a signal event or -1 if it is a background event;
- the tree response r_{ij}: it is equal to +1 if the i-th event is classified by the j-th tree as signal or -1 if it is classified as background;
- the tree weight w_j, which is assigned to each tree j in the ensemble of trees (forest) at each stage of the BDT training;
- the event score s_i:

  $$s_i = \frac{\sum_j^{N_{tree}} w_j r_{ij}}{N_{tree}},$$

  where N_{tree} is the number of trees in the forest;
- the AdaBoost loss function:

  $$L_{ada} = \sum_i^N exp(-\gamma_i s_i),$$

  where N is the number of events in the training;
- the flatness loss function, defined by dividing the $\vec{y}$ space in bins (b):

  $$L_{flat} = \sum_b f_b \int |F_b(s) - F(s)|^p ds,$$

  where $F(s)$ is the cumulative distribution of the score for the events in the training, $F_b(s)$ is the cumulative distribution of the score for the events in the bin b, f_b is the fraction of signal events in the bin and p is a parameter to be decided;
- the loss function:

  $$L_{loss} = L_{ada} + \alpha L_{flat},$$

  where α is a parameter to be decided.

L_{loss} is a function of the event scores s_i. The scores depend from the weights w_j, therefore L_{loss} can be treated as a function of w_j. At each stage of the BDT training the w_j are chosen as the values that minimize L_{loss}. Minimizing L_{ada} provides discrimination, while minimizing L_{flat} provides uniformity with respect to $\vec{y}$. In this analysis the dijet invariant mass m_{jj} is the only observable one wants uncorrelated from the BDT output, therefore $\vec{y}$ is a 1-dimensional space.

4.3.6 *uGB* optimization and training result

In the MVA training MC events of $Z \to b\bar{b}$ are used as signal sample. Data is considered representative of the QCD background, as explained at the beginning of Section 4.3. A certain fraction (f) of data events are randomly extracted from the data sample, used in the training as background sample, and then removed from the measurement to avoid a bias. This fraction has been chosen after some tests in order to not have an overtrained classifier ($f = 10\%$). The observables used in input to the algorithm are six and they are listed in table 4.5. The linear correlations between these observables and the dijet invariant mass are shown in figure 4.22 for the signal and background samples.

Table 4.5: Observables used in the MVA training.

Observables
$\lvert\Delta\eta_{12}\rvert$
$p_T(jet_3)$
$\lvert\Delta\phi_{bb3}\rvert$
$\lvert\Delta\Theta_{bb3}\rvert$
T
S

Two parameters defined in Section 4.3.5 have to be optimized in the algorithm: p and α. The latter represents the ratio between the discriminating power and uniformity of the classifier, as shown in the L_{loss} definition. The values of p and α have been decided after some tests, verifying the classifier discriminating power and uniformity with respect to the dijet mass ($p = 1.5$ and $\alpha = 10$).

From the training samples 30% of the events are randomly selected and used as test samples.

The distributions of the classifier for signal and background are shown in figure 4.23, superimposed with the distributions of the test samples. Since the training and test distributions are compatible within the statistical errors, it is possible to exclude a significant overtraining. The Receiver Operating Characteristic (ROC) curves, where on the x axis there is the signal efficiency and on the y axis the background rejection for different values of the classifier cut, are shown in figure 4.24 for the test and the training

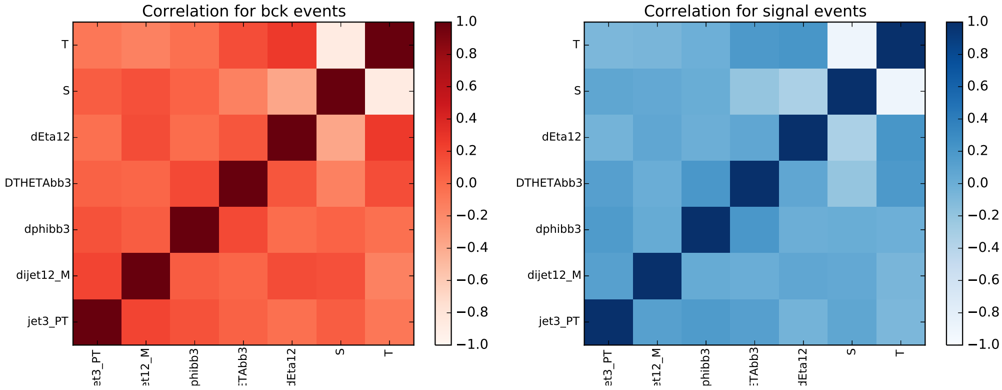

Figure 4.22: Linear correlations between the input observables and the dijet invariant mass.

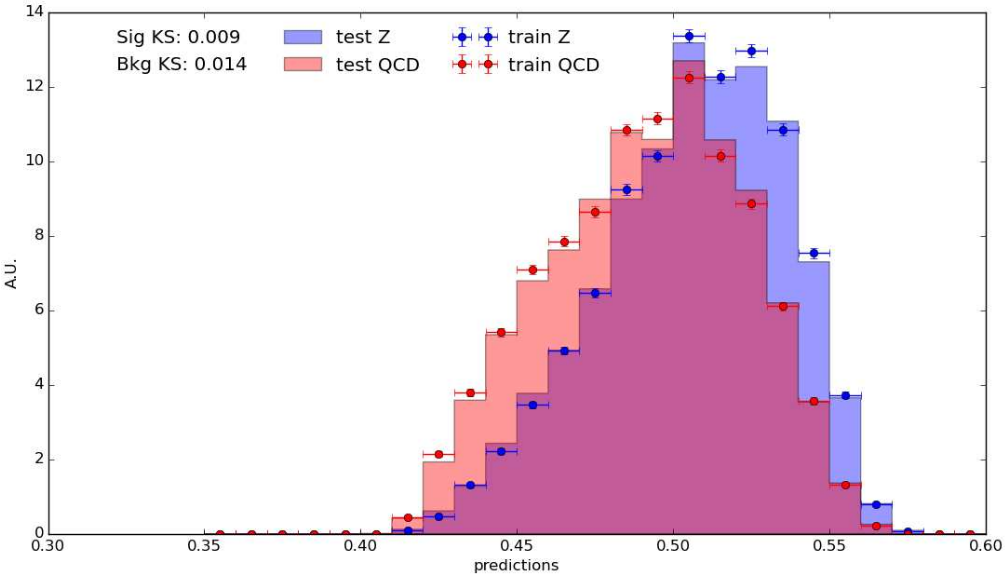

Figure 4.23: Distributions of the classifier obtained from the *uGB* traning, for signal and background. Distributions of the test samples are superimposed.

samples. In the rest of the thesis the output classifier of the training is simply called *uGB*. The distributions of the *uGB* applied to the full data sample and to the $Z \to b\bar{b}$ MC are shown in figure 4.25. It shows that a separation between the signal and the background has been obtained.

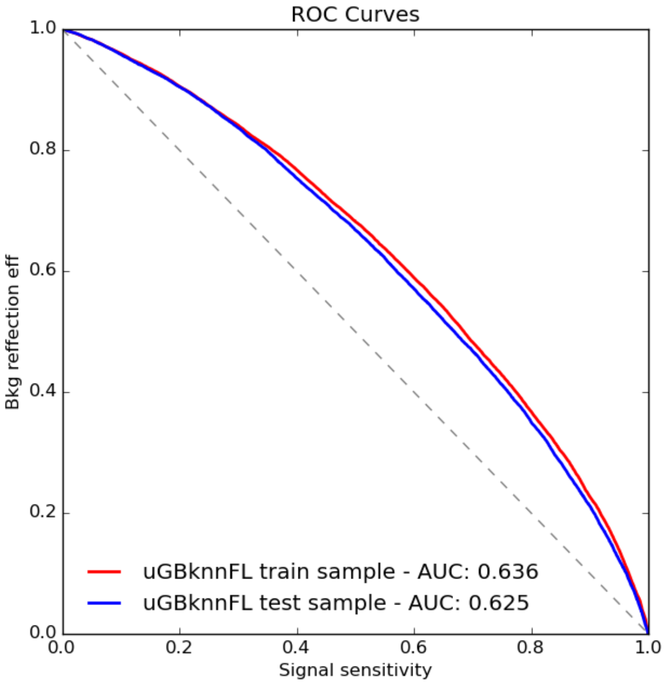

Figure 4.24: ROC curves for the test and training samples.

4.3.7 *uGB* correlations with the dijet invariant mass

Data is used to study the correlation of the *uGB* with the dijet invariant mass in the QCD background. The mean and the RMS of the dijet invariant mass distribution in data as a function of the *uGB* output are shown in figure 4.26. The figure shows that these observables are almost uncorrelated with the *uGB*.

The efficiencies of the *uGB* cut are defined as

$$\epsilon = \frac{N(uGB > x)}{N_{tot}},$$

where x is the value of the *uGB* cut, $N(uGB > x)$ the number of events that satisfy the requirement and N_{tot} is the number of events before the cut. The *uGB* efficiencies are computed for data events within different intervals of dijet invariant mass. The statistical errors are evaluated by using the error formula of a binomial distribution:

$$\Delta\epsilon(m_{jj}) = \sqrt{\frac{\epsilon(m_{jj})[1 - \epsilon(m_{jj})]}{N_{tot}(m_{jj})}},$$

where m_{jj} is the center of the dijet mass interval, $N_{tot}(m_{jj})$ and $\epsilon(m_{jj})$ are respectively the total number of data events in that particular bin and the *uGB* cut efficiency. The

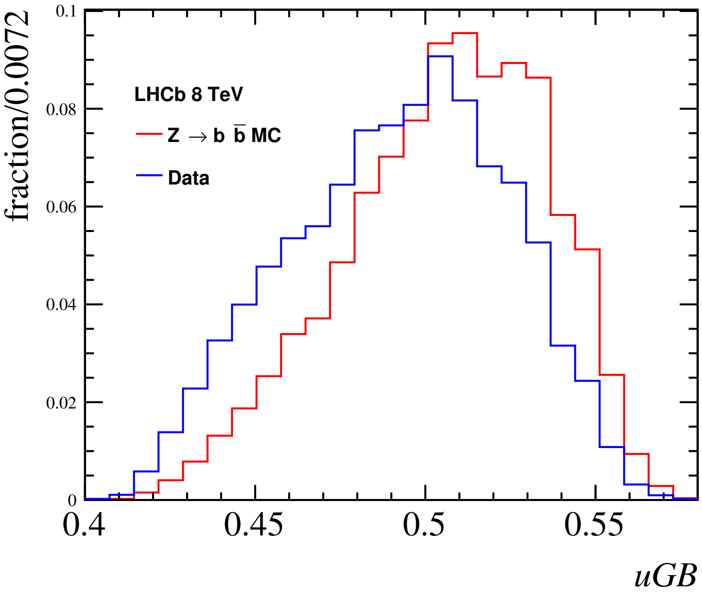

Figure 4.25: Distribution of uGB for data and $Z \to b\bar{b}$ MC events, after the full selection. Only events with a dijet invariant mass in the [65,95] GeV range are shown. The distributions are normalized to unit area.

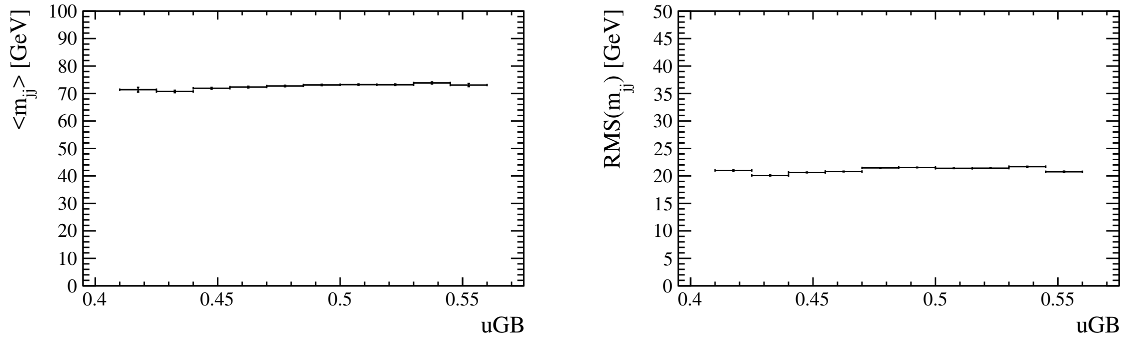

Figure 4.26: Mean and the RMS of the dijet invariant mass distribution in data for different intervals of uGB output.

efficiencies of different uGB cuts as a function of the dijet invariant mass can be seen in figure 4.27. In this study events in the [60,105] GeV interval of dijet mass are removed: there the presence of the signal can modify the QCD background efficiency. In the plot six bins in the [45,60] GeV range and twenty-four bins in the [105,165] GeV are used, with a uniform width of 2.5 GeV. It can be seen that the residual correlation of the uGB with the dijet mass is low, since the efficiencies are almost constant for different mass intervals.

The continuity of the efficiency in the [60,105] GeV mass range can be probed by using

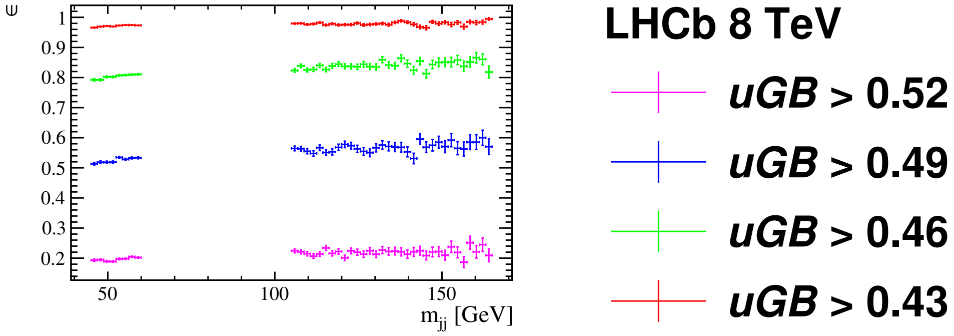

Figure 4.27: Efficiencies of different uGB cuts as a function of the dijet invariant mass computed with the selected data sample. Events in the [60,105] GeV interval of dijet mass are removed.

the QCD MC sample described in Section 4.2.1. Here the simulated processes are quite different from the real case, but no resonant signal is present. The efficiencies of different uGB cuts as a function of the dijet invariant mass for the QCD MC are shown in figure 4.28. This time the points in the [60,105] GeV mass range are shown. The continuity of the efficiency can be asserted within the statistical uncertainties.

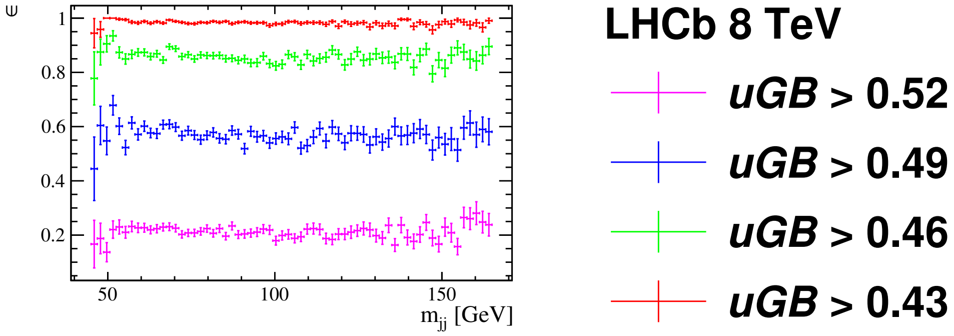

Figure 4.28: Efficiencies of different uGB cuts as a function of the dijet invariant mass computed with the QCD mixture simulation sample.

4.4 Signal Extraction

The MVA discriminator described in Section 4.3.6 is used to define a signal region, optimized for the $Z \to b\bar{b}$ significance, and a control region, orthogonal to the signal region, with low signal contribution. The control region is used to probe the QCD background dijet model.

A simultaneous fit to the dijet invariant mass distributions of signal and control regions is performed to extract the $Z \to b\bar{b}$ yield and the Jet Energy Scale factor. In the fit the dijet mass window [45,165] GeV is used.

The invariant mass distribution of the QCD background can be inferred from that obtained in the control region by applying a correction, called transfer function. This is possible because the uGB dependency on the dijet invariant mass is low and the transfer function has to be applied to correct for the residual correlations.

4.4.1 Definition of signal and control regions

The uGB distributions shown in figure 4.29 are used to define the signal region, with enhanced $Z \to b\bar{b}$ contribution, and the control region, dominated by the QCD background:

- the signal region is formed by the events that have a uGB value above x_s;
- the control region is formed by the events that have a uGB value below x_c.

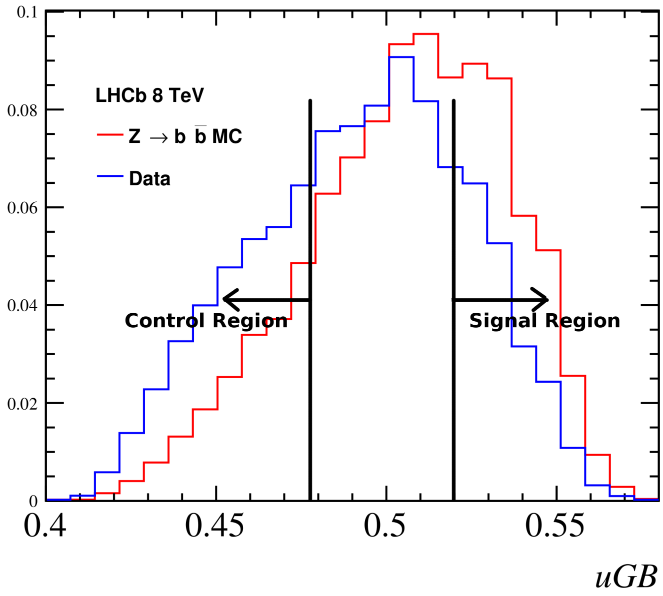

Figure 4.29: uGB distributions for data and $Z \to b\bar{b}$ MC. Signal region $uGB > x_s$ and control region $uGB < x_c$ definitions are shown.

The values x_s and x_c are optimized with the procedure described next. The following quantities are defined:

- N_Z^s: the number of expected $Z \rightarrow b\bar{b}$ events in the signal region, computed using the selection efficiency obtained in MC and the theoretical cross section as explained in Section 4.2.3;
- N_Z^c: the number of expected $Z \rightarrow b\bar{b}$ events in the control region, computed using the selection efficiency obtained in MC and the theoretical cross section as explained in Section 4.2.3;
- N_{TOT}^s: the number of data events in the signal region;
- N_{TOT}^c: the number of data events in the control region;
- the $Z \rightarrow b\bar{b}$ significance in the signal region, $S = \frac{N_Z^s}{\sqrt{N_{TOT}^s}}$;
- the $Z \rightarrow b\bar{b}$ purity in the signal region, $P^s = \frac{N_Z^s}{N_{TOT}^s}$;
- the $Z \rightarrow b\bar{b}$ purity in the control region, $P^c = \frac{N_Z^c}{N_{TOT}^c}$.

The value of x_s is determined by maximizing the following figure of merit:

$$S' = S \cdot P^s.$$

Several values of S' are computed by varying x_s, as shown in figure 4.30. The x_s value that makes S' maximum is $x_s = 0.52$.

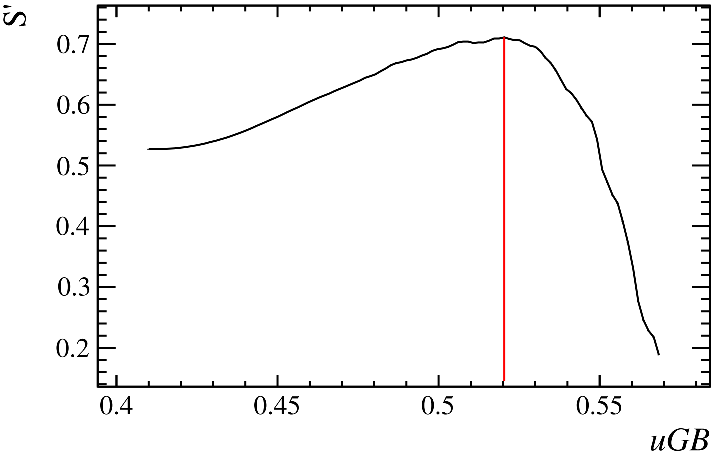

Figure 4.30: Expected S' as a function of x_s, computed as explained in the text.

Given the invariant mass resolution and the high level of $b\bar{b}$ background it is impossible to completely remove the signal from the control region. The signal purity in the control region (P^c) calculated by varying x_c is shown in figure 4.31. It is evident that the signal contribution lowers by reducing the value of x_c. On the other hand lowering x_c the number of data events in the control region decreases. It is possible to define a working point w_p as

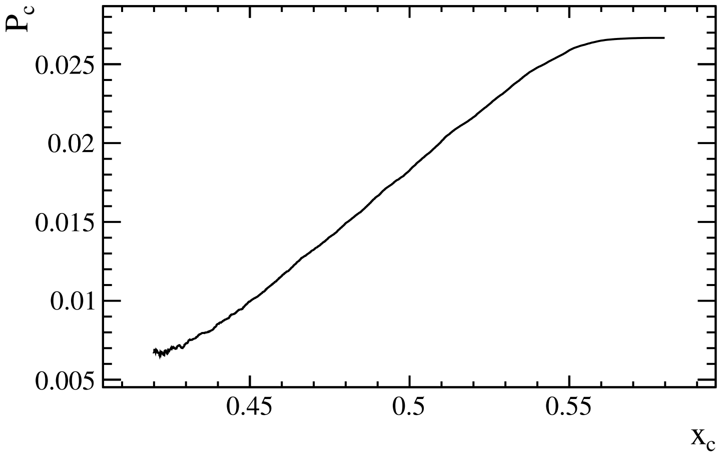

Figure 4.31: Expected signal purity in the control region as a function of x_c, computed as explained in the text.

$$w_p = \frac{N^c_{TOT}}{N^s_{TOT}}.$$

The value of w_p should be greater than 1, in order to have a control sample with a greater number of events with respect to the signal sample. With x_s fixed as explained above, the maximum value of w_p is obtained in the case $x_c = x_s = 0.52$, and it is equal to 3.76. In the measurement w_p is arbitrary set to 1.5, resulting to $x_c = 0.48$. Systematics uncertainties associated to the control region definition are discussed in Section 4.5, where the invariant mass fit is repeated by varying the value of w_p.

In table 4.6 the number of selected data events in signal and control region as well as the number of expected $Z \to b\bar{b}$ and background events are shown.

4.4.2 Signal invariant mass distribution model

The $Z \to b\bar{b}$ dijet invariant mass distribution is studied in MC simulation. The distribution is asymmetric with a tail at high values of mass. A triple gaussian model is used to describe it:

$$Z(m_{jj}) = N_Z[f_1 G(m_{jj};\mu_1,\sigma_1) + f_2 G(m_{jj};\mu_2,\sigma_2) + (1 - f_1 - f_2) G(m_{jj};\mu_3,\sigma_3)],$$

where N_Z is the normalization, $G(m_{jj};\mu_i,\sigma_i)$ are gaussian distributions with mean μ_i and width σ_i, f_i are their relative fractions. MC events with uGB $> x_s$ and uGB $< x_c$ are fitted

Table 4.6: Expected yields of signal and backgrounds processes in the signal and control regions. The yields of selected data events are also reported.

Process	Signal region yield	Control region yield
$Z \to b\bar{b}$	5469	2980
$Z \to c\bar{c}$	177	112
$W \to q\bar{q}'$	48	32
$t\bar{t}$	14	11
Data	116478	175065

separately using $Z(m_{jj})$, to obtain the $Z \to b\bar{b}$ invariant mass templates for the signal and the control regions. The fits are performed using the *RooFit* package of ROOT [79] with the Unbinned Maximum Likelihood technique. In the fits N_Z, f_i, μ_i and σ_i are free parameters. MC events are weighted to take into account of the differences with data in the b-tagging efficiencies and jet identification efficiencies. These corrections have been evaluated in Chapter 3 and they depends on the jets p_T and η, therefore they can affect the invariant mass distribution. The dijet invariant mass distributions of MC events in signal and control regions are shown in figures 4.32 and 4.33, superimposed with the fits results. A χ^2/ndof is computed for each fit result. The χ^2/ndof is equal to 1.37 for the

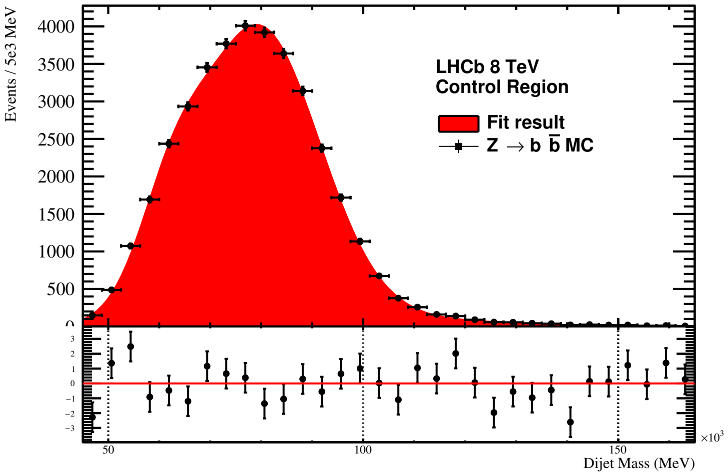

Figure 4.32: Dijet invariant mass distributions of $Z \to b\bar{b}$ MC events in the signal region and fit result. The pulls, displayed below the plot, are the difference between the number of MC events in each bin and the value of the fitted distribution in the center of the bin, divided for the statistical error associated to the number of MC events.

signal region fit, and it is equal to 0.80 for the control region fit. This model describes

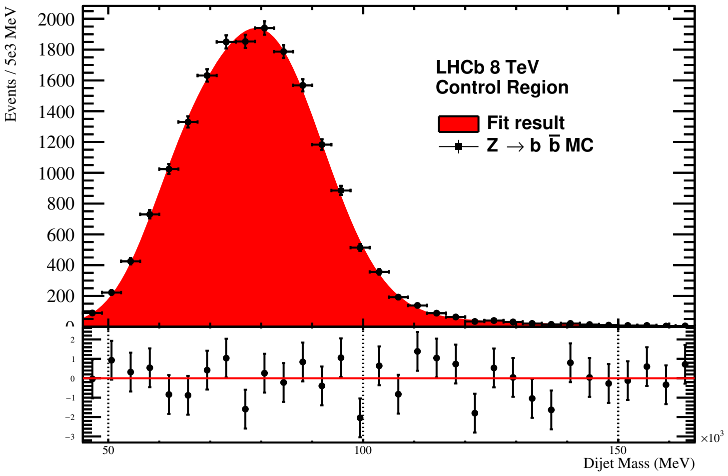

Figure 4.33: Dijet invariant mass distributions of $Z \to b\bar{b}$ MC events in the control region and fit result. The pulls, displayed below the plot, are the difference between the number of MC events in each bin and the value of the fitted distribution in the center of the bin, divided for the statistical error associated to the number of MC events.

well the invariant mass distributions. The values returned by the fits for the parameters and their statistical uncertainties are presented in tables 4.7 and 4.8.

Table 4.7: Value returned by the fit for the parameters and statistical uncertainties obtained for $Z \to b\bar{b}$ dijet invariant mass in the signal region.

Parameter	Fitted Value	Stat. Uncertainty
f_1	0.126	0.022
μ_1	62.14 GeV	0.51 GeV
σ_1	6.49 GeV	0.41 GeV
f_2	0.824	0.026
μ_2	79.48 GeV	0.40 GeV
σ_2	11.95 GeV	0.23 GeV
μ_3	100.0 GeV	4.4 GeV
σ_3	24.5 GeV	1.5 GeV

These templates are used to fit data, with parameters fixed to the MC values, but a Jet Energy Scale parameter is introduced to take into account of possible data/MC differences. The Jet Energy Scale factor k_{JES} is defined as

$$(E, \vec{p})_{data} = k_{JES} \cdot (E, \vec{p})_{MC},$$

where $(E, \vec{p})_{data}$ is the jet four-momentum in data and $(E, \vec{p})_{MC}$ is the jet four-momentum

Table 4.8: Value returned by the fit for the parameters and statistical uncertainties obtained for $Z \to b\bar{b}$ dijet invariant mass in the control region.

Parameter	Fitted Value	Stat. Uncertainty
f_1	0.144	0.032
μ_1	63.88 GeV	0.80 GeV
σ_1	7.78 GeV	0.50 GeV
f_2	0.798	0.037
μ_2	80.37 GeV	0.57 GeV
σ_2	11.81 GeV	0.32 GeV
μ_3	97.92 GeV	7.1 GeV
σ_3	27.2 GeV	2.6 GeV

in simulation. The effect of this transformation on the dijet invariant mass is:

$$(m_{jj})_{data} = k_{JES} \cdot (m_{jj})_{MC}.$$

The transformation maps a gaussian distribution with another gaussian distribution but with different mean and width, related to the original ones through the following expressions:

$$\mu_{data} = k_{JES} \cdot \mu_{MC},$$

$$\sigma_{data} = k_{JES} \cdot \sigma_{MC}.$$

Therefore the signal invariant mass model used in the fit to data is

$$\begin{aligned} Z(m_{jj}; k_{JES}) =& N_Z[f_1 G(m_{jj}; k_{JES}\mu_1, k_{JES}\sigma_1) + f_2 G(m_{jj}; k_{JES}\mu_2, k_{JES}\sigma_2) \\ &+ (1 - f_1 - f_2) G(m_{jj}; k_{JES}\mu_3, k_{JES}\sigma_3)], \end{aligned}$$

where k_{JES} and N_Z are the parameters left free in the fit, while the other are fixed to the values in tables 4.7 and 4.8.

4.4.3 QCD combinatorial background invariant mass model

Data events in the control region are used to verify the invariant mass model of the QCD combinatorial background. This sample is contaminated by the signal, as demonstrated in Section 4.4.1, and this has to be taken into account in the final fit. The combinatorial background mass shape is parametrized with a Pearson IV distribution, typically used to describe processes where a pair of quarks is produced in the hard interaction [71]:

$$Q(m_{jj}; a_1, a_2, a_3, a_4) = N_Q \left[1 + \left(\frac{m_{jj} - a_1}{a_2} \right)^2 \right]^{-a_3} exp \left[-a_4 tan^{-1} \left(\frac{m_{jj} - a_1}{a_2} \right) \right],$$

where a_1, a_2, a_3 and a_4 are the Pearson IV parameters and N_Q is the normalization. $Q(m_{jj})$ is assumed to be the QCD shape before any uGB cut. If the uGB is completely uncorrelated with the dijet invariant mass then $Q(m_{jj})$ can represent the distribution of the QCD background in the signal and control regions. In our case this is not completely true: there is a small correlation between the uGB and the dijet invariant mass, as seen in Section 4.3.7. The correction as a function of the dijet invariant mass can be evaluated from data events in the [45,60] GeV and [105,165] GeV invariant mass intervals, to relate the QCD background shape before the uGB cut to the shape after the uGB cut. This correction is called "transfer function" and it is computed as follow:

1. the efficiency of the uGB $> x_s$ cut is calculated in data for different dijet invariant mass bins, as described in Section 4.3.7. Points in the [60,105] GeV mass range are removed since a not negligible signal contribution is present;

2. the linear transfer function for the signal region is defined by:

$$f^s(m_{jj}) = q_s + h_s \cdot m_{jj},$$

   where q_s and h_s are parameters to be determined. These are obtained with a fit to the efficiency points, as shown in figure 4.34;

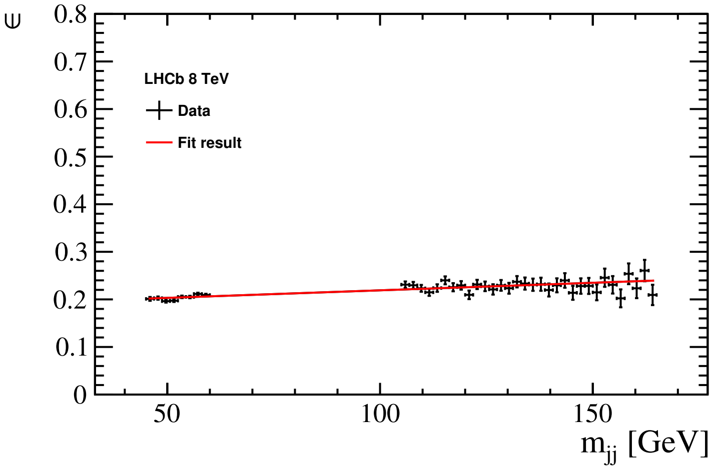

Figure 4.34: Linear fit to the $uGB > x_s$ efficiency as a function of the dijet invariant mass. The efficiency points are computed using data events in the [45,60] GeV and [105,165] GeV sidebands.

3. the efficiency is assumed to be continuous also in the [60,105] invariant mass range, as verified in Section 4.3.7 using the QCD MC. Therefore the transfer function is assumed valid in this invariant mass range;

4. the same procedure is applied to obtain the transfer function for the control region:

$$f^c(m_{jj}) = q_c + h_c \cdot m_{jj},$$

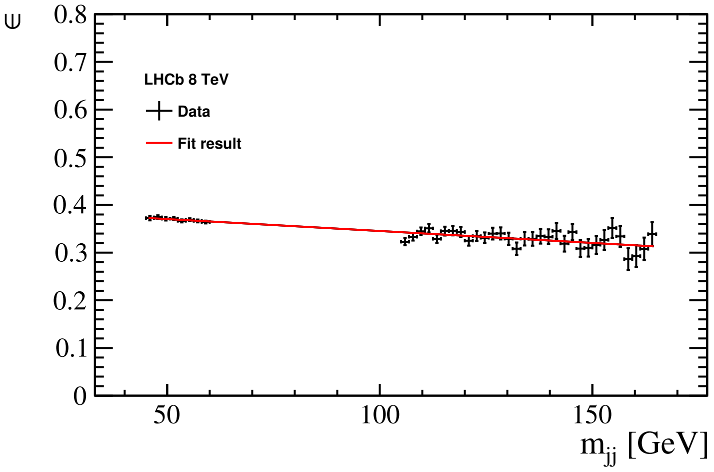

Figure 4.35: Linear fit to the $uGB < x_c$ efficiency as a function of the dijet invariant mass. The efficiency points are computed using data events in the [45,60] GeV and [105,165] GeV sidebands.

where q_c and h_c are parameters obtained from the fit shown in figure 4.35.

The parameters of the transfer functions obtained from the fits are summarized in table 4.9. The QCD invariant mass shape in the signal and control regions is described by the following models:

Table 4.9: Parameters of the linear transfer functions. They are obtained from the fits to the uGB requirements efficiency as a function of the dijet invariant mass.

Parameter	Fitted Value	Stat. Uncertainty
h_s	-0.000317 GeV^{-1}	0.000022 GeV^{-1}
q_s	0.1873	0.0022
h_c	-0.000500 GeV^{-1}	0.000034 GeV^{-1}
q_c	0.3954	0.0026

$$Q^s(m_{jj}; a_1, a_2, a_3, a_4) = Q(m_{jj;a_1,a_2,a_3,a_4}) f^s(m_{jj}) \quad \text{(signal region)},$$

$$Q^c(m_{jj}; a_1, a_2, a_3, a_4) = Q(m_{jj;a_1,a_2,a_3,a_4}) f^c(m_{jj}) \quad \text{(control region)}.$$

To validate the QCD background shape, the invariant mass distribution in the control region is fitted with this model. In this cross-check the $Z \to b\bar{b}$ contamination in the control sample is assumed negligible. The fit result is shown in figure 4.36, with a χ^2/ndof equal to 1.20. The a_1, a_2, a_3 and a_4 values obtained in the fit are used as starting values in the invariant mass fit to data for the extraction of the $Z \to b\bar{b}$ yield. They are reported in table 4.10.

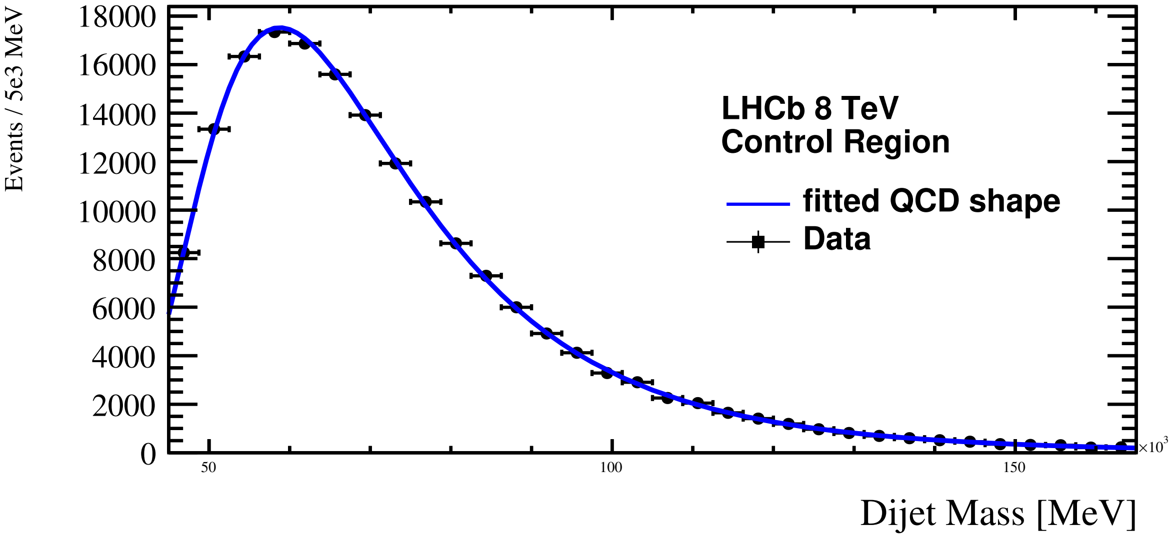

Figure 4.36: Fit to the dijet invariant mass distribution of data events in the control region ($uGB < x_c$). Only the QCD background model is used, signal and other backgrounds contamination are considered negligible.

Table 4.10: QCD mass distribution parameters obtained from the control region fit.

Parameter	Result	Stat. uncertainty
a_1	26.033 GeV	0.357 GeV
a_2	3.769 GeV	0.268 GeV
a_3	3.1410	0.0438
a_4	55.21	4.00

4.4.4 Dijet invariant mass distribution fit to data

The dijet invariant mass distribution of data is simultaneously fitted in the signal and in the control regions, to extract the number of $Z \to b\bar{b}$ events and k_{JES}. The mass window considered in the fit is [45,165] GeV. The invariant mass model used in the fit to describe the signal region is the following:

$$f^s(m_{jj}) = N^s_Q Q^s(m_{jj}; a_1, a_2, a_3, a_4) + N^s_Z Z^s(m_{jj}; k_{JES}) + N^s_Z \cdot k^s_{Zcc} Z^s(m_{jj}; k_{JES}),$$

where N^s_Q and N^s_Z are respectively the number of QCD events and the number of $Z \to b\bar{b}$ events in the signal region, Q^s and Z^s are respectively the QCD and $Z \to b\bar{b}$ models in the signal region, and $N^s_Z \cdot k^s_{Zcc}$ is the number of $Z \to c\bar{c}$ events, with

$$k^s_{Zcc} = \frac{\epsilon^s_{Zcc}}{\epsilon^s_{Zbb}}, \frac{BR(Z \to c\bar{c})}{BR(Z \to b\bar{b})},$$

where ϵ^s_{Zcc} and ϵ^s_{Zbb} are the signal region selection efficiencies respectively for $Z \to c\bar{c}$ and $Z \to b\bar{b}$, $BR(Z \to c\bar{c})$ and $BR(Z \to b\bar{b})$ are the branching ratios. The $Z \to c\bar{c}$ mass shape is assumed to be equal to the $Z \to b\bar{b}$ mass shape. Resonant and quasi-resonant backgrounds other than $Z \to c\bar{c}$ are neglected in this fit. A systematic uncertainty is assigned to this approximation as explained in Section 4.5. The function that describes the invariant mass distribution in the control region is

$$f^c(m_{jj}) = N^c_Q Q^c(m_{jj}; a_1, a_2, a_3, a_4) + R \cdot N^s_Z Z^c(m_{jj}; k_{JES}) + R \cdot N^s_Z \cdot k^c_{Zcc} Z^c(m_{jj}; k_{JES}),$$

where N^Q_c is the number of QCD events in the control region, Q_c and Z_c are respectively the QCD and $Z \to b\bar{b}$ models in the control region, R is the ratio of the number of $Z \to b\bar{b}$ events in the control region to the number of $Z \to b\bar{b}$ events in the signal region and $R \cdot N^s_Z \cdot k^c_{Zcc}$ is the number of $Z \to c\bar{c}$ events with

$$k^c_{Zcc} = \frac{\epsilon^c_{Zcc}}{\epsilon^c_{Zbb}}, \frac{BR(Z \to c\bar{c})}{BR(Z \to b\bar{b})},$$

where ϵ^c_{Zcc} and ϵ^c_{Zbb} are the control region selection efficiencies respectively for $Z \to c\bar{c}$ and $Z \to b\bar{b}$. The fit is performed with the *RooFit* package with the Unbinned Maximum Likelihood technique. N^Q_s, N^Q_c, a_1, a_2, a_3, a_4, N^Z_s and k_{JES} are free parameters. R is fixed to 0.545, obtained in MC with the formula:

$$R = \frac{N^Z(uGB < x_c)}{N^Z(uGB > x_s)},$$

where $N^Z(uGB < x_c)$ is the number of events that pass the $uGB < x_c$ cut and $N^Z(uGB > x_s)$ is the number of events that pass the $uGB > x_s$ cut. The agreement between data and simulation on R depends on how well the uGB is simulated. Moreover R has low correlation with the Z boson cross section, because it cancels out in the ratio, and it has low correlation with k_{JES} for uGB construction. Anyway a systematic uncertainty associated to the choice of fixing its value is considered, as described in section 4.5. k^s_{Zcc} and k^c_{Zcc} are fixed to the values obtained with the MC efficiencies and the branching ratio in [21].

4.4.5 Test of the fitting procedure

To verify the fit stability and that it does not introduce bias in the result, fits to pseudo-experiments are performed. The pseudo-experiments are generated in the signal and in the control regions, according to the QCD background invariant mass model obtained in figure 4.36 plus a $Z \to b\bar{b}$ contribution, with yield fixed to the expected value and distribution taken from the $Z \to b\bar{b}$ MC (k_{JES} is set to 1). An example of toy fit to a set of pseudo-experiments is shown in figure 4.37. 10000 set of toy data samples are generated and for each of them a fit is performed.

For each free parameter the following quantities are evaluated:

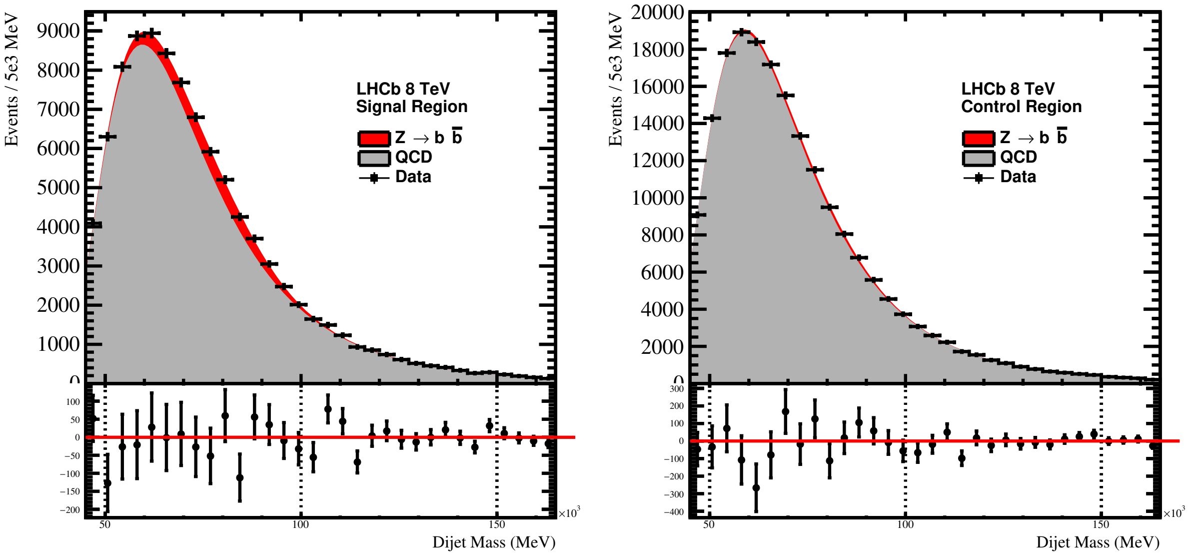

Figure 4.37: Result of a simultaneous dijet invariant mass fit to a toy data sample.

- residual: difference between the value used to generate the pseudo-experiments and the fit result;
- pull: the residual divided by the statistical uncertainty obtained in the fit.

Their distributions are fitted with a Gaussian model. The distributions of residuals and pulls for the QCD model parameters a_1, a_2, a_3 and a_4 are shown in figures 4.38 and 4.39 respectively. It can be seen that no significant bias is present. The distributions of residuals and pulls for the Jet Energy Scale k_{JES} and the signal yield N_Z^s are shown in figures 4.40 and 4.41 respectively. Only a small bias can be found in the signal yield result, of about -90 events, which is negligible with respect to the statistical error..

4.4.6 Fit result

Data are fitted by using the procedure described in section 4.4.4. The fit converges and the error matrix is positive-definite. In figure 4.42 the fit result superimposed to data in the signal and in the control regions is presented. χ^2/ndof are computed separately in the signal and control regions: they are equal to 0.75 and 1.26 respectively. The parameters returned by the fit are summarized in table 4.11 with their statistical uncertainties. The $Z \to b\bar{b}$ yield has been determined with a statistical uncertainty of about 15% while k_{JES} has a statistical uncertainty of about 1.7%. The yield is 4759 ± 730 which is compatible, within its uncertainty, with the prediction of 5469. The significance of the extracted signal peak can be computed with the formula (Wilk's theorem)

$$S = \sqrt{2\Delta LL_{min}},$$

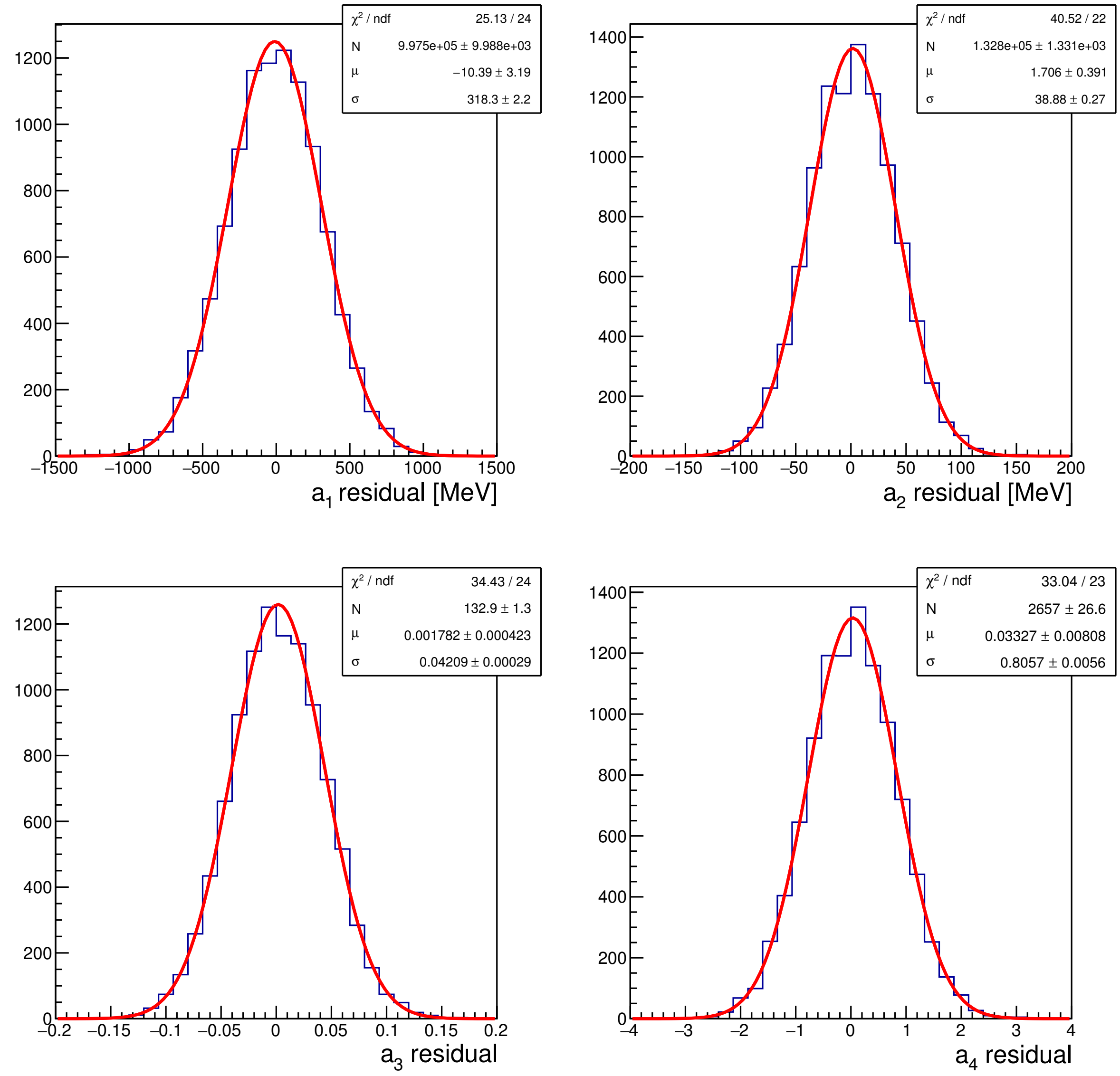

Figure 4.38: Distributions of residuals for the QCD model parameters a_1, a_2, a_3 and a_4, fitted with gaussian functions.

where ΔLL_{min} is the difference between the minimum of the log-likelihood when the signal is removed from the fitting model and the minimum of the log-likelihood when the signal is included. In the fits employed in the computation of ΔLL_{min} the k_{JES} factor is fixed to 1. The significance obtained in this way is 6.6. In figure 4.43 the backgrounds subtracted data, together with the result of the fit is shown.

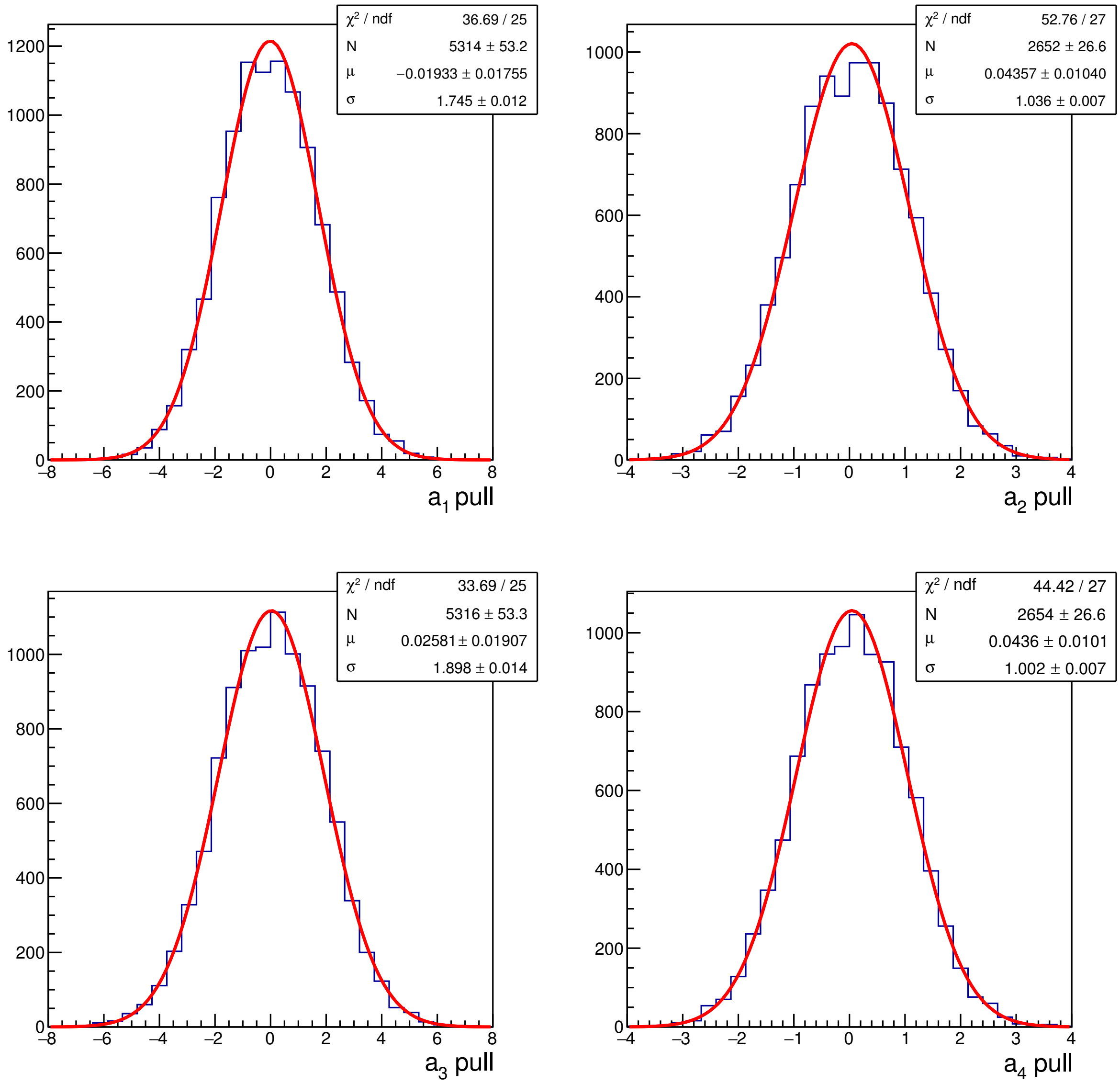

Figure 4.39: Distributions of pulls for the QCD model parameters a_1, a_2, a_3 and a_4, fitted with gaussian functions.

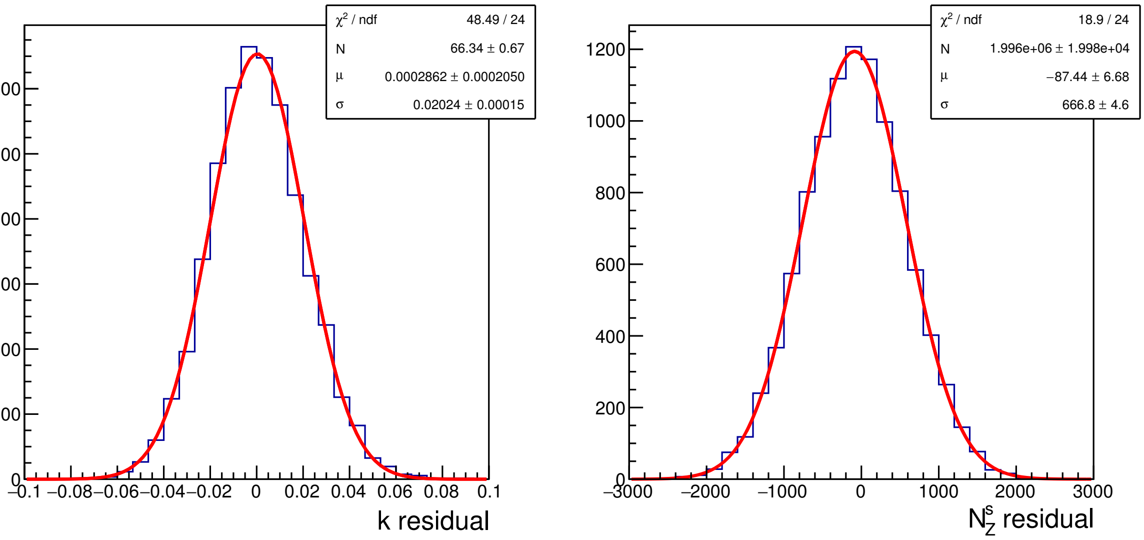

Figure 4.40: Distributions of residuals for the QCD model parameters a_1, a_2, a_3 and a_4, fitted with gaussian functions.

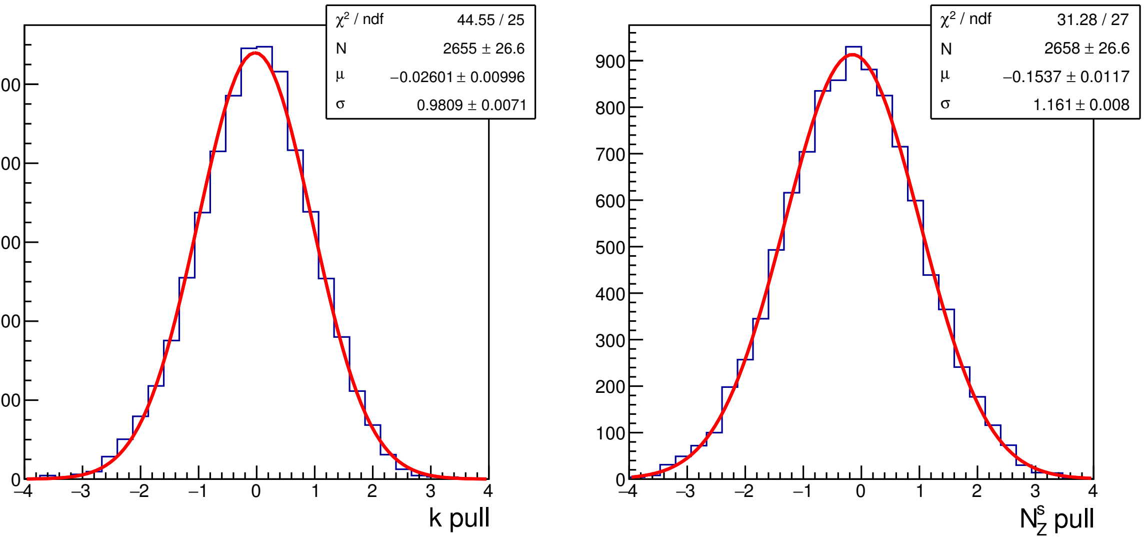

Figure 4.41: Distributions of pulls for the Jet Energy Scale k_{JES} and the signal yield in the signal region N_Z^s, fitted with gaussian functions.

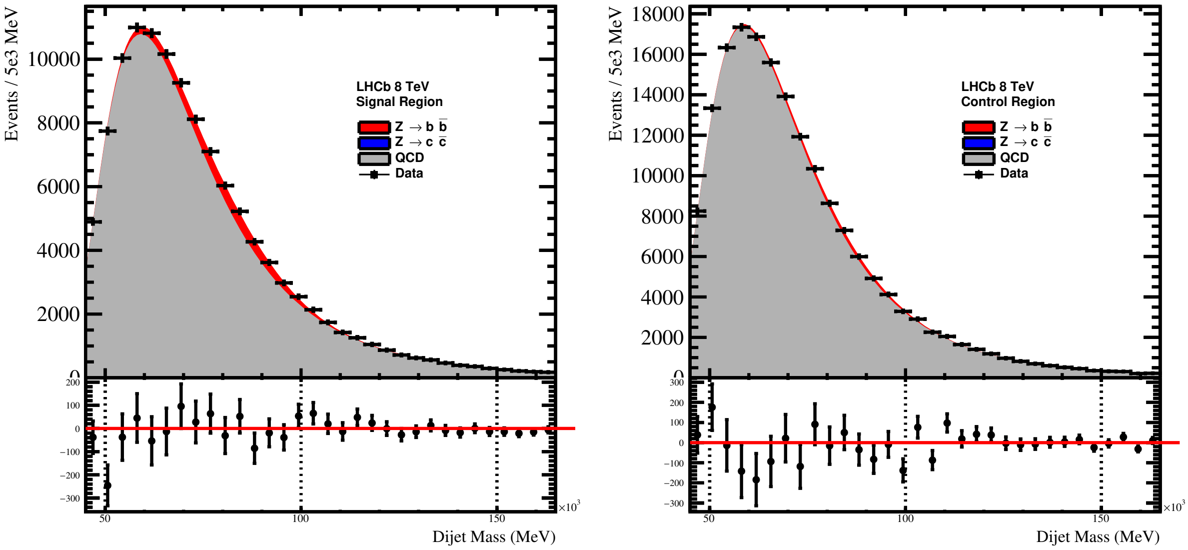

Figure 4.42: Result of the simultaneous dijet invariant mass fit to data superimposed with the data distribution in the signal and control regions. Residuals, obtained as the difference between the number of data events in each mass interval and the value of the fitted model, are displayed below the plot.

Table 4.11: Parameters obtained from the dijet invariant mass fit.

Parameter	Result	Stat. uncertainty
a_1	26.324 GeV	0.325 GeV
a_2	3.218 GeV	0.160 GeV
a_3	3.1550	0.0430
a_4	63.85	2.58
N_Q^c	172723	534
N_Q^s	111553	826
N_Z^s	4759	7308
k_{JES}	1.0194	0.0176

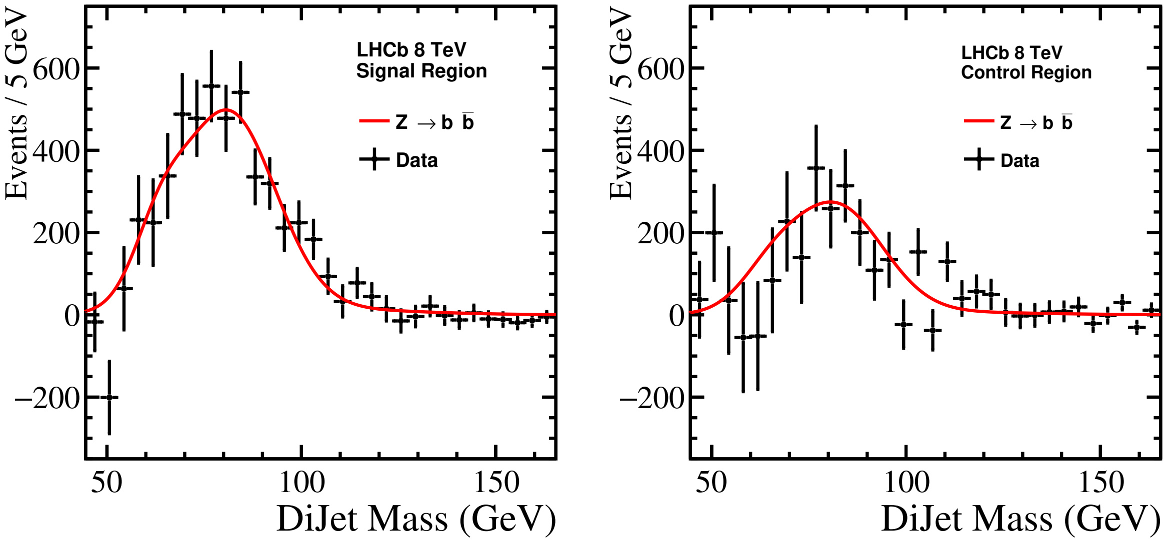

Figure 4.43: Backgrounds subtracted data events compared with the $Z \to b\bar{b}$ mass model. Statistical uncertainties are shown.

4.5 Systematic uncertainties

4.5.1 $Z \to b\bar{b}$ cross section determination and systematics

The $Z \to b\bar{b}$ cross section in the forward region is calculated using the formula:

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = \frac{N_Z^s}{\mathcal{L} \cdot \epsilon_Z^s \cdot (1-f)},$$

where N_Z^s is the signal yield, $\mathcal{L}$ the integrated luminosity, ϵ_Z^s the signal selection efficiency, f is the fraction of data events removed for the *uGB* training (10%). ϵ_Z^s is determined as explained in Section 4.2.3 and takes into account also the signal acceptance defined by the request that the two b quarks from the Z boson decay are reconstructed in the $2 < \eta < 5$ pseudo-rapidity range.

In this chapter the systematic uncertainties in the $Z \to b\bar{b}$ cross section and in the JES measurements are studied. Some of them are related to data/MC differences and can modify the signal template and ϵ_Z^s, which are determined using MC. Other sources are related to the signal extraction procedure. The list of systematics sources considered in this analysis is in table 4.12. The systematics may affect N_Z^s, k_{JES} or ϵ_Z^s as indicated in the table.

Table 4.12: List of systematics uncertainties studied in the $Z \to b\bar{b}$ analysis. The table indicates with a X if the systematic source affect the signal yield, k_{JES} or the signal efficiency

Systematic source	N_Z	k_{JES}	ϵ_Z
Jet b-tagging	X	X	X
Jet Identification	X	X	X
Jet Resolution	X	X	X
Control region bias	X	X	
R	X	X	
$Z \to c\bar{c}$ yield	X	X	
Sub-dominant backgrounds	X	X	
Luminosity			

4.5.2 Jet b-tagging

Data/MC corrections for jet b-tagging efficiencies are applied to MC events in the evaluation of the signal template and of ϵ_Z^s. These corrections and corresponding uncertainties are measured as explained in Section 3.3.3. Two weights are applied, one for the jet tagged at trigger level (jet$_1$), and the other for the jet tagged with the SV algorithm (jet$_2$). Moreover these corrections depend on the jet p_T. The uncertainties varies from 5% to 10% on most of the jets.

The impact on the measurements is evaluated as follow:

1. a signal template is calculated as explained in Section 4.4.2 but applying weights corresponding to the upper variation of their uncertainties;
2. the fit to the dijet invariant mass described in Section 4.4.4 is performed using the template of point 1 as signal model. The Z yield and JES obtained in this fit are labeled as N_Z^{tag-up} and k_{JES}^{tag-up};
3. points 1 and 2 are repeated applying weights corresponding to the lower variation of the uncertainty. N_Z^{tag-dw} and k_{JES}^{tag-dw} are obtained;
4. the maximum between $|k_{JES} - k_{JES}^{tag-up}|$ and $|k_{JES} - k_{JES}^{tag-dw}|$ is taken as systematic uncertainty for the JES;
5. the efficiency calculation is repeated by applying weights to MC events according to the upper (lower) variation of their uncertainty obtaining ϵ_Z^{tag-up} (ϵ_Z^{tag-dw});
6. the Z cross section is calculated using N_Z^{tag-up} and ϵ_Z^{tag-up} (σ_Z^{tag-up}) and using N_Z^{tag-dw} and ϵ_Z^{tag-dw} (σ_Z^{tag-dw});
7. the uncertainty on σ_Z is calculated as $max(|\sigma_Z - \sigma_Z^{tag-up}|, |\sigma_Z - \sigma_Z^{tag-dw}|)$.

4.5.3 Jet Identification

Data/MC corrections for jet identification efficiencies are applied to MC events in the evaluation of the signal templates and of ϵ_Z^s. These corrections and their uncertainties are obtained as explained in 3.2.6. Three weights are applied, one for each of the two b-jets and one for the third jet. The uncertainty is found to be below 2% for jets with p_T greater than 10 GeV.

The impact on the measurements is calculated as for the b-tagging uncertainty. A signal template is built for the upper (lower) variation of the weights, then the fit is repeated to obtain the upper (lower) values of N_Z^s and k_{JES}. The maximum deviation from the nominal k_{JES} is taken as systematic uncertainty for the JES. The impact on ϵ_Z^s is obtained in a similar way: in the efficiency calculation weights to MC events are applied according to the upper (lower) variation of their uncertainty. The upper (lower) Z yield and efficiency are used to compute the upper (lower) value of the cross section. The maximum deviation from the nominal measurement is taken as systematic.

4.5.4 Jet energy resolution

The uncertainty associated to data/MC differences in the jet resolution affects the signal shape and the signal efficiency. This is evaluated as explained in [77]: in this study events of a $Z \rightarrow \mu\mu$+jet data sample are used to evaluate the maximum smearing one need to apply to the jet p_T in MC to have a an agreement with data less than 1σ. This smearing is found to be around 10%. To determine the impact of this uncertainty on the fit the p_T of the two b-jets and of the recoil jet is smeared in the $Z \to b\bar{b}$ MC, before applying the selection. A signal template is evaluated with this sample and the fit is repeated, obtaining N_Z^{JER} and k_{JES}^{JER}. The difference between k_{JES}^{JER} and the nominal k_{JES} is taken as uncertainty for the JES. In the same way the efficiency ϵ_Z^{JER} is calculated by smearing the p_T of jets in MC, before evaluating the number of events that pass the selection. The variation of the cross section calculated using the N_Z^{JER} and ϵ_Z^{JER} with respect to the nominal measurement is taken as uncertainty.

4.5.5 Jet Energy Correction

The uncertainty due to the Jet Energy Correction is explained in Section 3.2.5. The JEC for the upper (lower) variation of its uncertainty is applied to jets in $Z \to b\bar{b}$ before applying the selection. The impact on the measurement is calculated as for the b-tagging uncertainty: signal templates for the upper (lower) JEC uncertainty are built and the invariant mass fit is repeated. The variation of ϵ_Z^s is calculated in an analogous way.

4.5.6 Control region bias

The control region definition ($uGB < x_c$) can introduce a bias in the fit result. In fact it is possible that the control sample obtained in this way does not properly reproduce the QCD shape in the signal region, although many cross-checks have been performed. To evaluate this bias the control region definition is varied between two working points ($w_p = 1.0$ and $w_p = 2.0$, defined in Section 4.4.1). For each w_P a new transfer function $f^c(m_{jj})$ is evaluated and the dijet invariant mass fit is repeated. The maximum deviation of k_{JES} and N_Z^s from the nominal result is taken as systematic uncertainty.

4.5.7 Uncertainty on R

The parameter R, defined in Section 4.4.4, is fixed in the fit to the value obtained from the simulation. The observables in input to the *uGB* classifier have been cross-checked with the $Z \to \mu\mu$ + jet data sample, therefore a significant deviation from the real value is not expected. To estimate the uncertainty on the fit, R is varied in the ±15% range (arbitrary chosen), and the maximum deviation of k_{JES} and N_Z^s from the nominal result is taken as systematic uncertainty.

4.5.8 $Z \to c\bar{c}$ yield

The k_{Zcc}^s and k_{Zcc}^c values are fixed as explained in Section 4.4.4. The uncertainty on these factors is dominated by the uncertainty on the $Z \to c\bar{c}$ branching fraction used in their evaluation [21]. The impact on the measurement is obtained repeating the fit by varying tk_{Zcc}^s and k_{Zcc}^c, at the same time and of the same fraction, in the ±4% range, the maximum deviation from the nominal result is taken as uncertainty.

4.5.9 Sub-dominant backgrounds

Resonant and quasi-resonant backgrounds other than $Z \to c\bar{c}$ have been considered negligible in the dijet invariant mass fit. The most significant backgrounds neglected are $W \to q\bar{q}'$ and $t\bar{t}$. To evaluate their impact, the invariant mass fit is repeated by introducing them in the invariant mass model as fixed contributions. The yields are fixed to the expectation while the shape are obtained applying the kernel density estimation *RooKeysPdf* [79] to MC events that pass the selection, before the *uGB* cuts. The *pdf* obtained for the signal region are shown in figures 4.44 and 4.45. The deviation of the fit results from the nominal result is taken as systematic uncertainty.

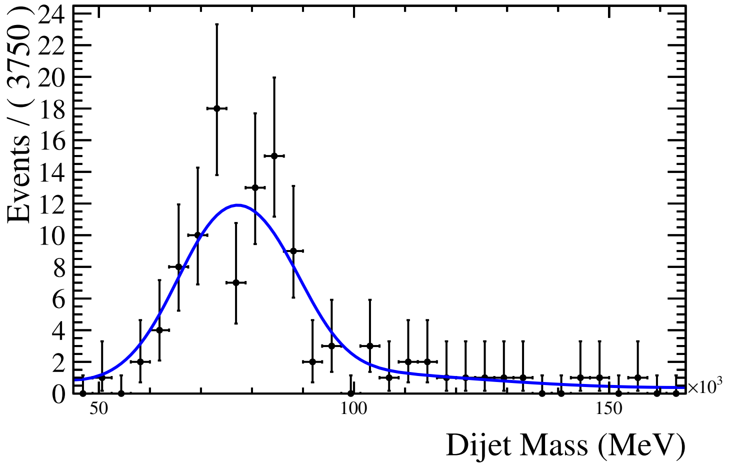

Figure 4.44: Kernel density estimation of the $W \to q\bar{q}'$ dijet invariant mass distribution, obtained using MC events.

4.5.10 Luminosity

The systematic uncertainty on the integrated luminosity is of about 1% as determined in [73]. It is propagated to the cross section measurement using the cross section formula.

4.5.11 Summary of systematics uncertainties

The systematics contributions from each source are shown in table 4.13. Since the different sources are considered uncorrelated, the total uncertainties is obtained by summing them in quadrature. N_Z^s and ϵ_Z^s uncertainties are then propagated through the cross section formula. The dominant systematic uncertainties are those related to the b-tagging and to the control region bias.

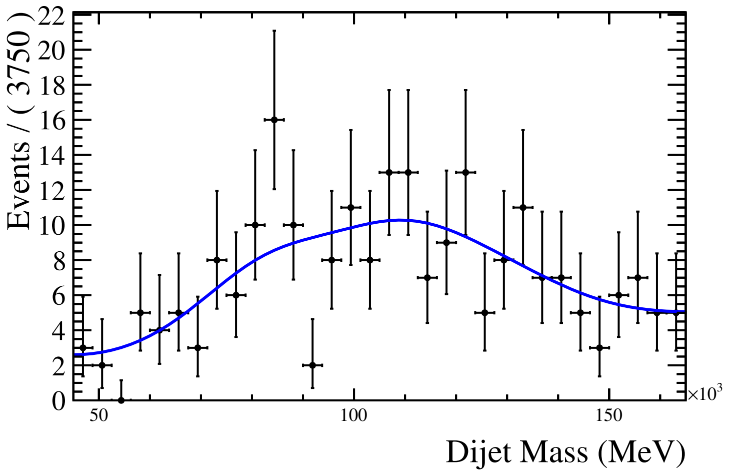

Figure 4.45: Kernel density estimation of the $t\bar{t}$ dijet invariant mass distribution, obtained using MC events.

Table 4.13: Systematics sources contributions to the measurements. The total systematics uncertainty is also reported.

Systematic source	$\frac{\Delta\sigma_Z}{\sigma_Z}$	$\frac{\Delta k_{JES}}{k_{JES}}$
Jet b-tagging	14.9%	0.824%
Jet Identification	2.72%	0.0589%
Jet Resolution	1.82%	0.255%
Jet Energy Correction	2.45%	0.269%
Control region bias	13.3%	0.834%
R	0.922%	0.134%
$Z \to c\bar{c}$ yield	0.210%	0.181%
Sub-dominant backgrounds	1.15%	0.0598%
Total	20.5%	1.42%

4.6 Results

The measured Jet Energy Scale factor is

$$k_{JES} = 1.019 \pm 0.018(\text{stat.}) \pm 0.014(\text{syst.}),$$

where the first uncertainty is statistical and the second systematic. The dominant uncertainty is the statistical one. The k_{JES} factor is compatible with 1: it is possible to say that the jet energy scale in simulation is compatible with the jet energy scale in data. The $Z \to b\bar{b}$ cross section in pp collisions at an energy of 8 TeV in the center-of-mass frame, with the two b quarks in the LHCb forward acceptance ($2 < \eta < 5$) is

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 712 \pm 108(\text{stat.}) \pm 146(\text{syst.}) \pm 7(\text{lum.}),$$

where the first uncertainty is statistical, the second systematic and the third is related to the integrated luminosity measurement. The cross section measurement is dominated by the systematic uncertainty and it is compatible, within the total uncertainty, with the theoretical prediction of 817 pb (table 4.3).

Chapter 5

Search for the Higgs in association with a W or Z boson

5.1 Introduction

The SM $H^0 \to b\bar{b}$ can be searched exploiting the production in association with a W or Z boson (V) that decays leptonically. The presence of an isolated high energy lepton reduces drastically the QCD background with respect to the inclusive search, where only two heavy-flavour tagged jets are required. In this Chapter the dataset collected by LHCb with the high p_T lepton (muon or electron) trigger during the 2012 period is analysed to set upper limits on $\sigma(V + H^0)BR(H^0 \to b\bar{b})$ and $\sigma(V + H^0)BR(H^0 \to c\bar{c})$. The main physical backgrounds in the search are the $W + b\bar{b}$ and $W + c\bar{c}$ processes, where the W boson decays in $W \to \ell\nu_\ell$ and the $t\bar{t}$ process, where the top decays in $t \to W(\ell + \nu_\ell) + b$. In this Chapter the symbol ℓ indicates a muon or an electron. Figures 5.1 and 5.2 show the Feynman diagrams of these processes. The analysis proceeds through the following steps:

1. a lepton+dijet selection is applied to the dataset collected in 2012, requiring two reconstructed jets identified as originating from heavy flavour quarks by the SV tagging algorithm, labeled as j_1 and j_2, and a lepton (electron or muon) that satisfies the high p_T lepton trigger requirements, labeled as ℓ;
2. the MC simulation is used to reproduce the SM backgrounds and it is compared with data;
3. a multivariate technique (MVA) is employed to build three different discriminators: one to separate $t\bar{t}$ from $W + b\bar{b}$, called $uGB(W + b\bar{b} + t\bar{t})$, one to separate $V + H^0(\to b\bar{b})$ from $W + b\bar{b}$, called $uGB(W/Z + H^0 \text{ vs } W + b\bar{b})$, and one to separate $V + H^0(\to b\bar{b})$ from $t\bar{t}$, called $uGB(W/Z + H^0 \text{ vs } t\bar{t})$;
4. a fit to the dijet invariant mass (m_{jj}), $uGB(W + b\bar{b} + t\bar{t})$, $\text{BDT}_{b|c}(j_1)$ and $\text{BDT}_{b|c}(j_2)$ ($\text{BDT}_{b|c}$ definition in Section 3.3.1) distributions is performed to measure the $t\bar{t}$,

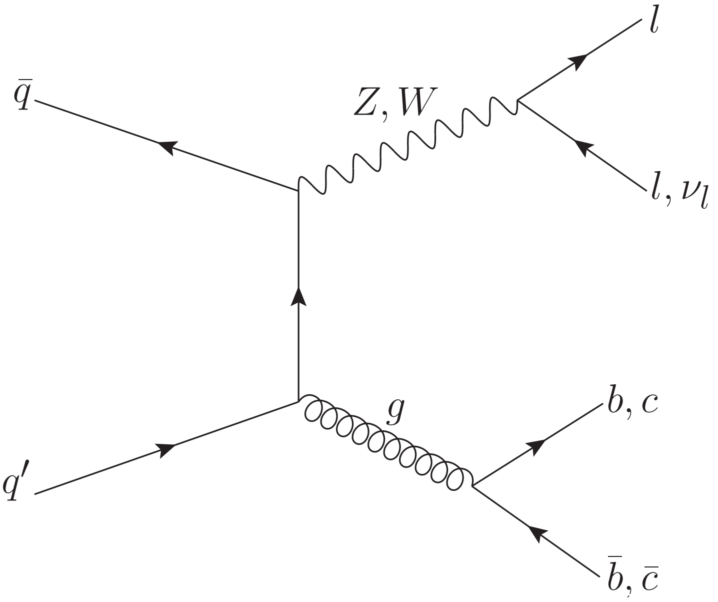

Figure 5.1: Feynman diagram for the production of $W + b\bar{b}$ and $W + c\bar{c}$.

$W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$. The yields are used to determine the $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ cross sections. The measurements are

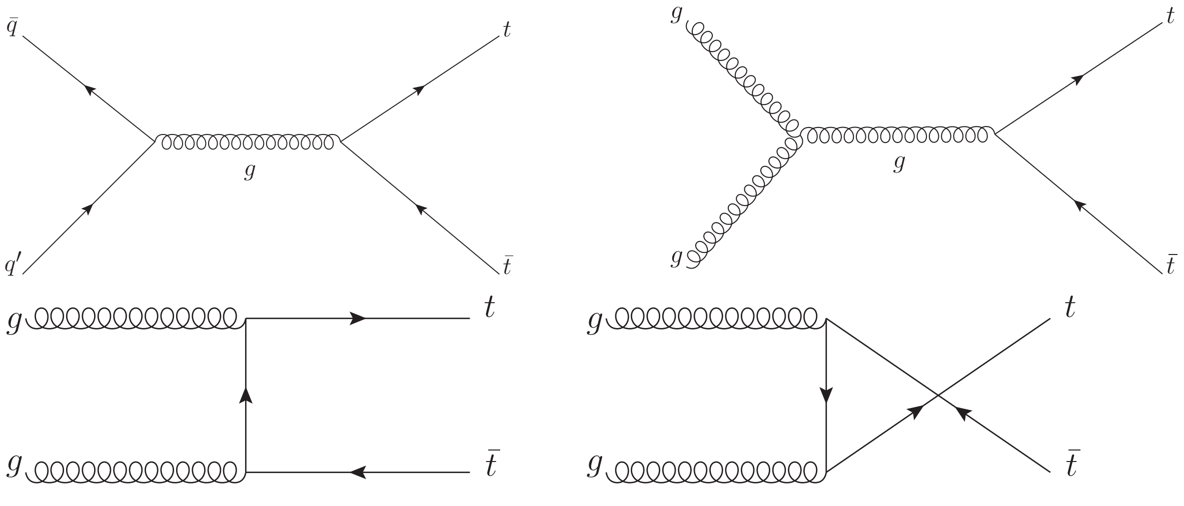

Figure 5.2: Feynman diagrams for the production of $t\bar{t}$ in $q\bar{q}$ (top left) and gg (top right and bottom right and left) production modes.

compared to the cross sections predicted by the SM.

5. the CL_s technique is used to set the upper limit to $\sigma(V + H^0)BR(H^0 \to b\bar{b})$ and $\sigma(V+H^0)BR(H^0 \to c\bar{c})$, where the Higgs has a mass of 125 GeV[1] and SM properties. The limit is calculated using the distributions of m_{jj}, $uGB(W/Z + H^0$ vs $W + b\bar{b})$, $uGB(W/Z + H^0$ vs $t\bar{t})$ and the lepton flavour.

5.2 Expected physical backgrounds

$W + b\bar{b}$, $W + c\bar{c}$ process, with $W \to \ell\nu_\ell$, and $t\bar{t}$, with $t \to W(\ell + \nu_\ell) + b$, are expected to be the main backgrounds in the $V + H^0(\to b\bar{b})$ search. Moreover the measurements of their production cross sections is of physical interest. In fact the $W + b\bar{b}$ and $W + c\bar{c}$ cross sections in the forward region provide experimental test of perturbative Quantum Chromodynamics (pQCD) [80–82] at a complementary phase space region to ATLAS and CMS. Previous studies of the $W + b\bar{b}$ final state have been performed by ATLAS [83] and CMS [84, 85] at centre-of-mass energies $\sqrt{s} = 7\,\text{TeV}$ and $8\,\text{TeV}$. The study of $W + c\bar{c}$ is novel, with no previous measurements of this cross-section available. Forward $t\bar{t}$ events can be used to constrain the gluon Parton Distribution Function (PDF) at large momentum fraction [86].

Other processes that contribute to the background for the $V + (H^0 \to b\bar{b})$ search, but with lower yields with respect to $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$, as demonstrated in the following Sections, are:

- single t production;
- $Z(\to \ell\ell) + b\bar{b}$;
- $Z(\to \ell\ell) + c\bar{c}$;
- $Z(\to \ell\ell) + Z(\to b\bar{b})$;
- $W(\to \ell\nu_\ell) + Z(\to b\bar{b})$;
- $W(\to \tau\nu_\tau) + b\bar{b}$;
- $W(\to \ell\nu_\ell)$+jets.

---

[1] in this Chapter natural units where $\hbar = c = 1$ are used.

5.3 Lepton + dijet selection

In this Section the selection of events with a high energy lepton (muon or electron) and a pair of jets generated by heavy flavour quarks (b or c) is described.

The muon track is selected at trigger level (Trigger On Signal, TOS) with the following requirements:

- the track must pass the *L0Muon* trigger line defined in Section 2.11.1;
- the track must pass the *Hlt1SingleMuonHighPT* trigger line, which requires a track identified as a muon with $p > 8$ GeV, $p_\mathrm{T} > 4.8$ GeV and $\chi^2/\mathrm{ndof} < 4$, where the χ^2 is that obtained in the track fit;
- the track must pass the *Hlt2SingleMuonHighPT* trigger line, which requires a track that pass the *Hlt1SingleMuonHighPT* and $p_\mathrm{T} > 10$ GeV.

The following requirements are applied to ensure the quality of the muon track:

- $\chi^2/\mathrm{ndof} < 2.5$, where the χ^2 is that obtained in the track fit;
- the p-value of the χ^2 obtained in the track fit, $P(\chi^2, \mathrm{ndof})$, must be greater than 0.01;
- the number of hits in the TT station must be greater than 0;
- the momentum resolution $\Delta p/p$ must be less than 10%.

Moreover the muon must satisfy the following requirements:

- $p_\mathrm{T}(\mu) > 20$ GeV, to remove the background from low energy muons, not coming from a vector boson decay;
- $2 < \eta(\mu) < 4.5$;
- $\frac{E_{ECAL}+E_{HCAL}}{p}$ must be less then 0.04, where E_{ECAL} and E_{HCAL} are the energy released by the track in the ECAL and HCAL respectively, to remove the background from hadrons that do not stop in the calorimeter;
- the muon impact parameter (IP) must be less than 0.04 mm to reject muons from heavy flavour hadrons decays;
- a requirement on the muon isolation I is applied; the isolation is defined as $I = \frac{p_\mathrm{T}(\mu)}{p_\mathrm{T}(j_\mu)}$, where $p_\mathrm{T}(j_\mu)$ is the transverse momentum of the jet containing the muon, built using the jets reconstruction described in Section 3.2. The isolation must be greater than 0.8 to remove muons not coming from a vector boson decay.

The electron selection is explained next. The electron track is selected at trigger level with the following requirements:

- the track must pass the *L0Electron* trigger line defined in Section 2.11.1;
- the track must pass the *Hlt1SingleElectronNoIP* trigger line, which requires a track identified as an electron with $p > 8$ GeV, $p_{\mathrm{T}} > 4.8$ GeV and $\chi^2/\mathrm{ndof} < 4$, where the χ^2 is that obtained in the track fit;
- the track must pass the *Hlt2SingleTFVHighPtElectron* trigger line, which requires a track that pass the *Hlt1SingleElectronNoIP*, $p_{\mathrm{T}} > 15$ GeV, $E_{\mathrm{ECAL}}/p > 0.1$, $E_{\mathrm{HCAL}}/p < 0.05$ and a number of hits in the Preshower detector greater than 50.

The following requirements are applied to ensure the quality of the electron track:

- $\chi^2/\mathrm{ndof} < 2.5$, where the χ^2 is that obtained in the track fit;
- the p-value of the χ^2 obtained in the track fit, $P(\chi^2, \mathrm{ndof})$, must be greater than 0.01;
- the number of hits in the TT station must be greater than 0;
- the momentum resolution $\Delta p/p$ must be less than 10%.

Moreover the electron must satisfy the following requirements:

- $p_{\mathrm{T}}(\mu) > 20$ GeV, to remove the background from low energy electrons, not coming from a vector boson decay;
- $2 < \eta(\mu) < 4.25$ to select electrons inside the full ECAL and HCAL acceptance;
- $E_{ECAL}/p > 0.2$, $E_{HCAL}/p < 0.01$ and $E_{HCAL} < 20$ GeV to remove hadrons misidentified as electrons;
- the electron impact parameter (IP) must be less than 0.04 mm to reject electrons from heavy flavour hadrons decays;
- a requirement on the electron isolation I is applied; the isolation is defined as $I = \frac{p_{\mathrm{T}}(e)}{p_{\mathrm{T}}(j_e)}$, where $p_{\mathrm{T}}(j_e)$ is the transverse momentum of the jet containing the electron, built using the jets reconstruction described in Section 3.2. The isolation must be greater than 0.8 to remove electrons not coming from a vector boson decay.

The jet candidates are reconstructed using the standard LHCb algorithm discussed in Section 3.2. The input particles are clustered into jets using the anti-k_{T} algorithm [64] with radius parameter $R = 0.5$. Two jets (j_1 and j_2 where j_1 has the highest p_T in the pair) are selected with the following requirements:

- the two jets must be identified as originating from a heavy flavour quark by the SV tagging algorithm (Section 3.3.1);
- $\mathrm{BDT}_{bc|udsg} > 0.2$ for both jets (definition in Section 3.3.1), to remove the background from light jets;

Table 5.1: Number of selected data events used in the cross section measurements and in the Higgs limit.

	cross sections meas.	limit
muon + dijet	108 events	44 events
electron+dijet	85 events	38 events

- $2.2 < \eta < 4.2$ to select jets with the cone completely inside the LHCb acceptance;
- $p_T > 12.5$ GeV and $p_T < 100$ GeV for both jets;
- the ΔR distance ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$) between the two jets must be greater than 0.5: $\Delta R(j_1, j_2) > 0.5$;
- the ΔR distance between the jets and the selected lepton must be greater than 0.5: $\Delta R(j_1, \ell) > 0.5$ and $\Delta R(j_2, \ell) > 0.5$;
- $p_T(\ell + j_1 + j_2) < 15$ GeV, where $p_T(\ell + j_1 + j_2)$ is the transverse component of the sum of j_1, j_2 and ℓ momenta; this cut is applied to select events where the dijet is balanced by the lepton momentum; if more than one lepton+dijet candidate is selected in the same event then the one with lowest $p_T(\ell + j_1 + j_2)$ is selected.

One extra requirement is applied to events used in the $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ cross sections measurements, but not in the $V + H^0$ limit:

- a Z veto is applied: events with two opposite sign leptons selected as described above are removed if their invariant mass is compatible with the Z mass.

One extra requirement is applied to events used in the $V + H^0$ limit, but not in the $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ cross sections measurements:

- $p_T > 20$ GeV for both jets, to remove the background at low dijet invariant mass.

The number of selected data events for the muon+dijet and the electron+dijet final states are reported in table 5.1

5.4 MC samples

Simulated data are analysed to determine the selection efficiencies and the templates used in the fit (Section 5.9) and in the limit extraction (Section 5.13). MC events of $V + H^0(\rightarrow b\bar{b})$, $V + H^0(\rightarrow c\bar{c})$, and of the processes described in Section 5.2 are generated. The generation step is performed with PYTHIA 8, but ALPGEN (for matrix elements) + PYTHIA 8 (for parton showers) are used for the processes where a vector boson is produced in association with heavy flavour jets ($W + b\bar{b}$, $W + c\bar{c}$, $Z + b\bar{b}$ and $Z + c\bar{c}$). The simulation configuration is defined in the software GAUSS, as described in Section 2.12. The following MC samples have been generated:

- $V(\rightarrow \ell{+}X){+}H^0(\rightarrow b\bar{b})$: generated with PYTHIA 8, requiring at least one lepton from V decay and the two b quarks from H^0 decay in the LHCb acceptance ($2 < \eta < 5$). The Higgs has a mass of 125 GeV and SM properties;
- $V(\rightarrow \ell + X) + H^0(\rightarrow c\bar{c})$: generated with PYTHIA 8, requiring at least one lepton from V decay and the two c quarks from H^0 decay in the LHCb acceptance. The Higgs has a mass of 125 GeV and SM properties;
- $W(\rightarrow \ell\nu_\ell) + b\bar{b}$: generated with ALPGEN+PYTHIA 8, requiring the lepton from W decay with $p_T > 10$ GeV and the two b quarks with $p_T > 3$ GeV in the LHCb acceptance;
- $W(\rightarrow \ell\nu_\ell) + c\bar{c}$: generated with ALPGEN+PYTHIA 8, requiring the lepton from W decay with $p_T > 10$ GeV and the two c quarks with $p_T > 3$ GeV in the LHCb acceptance;
- $t\bar{t}$, with $t \rightarrow W(\rightarrow \ell\nu_\ell) + b$: generated with PYTHIA 8, requiring at least one lepton from t decays with $p_T > 17$ GeV and the two b quarks coming from t quarks in the LHCb acceptance;
- single t: generated with PYTHIA 8, requiring the lepton from t decay with $p_T > 17$ GeV and the two b quarks in the final state in the LHCb acceptance;
- $Z(\rightarrow \ell\ell) + b\bar{b}$: generated with ALPGEN+PYTHIA 8, requiring at least one lepton from Z decay with $p_T > 10$ GeV and the two b quarks with $p_T > 3$ GeV in the LHCb acceptance;
- $Z(\rightarrow \ell\ell) + c\bar{c}$: generated with ALPGEN+PYTHIA 8, requiring at least one lepton from Z decay with $p_T > 10$ GeV and the two c quarks with $p_T > 3$ GeV in the LHCb acceptance;
- $Z(\rightarrow \ell\ell) + Z(\rightarrow b\bar{b})$: generated with PYTHIA 8, requiring at least one lepton from Z decay with $p_T > 5$ GeV and the two b quarks from Z decay in the LHCb acceptance;
- $W(\rightarrow \ell\nu_\ell) + Z(\rightarrow b\bar{b})$: generated with PYTHIA 8, requiring the lepton from W decay with $p_T > 5$ GeV and the two b quarks from Z decay in the LHCb acceptance;
- $W(\rightarrow \tau\nu_\tau) + b\bar{b}$: generated with PYTHIA 8, requiring the lepton from τ decay with $p_T > 10$ GeV and the two b quarks with $p_T > 3$ GeV in the LHCb acceptance;
- $W(\rightarrow \ell\nu_\ell)$+jets: generated with PYTHIA 8, requiring the lepton from W decay with $p_T > 17$ GeV in the LHCb acceptance.

In order to have MC simulation reproducing the data the Global Event Cut (GEC) efficiency, the trigger efficiency and heavy flavour efficiency need to be corrected. The Global Event Cut is a hardware requirement on the number of hits in the SPD subdetector (nSPD) described in Section 2.8. In the 2012 data taking period the GEC efficiency calculated with $Z \to \mu^+\mu^-$ was 88% [87]. In simulation GEC efficiency ranges between 94% and 98% for the relevant SM processes therefore is corrected by a factor 0.92, to approximately match the observed efficiency. A systematic uncertainty will be assigned to this cut. Muon and electron show similar nSPD distribution as can be seen in Fig. 5.3, so the same correction factor is applied to muon+dijet and electron+dijet final states.

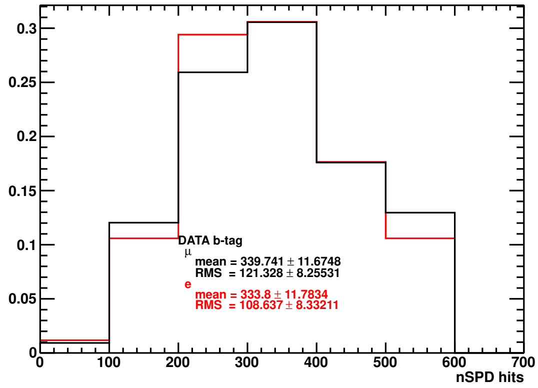

Figure 5.3: nSPD distribution in $W + b\bar{b}$ simulation events, for the electron channel and muon channel.

According to $Z \to \mu\mu$ studies at LHCb [87], the simulated muon trigger efficiencies are different from real ones. They have been measured using a tag and probe technique on $Z \to \mu\mu$ data events. One track that satisfies the muon trigger requirements is selected as tag and a second opposite charge track is selected as probe. The invariant mass of the two particles must be in the [60,120] GeV range to select genuine muons from Z decay and to make the background contamination negligible even if the second track is not selected by the muon trigger. The trigger efficiency is computed using the probe, as the the number of tracks that satisfy the trigger requirements to the total number of selected tracks. The p_{T} dependence of the trigger efficiency is measured in data and simulation in ten η bins from 2.0 to 4.5 and ten p_{T} bins, from 20 to 70 GeV. The ratio data to MC in each bin is used as correction factor in simulation.

The same technique is employed to measure the electron trigger efficiencies in $Z \to ee$ decays [88]. The efficiency has been calculated separately for electrons and positrons, in bins of electron η and p_{T}. Using the tag and probe technique, where a TOS electron is considered as tag to probe positron efficiency and vice versa, the p_{T} dependence of the trigger efficiency is measured in data and simulation in nine η bins from 2.0 to 4.25 (see Fig. 5.4 for the two outermost η bins). The ratio data to MC in each bin is used to correct the simulation.

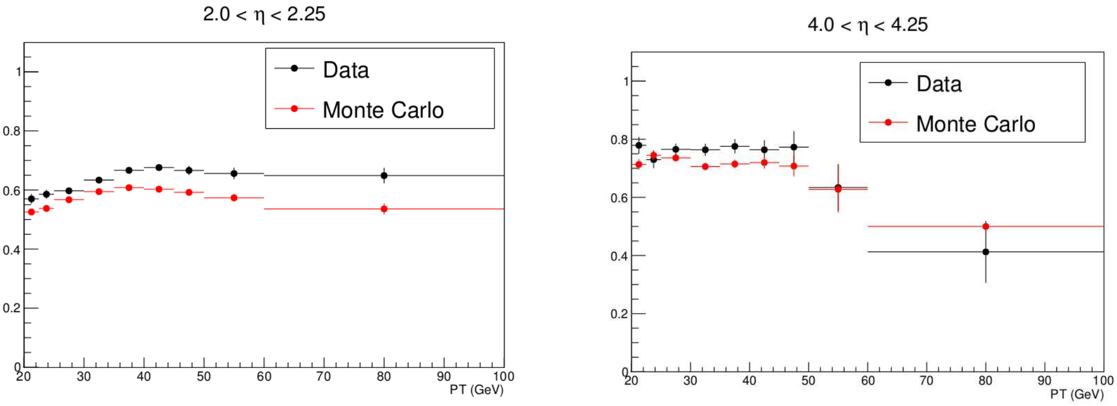

Figure 5.4: Data and simulated electron trigger efficiencies as a function of p_{T} for $2.0 < \eta < 2.25$ (left) and $4.0 < \eta < 4.25$ (right).

The heavy flavour tagging efficiencies in MC are corrected with factors obtained as described in section 3.3.

5.5 Yields prediction

The expected yields of $V + H^0(\to b\bar{b})$, and of the processes described in Section 5.2 is calculated with the method described in Section 4.2.3. The expected yields are computed separately for muon+dijet and the electron+dijet final states. For each process the following quantities are determined:

- the selection efficiency ϵ: it is calculated applying the selection described in Section 5.3 to MC events, as the number of selected MC events to the number of generated events. The corrections explained in Section 5.4 are applied to MC events;
- the theoretical cross section in the LHCb acceptance, $\sigma \cdot A$: it has been obtained at next-to-leading-order (NLO) using the software MCFM v8.6 [89]. The Parton Distribution Functions used in the calculation are those defined in [90]. The lepton universality is assumed, so cross sections related to the muon and electron decay channels are considered to be the same. The acceptance for each process is defined in Section 5.4. The MCFM software provides also the uncertainty on the theoretical prediction. The Higgs cross section is obtained as explained in 1.4 and the acceptance is calculated using PYTHIA 8;
- the integrated luminosity of the collected dataset, $\mathcal{L}$, that is 1.98 ± 0.02 fb^{-1}. It has been measured in [73].

Table 5.2: Theoretical cross sections with relative uncertainties (obtained from MCFM), efficiencies and yield predictions for the studied processes in the muon+dijet final state. The total expected events (QCD background not included) and the observed yield are also reported.

process	$\sigma \cdot A$ (pb)	ϵ	N_{SM}
$W(\mu\nu) + b\bar{b}$	$4.80^{+39.8\%}_{-21.5\%}$	0.50%	47.2
$W(\mu\nu) + c\bar{c}$	10.38	0.052%	10.6
$t\bar{t}$	$12.19^{+26.0\%}_{-21.3\%}$	0.061%	14.8
single t	$2.57^{+11.4\%}_{-25.0\%}$	0.053%	2.7
$Z(\mu\mu) + c\bar{c}$	$1.36^{+22.0\%}_{-31.2\%}$	0.030%	0.8
$Z(\mu\mu) + b\bar{b}$	$0.82^{+110.5\%}_{-100.0\%}$	0.30%	4.8
$Z(\mu\mu)Z(bb) + W(\mu\nu)Z(bb)$	$0.10^{+34.5\%}_{-100.0\%}$	0.50%	2.5
$W(\mu\nu) + jet$	2785	1.8×10^{-5}%	1.0
$V(\mu X) + H^0(\rightarrow b\bar{b})$	0.00237	2.3%	0.11
total (without QCD)			84.5
data			108

Table 5.3: Theoretical cross sections with relative uncertainties, efficiencies and yield predictions for the studied processes in the electron+dijet final state. The total expected events (QCD background not included) and the observed yield are also reported.

process	$\sigma \cdot A$ (pb)	ϵ	N_{SM}
$W(e\nu) + b\bar{b}$	$4.80^{+39.8\%}_{-21.5\%}$	0.21%	20.3
$W(e\nu) + c\bar{c}$	10.38	0.022%	4.5
$t\bar{t}$	$12.19^{+26.0\%}_{-21.3\%}$	0.027%	6.6
single t	$2.57^{+11.4\%}_{-25.0\%}$	0.024%	1.2
$Z(ee) + c\bar{c}$	$1.36^{+22.0\%}_{-31.2\%}$	0.015%	0.4
$Z(ee) + b\bar{b}$	$0.82^{+110.5\%}_{-100.0\%}$	0.16%	2.6
$Z(ee)Z(bb) + W(e\nu)Z(bb)$	$0.10^{+34.5\%}_{-100.0\%}$	0.61%	1.2
$W(e\nu) + jet$	2785	3.6×10^{-6}%	0.2
$V(eX) + H^0(\rightarrow b\bar{b})$	0.00237	1.1%	0.0518
total (without QCD)			37
data			85

The values of ϵ, $\sigma \cdot A$ and the expected yields (N_{SM}) are reported in table 5.2 for the muon+dijet selection and in table 5.3 for the electron+dijet selection. The selection requirements applied to calculate the expected yields in the tables are those used in the $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ cross-section measurements, as explained in Section 5.3. Systematic uncertainties on the expected yields are discussed in Section 5.10.

The selection of the electron final state is less efficient than the muon one, due to the p_{T} threshold and lower trigger efficiencies. Figure 5.5 shows the electron and muon p_{T} distributions after the selection in the $V + H^0$ MC. It shows that the electron spectrum is softer than the muon one, in agreement with the electron energy loss due to bremsstrahlung radiation. The effiency of the p_T cut is calculated using $V + H^0$ MC events as:

$$\epsilon(p_T > 20\text{GeV}; \ell) = \frac{N(p_T > 20\text{GeV}; \ell)}{N_{gen}(\ell)},$$

where $N(p_T > 20\text{GeV}; \ell)$ is the number of MC events where the lepton has $p_T > 20$ GeV and flavour ℓ, and $N_{gen}(\ell)$ is the total number of generated events with lepton flavour ℓ. The following is obtained:

$$\frac{\epsilon(p_T > 20\text{GeV}; e)}{\epsilon(p_T > 20\text{GeV}; \mu)} \approx 0.75.$$

The lepton trigger efficiency on events that satisfy the lepton p_T cut is calculated using $V + H^0$ MC events as:

$$\epsilon(TOS; \ell) = \frac{N(TOS \text{ and } p_T > 20\text{GeV}; \ell)}{N(p_T > 20\text{GeV}; \ell)},$$

where TOS (Trigger On Signal) indicates that the lepton is selected by the trigger. The following is obtained:

$$\frac{\epsilon(TOS; e)}{\epsilon(TOS; \mu)} \approx 0.78.$$

The total loss of efficiency due to the p_{T} threshold and lower trigger efficiencies in the electron process with respect to the muon one is:

$$1 - \frac{\epsilon(p_T > 20 \text{ GeV}; e)}{\epsilon(p_T > 20\text{GeV}; \mu)} \cdot \frac{\epsilon(TOS; e)}{\epsilon(TOS; \mu)} \approx 41\%.$$

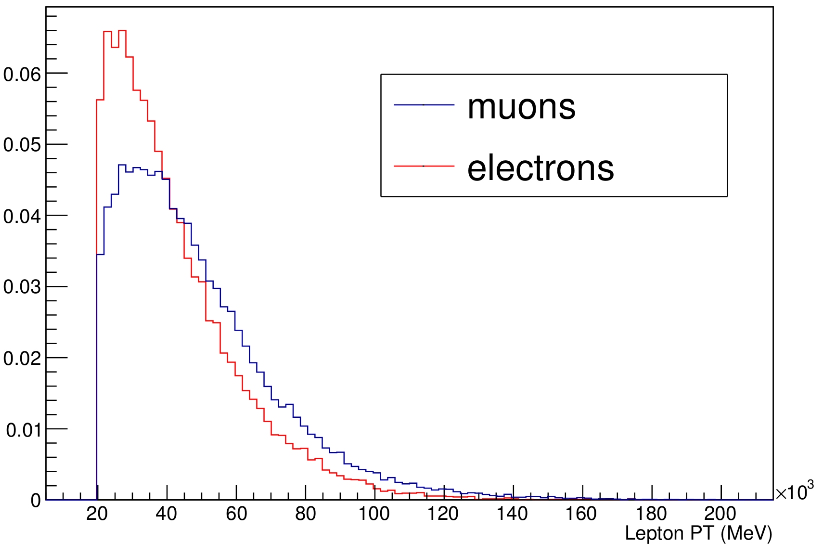

Figure 5.5: p_{T} distribution of leptons in $V + H^0$ simulated events.

5.6 Background from lepton misidentification and heavy flavour decays

The main background sources to $V + b\bar{b}$, other than physical processes described in Section 5.2, are:

- fake electrons: hadrons misidentified as electron by the calorimeter system, as explained in Section 2.8;
- real leptons not from W/Z decays, originated from the decays of heavy flavour mesons and baryons.

Observables templates for these two type of background have been obtained in data sidebands, defined as:

- electron misidentification sideband: data sample selected by applying the electron+dijet selection described in Section 5.3 but with $E_{ECAL}/p < 0.2$, $E_{HCAL}/p > 0.01$ instead of $E_{ECAL}/p > 0.2$, $E_{HCAL}/p < 0.01$. This sample is dominated by hadrons and the contribution from real electrons is assumed negligible, as demonstrated in [91]. The event topology is assumed to not depend on the calorimeters observables related to the lepton;
- heavy flavour sideband: data sample selected by applying the lepton+dijet selection described in Section 5.3 but with $IP(\ell) > 0.04$ mm instead of $IP(\ell) < 0.04$ mm. This sample is dominated by displaced leptons coming from heavy flavour decays and the contribution of leptons from W/Z decays is assumed negligible, as demonstrated in [91]. The event topology is assumed to not depend on the lepton IP.

The number of expected misidentification and heavy flavour events is obtained with a data driven technique. It is determined separately for the muon+dijet sample and for the electron+dijet sample. For the muon+dijet sample:

1. data events are selected with the muon+dijet requirements used in the $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ cross sections measurements as explained in Section 5.3, but the $I > 0.8$ cut is not applied;
2. the following muon isolation templates are built:
    - W/Z template $T_{W/Z}(I)$: obtained from the distribution of $V + H^0(\to b\bar{b})$ MC events;
    - heavy flavour template $T_{hf}(I)$: obtained from the distribution of data events in the heavy flavour sideband;
3. a fit to the isolation distribution is performed using the model

$$f(I) = N_{W/Z}T_{W/Z}(I) + N_{hf}T_{hf}(I),$$

Table 5.4: Isolation fit result for the muon+dijet sample.

event type	yield
W/Z	153.0 ± 15.6
heavy flavour	99.0 ± 18.6

where $N_{W/Z}$ and N_{hf} are the number of W/Z events and heavy flavour events respectively. $N_{W/Z}$ and N_{hf} are free parameters in the fit. The fit result is shown in figure 5.6 and reported in table 5.4.

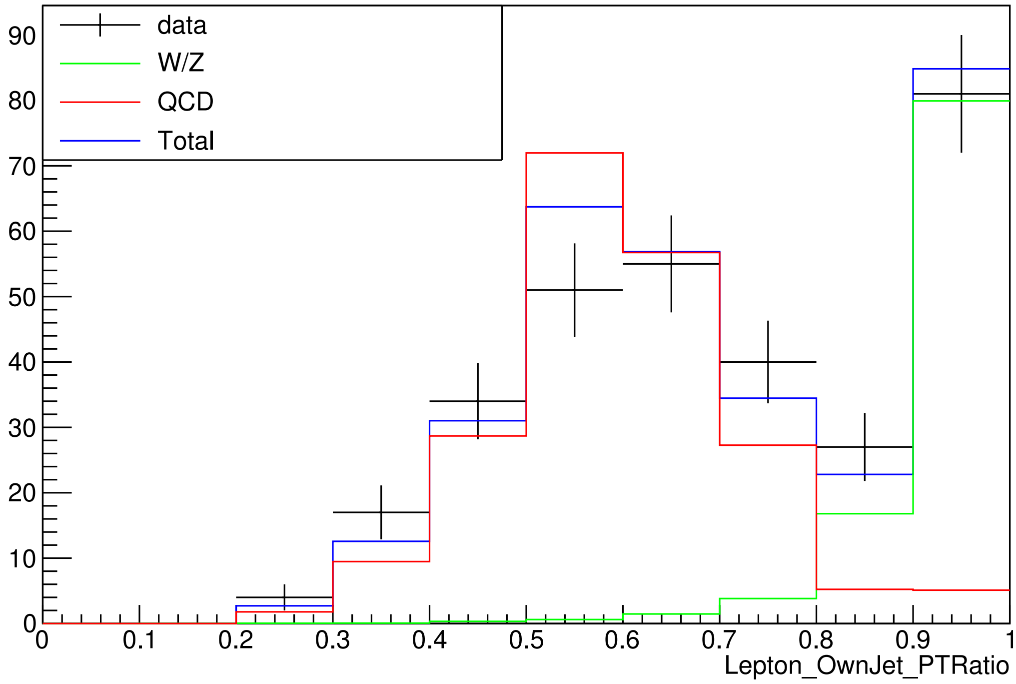

Figure 5.6: Isolation fit result for the muon+dijet sample.

4. the number of heavy flavour events in the $I > 0.8$ region is extrapolated using N_{hf} and $T_{hf}(I)$.

For the electron+dijet sample:

1. data events are selected with the electron+dijet requirements used in the $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ cross sections measurements as explained in Section 5.3, but the $I > 0.8$ cut is not applied;

2. the following electron isolation-p_T templates are built:
   - $W/Z(\rightarrow e+X)$ template $T_{W/Z}(I,p_T)$: obtained from the distribution of $V + H^0(\rightarrow b\bar{b})$ MC events that pass the selection;
   - $W/Z(\rightarrow \tau+X)$ template $T_{W/Z}(I,p_T)$: obtained from the distribution of $W(\tau\nu)$+jets MC events that pass the selection;
   - heavy flavour template $T_{hf}(I,p_T)$: obtained from the distribution of data events in the heavy flavour sideband;
   - misidentification template $T_{mis}(I,p_T)$: obtained from the distribution of data events in the misidentification sideband;

3. a fit to the (I,p_T) distribution is performed using the model

   $$f(I,p_T) = N_{W/Z(e)}T_{W/Z(e)}(I,p_T)+N_{W/Z(\tau)}T_{W/Z(\tau)}(I,p_T)+N_{hf}T_{hf}(I,p_T)+N_{mis}T_{mis}(I,p_T),$$

   where $N_{W/Z}(e)$ is the number of $W/Z(\rightarrow e+X)$ events, $N_{W/Z}(\tau)$ is the number of $W/Z(\rightarrow \tau+X)$ events, N_{hf} is the number of heavy flavour events and N_{mis} is the number of misidentificated events. $N_{W/Z}(e)$, N_{hf} and N_{mis} are free parameters in the fit, $N_{W/Z}(\tau)$ is fixed to the estimation obtained as explained in Section 5.5. The fit result is shown in figure 5.7 and reported in table 5.5.

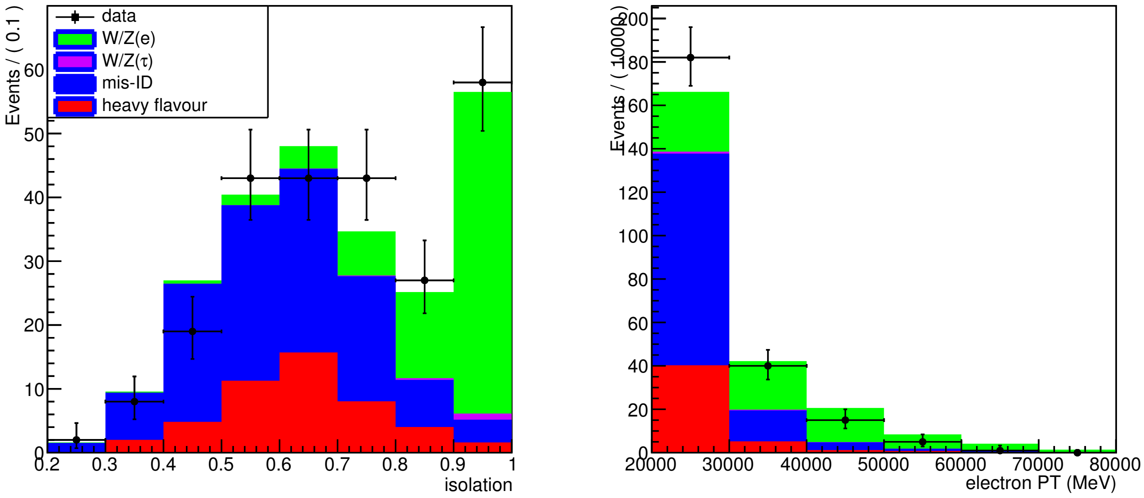

Figure 5.7: Isolation-p_T fit result projections for the electron+dijet sample. The isolation is displayed on the left and the p_T on the right.

4. the number of heavy flavour events in the $I > 0.8$ region is extrapolated using N_{hf} and $T_{hf}(I,p_T)$ and the number of misidentificated events in the $I > 0.8$ region is extrapolated using N_{hf} and $T_{hf}(I,p_T)$.

Table 5.5: Isolation-p_{T} fit result for the electron+dijet sample.

event type	yield
W/Z	77 ± 11
heavy flavour	47 ± 33
mis-identification	117 ± 34

In the rest of the thesis the heavy flavour background for the muon+dijet sample and the heavy flavour+misidentification background for the electron+dijet sample are both called “QCD background”. The number of QCD background events (N_{QCD}) obtained with the technique described above is reported in table 5.6.

Table 5.6: N_{QCD} values are presented with statistical errors.

sample	N_{QCD}
dijet+muon	5.4 ± 1.5
dijet+electron	17.0 ± 4.6

5.7 Study of kinematical and global event distributions

The $V+H(\to b\bar{b})$ is separated from $t\bar{t}$, $W+b\bar{b}$ and $W+c\bar{c}$ by looking at several observables:

- the dijet invariant mass m_{jj};
- the jets transverse momenta $p_T(j_1)$ and $p_T(j_2)$;
- the jets pseudo-rapidities $\eta(j_1)$ and $\eta(j_2)$;
- the jets invariant masses $m(j_1)$ and $m(j_2)$;
- the lepton transverse momentum $p_T(\ell)$;
- the lepton pseudo-rapidity $\eta(\ell)$;
- $\mathrm{BDT}_{bc|udsg}$ for both jets;
- $\mathrm{BDT}_{b|c}$ for both jets;
- the ΔR distance between the two jets;

- the angle between the dijet momentum in the laboratory frame and the lepton momentum in the dijet rest frame, $cos(\theta^l_{jj})$

The same observables are used to separate $t\bar{t}$ from $W + b\bar{b}$. For each final state (muon+dijet or electron+dijet) data and MC are compared, where the MC is obtained as the sum of all the expected contributions. Figure 5.8 shows the distributions of the variables for the muon+dijet dataset with the expected composition. It shows that m_{jj} is the observable that separates better $V + H(\rightarrow b\bar{b})$ from the background and $t\bar{t}$ from $W + b\bar{b}$. It is also evident that MC is not in perfect agreement with data in several distributions, within the statistical uncertainty. However systematics uncertainties on the prediction are not displayed and should be considered in the comparison. Figure 5.8 shows the distributions of the variables for the electron+dijet dataset with the expected composition. Again m_{jj} is the best variable to separate $V + H(\rightarrow b\bar{b})$ from the background and MC is not in perfect agreement with data in several distributions, within the statistical uncertainty.

5.8 Multivariate analyses for VH vs $t\bar{t}$, VH vs $W + b\bar{b}$ and $t\bar{t}$ vs $W + b\bar{b}$ separation

Multivariate classifiers (MVA) are used to separate VH from $t\bar{t}$, VH from $W + b\bar{b}$ and $t\bar{t}$ from $W + b\bar{b}$. The Bosted Decision Trees (BDT) algorithm is used for this purpose. The MVA outputs are chosen to be uncorrelated with the dijet invariant mass, therefore the *uGB* method described in Section 4.3.5 is used. The following classifiers are trained:

- $uGB[V(\rightarrow \mu + X) + H^0$ vs $W(\rightarrow \mu + X) + b\bar{b}]$: trained with the $V + H^0$ MC as signal sample and the $W + b\bar{b}$ MC as background sample, both in the muon+dijet final state;
- $uGB[V(\rightarrow \mu + X) + H^0$ vs $t\bar{t}(\rightarrow \mu + b\bar{b} + X)]$: trained with the $V + H^0$ MC as signal sample and the $t\bar{t}$ MC as background sample, both in the muon+dijet final state;
- $uGB[W(\rightarrow \mu + X) + b\bar{b}$ vs $t\bar{t}(\rightarrow \mu + b\bar{b} + X)]$: trained with the $W + b\bar{b}$ MC as signal sample and the $t\bar{t}$ MC as background sample, both in the muon+dijet final state;
- $uGB[V(\rightarrow e + X) + H^0$ vs $W(\rightarrow e + X) + b\bar{b}]$: trained with the $V + H^0$ MC as signal sample and the $W + b\bar{b}$ MC as background sample, both in the electron+dijet final state;
- $uGB[V(\rightarrow e + X) + H^0$ vs $t\bar{t}(\rightarrow e + b\bar{b} + X)]$: trained with the $V + H^0$ MC as signal sample and the $t\bar{t}$ MC as background sample, both in the electron+dijet final state;
- $uGB[W(\rightarrow e + X) + b\bar{b}$ vs $t\bar{t}(\rightarrow e + b\bar{b} + X)]$: trained with the $W + b\bar{b}$ MC as signal sample and the $t\bar{t}$ MC as background sample, both in the electron+dijet final state;

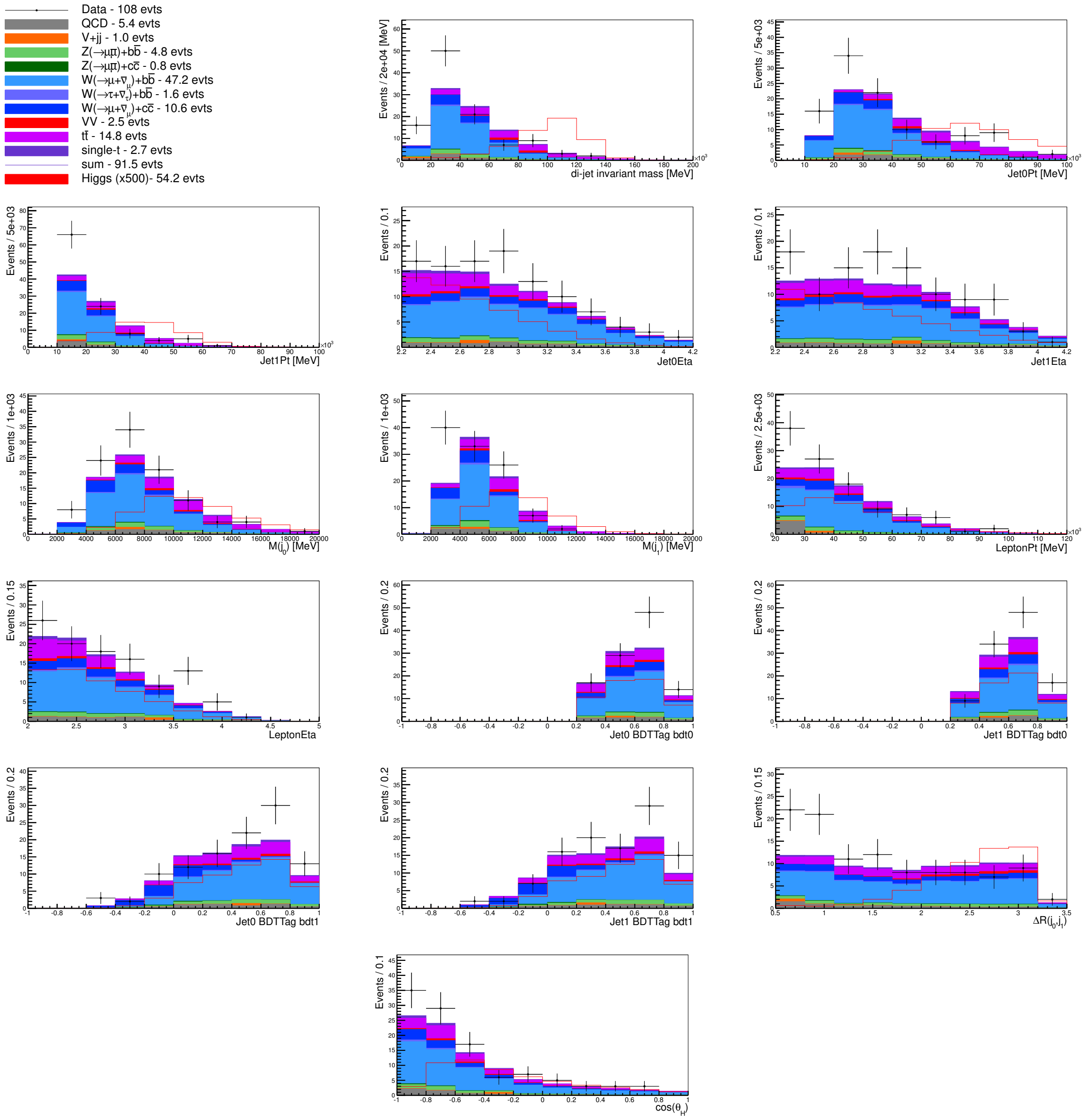

Figure 5.8: Observed and expected observables distributions for the dijet+muon data sample.

Since the first three classifier are used exclusively in the muon+dijet sample and the last three in the electron+dijet sample, in the rest of the thesis the output of the BDTs are called $uGB(W/Z + H^0 \text{ vs } W + b\bar{b})$, $uGB(W/Z + H^0 \text{ vs } t\bar{t})$ and $uGB(W + b\bar{b} \text{ vs } t\bar{t})$, without explicitely indicate the ℓ flavour. The variables in input to the classifiers are:

- the invariant mass of the two jets $m(\text{jet}_1)$ and $m(\text{jet}_2)$;
- the p_T of the two jets, $p_T(\text{jet}_2)$, $p_T(l)$;
- the pseudo-rapidity of the lepton, $\eta(l)$;
- the ΔR distance between the dijet and the lepton, $\Delta R(jj, l)$;
- the ΔR distances between the dijet and the jets, $\Delta R(jj, jet_1)$ and $\Delta R(jj, jet_2)$;

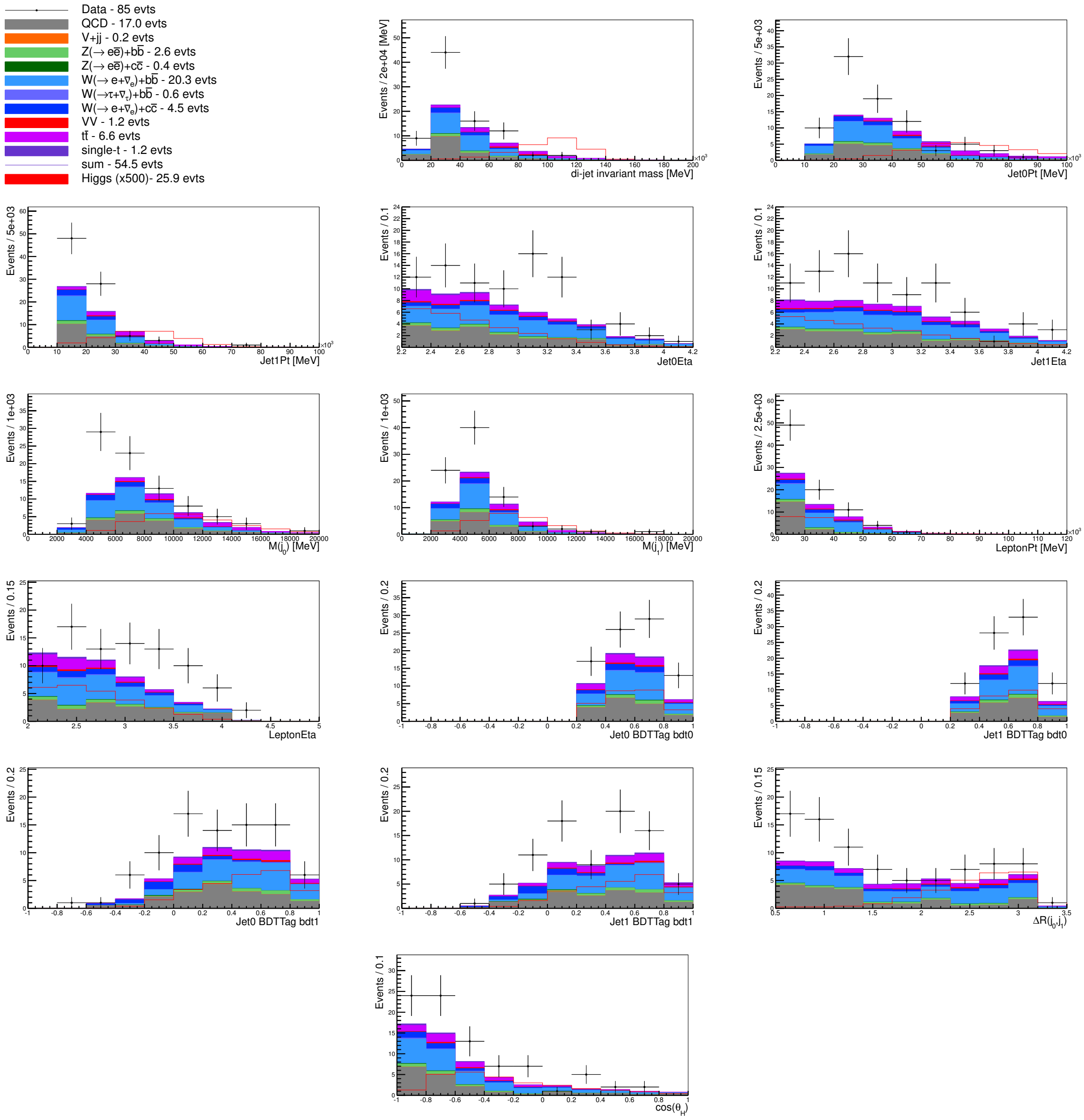

Figure 5.9: Observed and expected observables distributions for the dijet+electron data sample.

- $cos(\theta^l_{jj})$, where θ^l_{jj} is the angle between the dijet momentum in the laboratory frame and the lepton momentum in the dijet rest frame.

The output of the classifiers in data compared with MC prediction are shown in figure 5.10 for muon+dijet events and in figure 5.11 for electron+dijet events.

In figure 5.12 the plot on the left shows the uGB distribution in a $t\bar{t}$ simulation sample generated with PYTHIA 8 compared with one sample generated with POWHEG. This shows that the results is independent to the MC generator, while it is known that these generators produce events with different track and SPD multiplicities. On the right the same plot is shown with the Jet Energy Correction smeared with a variation of 10%. Again no significant difference between the distributions is found. For these reasons the uncertainty associated to the uGB shape is considered negligible.

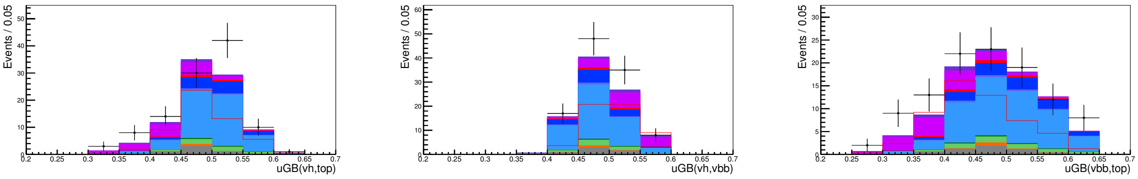

Figure 5.10: Classifier distribution for data events in the muon+dijet final state, compared with MC prediction.

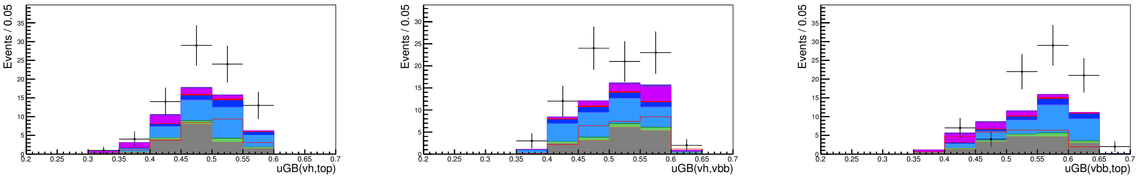

Figure 5.11: Classifier distribution for data events in the electron+dijet final state, compared with MC prediction.

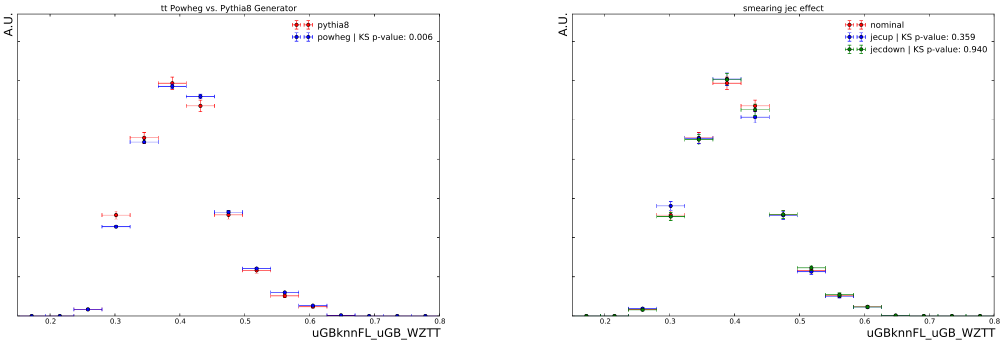

Figure 5.12: Shape comparisons for different MC generators in the $t\bar{t}$ MC sample (right) and different smearing (10 % up vs down) in the JEC applied to the jet momentum (left).

5.9 Fit for the extraction of $t\bar{t}$, W^{+}+$b\bar{b}$, W^{-}+$b\bar{b}$, W^{+}+$c\bar{c}$ and $W^{-}+c\bar{c}$ yields

A fit to the dijet invariant mass m_{jj}, $BDT_{b|c}(\mathrm{jet}_1)$, $BDT_{b|c}(\mathrm{jet}_2)$ and $uGB(W+b\bar{b}$ vs $t\bar{t})$ distributions is performed to measure the yield of $t\bar{t}$, $W^{+}+b\bar{b}$, $W^{-}+b\bar{b}$, $W^{+}+c\bar{c}$ and $W^{-}+c\bar{c}$ events in the selected data sample. The results of the fit are used in Section 5.11 to measure the cross sections of these processes. The data sample selected as explained in Section 5.3 is divided into four sub-samples that are fitted simultaneously:

- a sample where the selected lepton is identified as a μ^{+};
- a sample where the selected lepton is identified as a μ^{-};
- a sample where the selected lepton is identified as a e^{+};
- a sample where the selected lepton is identified as a e^{-};

In the fit the following are considered “signal” processes:

- $t\bar{t}$;
- $W^{+}+b\bar{b}$;
- $W^{-}+b\bar{b}$;
- $W^{+}+c\bar{c}$;
- $W^{-}+c\bar{c}$.

The “background” processes are:

- single t;
- $Z(\rightarrow \ell\ell)+b\bar{b}$;
- $Z(\rightarrow \ell\ell)+c\bar{c}$;
- $Z(\rightarrow \ell\ell)+Z(\rightarrow b\bar{b})$;
- $W(\rightarrow \ell\nu_{\ell})+Z(\rightarrow b\bar{b})$;
- $W(\rightarrow \tau\nu_{\tau})+b\bar{b}$;
- $W(\rightarrow \ell\nu_{\ell})$+jets;
- QCD.

For each of the signal and background processes 1-dimesional binned templates of the m_{jj}, $BDT_{b|c}(\text{jet}_1)$, $BDT_{b|c}(\text{jet}_2)$ and $uGB(W+b\bar{b}\text{ vs }t\bar{t})$ distributions are built using MC events. The only exception is the QCD background template, which is built as explained in Section 5.6.

The model used in the fit is the following:

$$f(\vec{x}) = \prod_{\ell,i}\left[f_\ell(x_i)\right] = \prod_{\ell,x_i}\left[\sum_{s}^{N_{sig}} K_s N^{exp}_{\ell,s} T_{\ell,s}(x_i) + \sum_{b}^{N_{bkg}} N^{exp}_{\ell,b} T_{\ell,b}(x_i)\right]$$

where

- $f_\ell(x_i)$ is the *pdf* of the variable x_i=m_{jj}, $BDT_{b|c}(j_1)$, $BDT_{b|c}(j_2)$ or $uGB(W+b\bar{b}\text{ vs }t\bar{t})$ for the subsample with lepton flavour ℓ=μ^+, μ^-, e^+ or e^-; a total of sixteen $f_\ell(x)$ distributions are fitted simultaneously, correlation are not included;
- the first sum runs on the signal processes, which are indicated with the index s; N_{sig} is the number of signal processes ($N_s = 5$);
- $T_{\ell,s}(x_i)$ is the template of the signal process s with lepton flavour ℓ for the x_i distribution;
- $N^{exp}_{\ell,s}$ is the expected yield of signal process s, calculated as explained in Section 5.5;
- K_s is defined as
$$K_s = \frac{\sigma^{true}_{\ell,s}}{\sigma^{exp}_{\ell,s}} = \frac{N^{true}_{\ell,s}}{N^{exp}_{\ell,s}},$$
where $\sigma^{true}_{\ell,s}$ is the real cross section, $\sigma^{exp}_{\ell,s}$ is the expected cross section and $N^{true}_{\ell,s}$ is the real yield of signal process s. K_s is assumed to be independent from the lepton flavour due to lepton universality. Therefore five K_s are included, $K(t\bar{t})$, $K(W^+ + b\bar{b})$, $K(W^- + b\bar{b})$, $K(W^+ + c\bar{c})$ and $K(W^- + c\bar{c})$ which are the free parameters of the fit;
- the second sum runs on the background processes, which are indicated with the index s; N_{bkg} is the number of background processes;
- $T_{\ell,b}(x_i)$ is the template of the background process b with lepton flavour ℓ for the x_i distribution;
- $N^{exp}_{\ell,b}$ is the expected yield of background process b, calculated as explained in Section 5.5.

The fit is performed with the binned maximum likelihood technique using the *RooFit* package of the software ROOT. Asymmetric statistical uncertainties are obtained using the *Minos* algorithm .

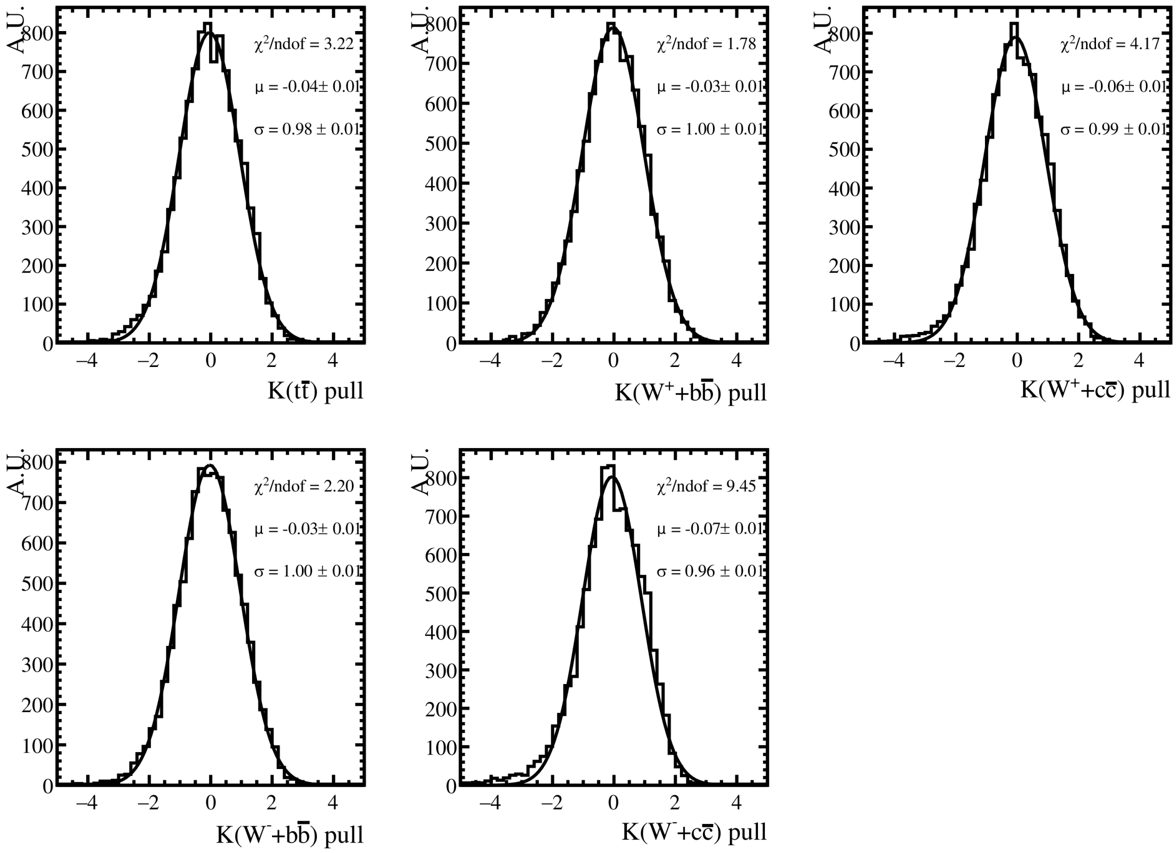

Figure 5.13: Pulls after running 10000 toys on the 5 fitted parameters ($K(t\bar{t})$, $K(W^+ + b\bar{b})$, $K(W^- + b\bar{b})$ $K(W^+ + c\bar{c})$ and $K(W^- + c\bar{c})$. The results of a gaussian fit are also shown.

Test of the fitting procedure

Toy MC sample are fitted to verify any fit bias. The procedure is the same described in Section 4.4.5. Pseudo-experiments have been generated using the fit model, with the K_s set to 1. Pulls of the K_s obtained from the fits with respect to $K_s = 1$ are calculated and their distributions are fitted with gaussians. Figure 5.13 shows the pulls distributions for the five fitted components after 10000 toys. It shows that no significant bias is present.

5.10 Systematics

In this section the systematic uncertainties considered in the $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ cross section measurements and Higgs limits are discussed. There may be two types of systematic errors, one that affects the expected yields of signal and background components and the other that affects the signal and background templates.

These systematics are introduced in the fit and in the limit computation as nuisance parameters. The templates $T_{\ell,p}$, where p indicates the process and ℓ the lepton flavour, are substituted with the effective templates:

$$T'_{\ell,p}(b) = T_{\ell,p}(b) + \sum_{i=0}^{nsyst} \left[\alpha_i \sigma_{i,\ell,p}(b) + \Delta_{i,\ell,p}(b)\right]$$

where

- $T'_{\ell,p}(b)$ is the value of the effective template in the bin b;
- $T_{\ell,p}(b)$ is the value of the template in the bin b;
- $nsyst$ is the number of nuisance parameters; each nuisance parameter is indicated with the index i in the sum;
- $\sigma_{i,\ell,p}(b)$ is the uncertainty in the bin b computed as

  $$\sigma^{up}_{i,\ell,p}(b) = |T^{up}_{i,\ell,p}(b) - T_{\ell,p}(b)|,$$
  $$\sigma^{dw}_{i,\ell,p}(b) = |T^{dw}_{i,\ell,p}(b) - T_{\ell,p}(b)|,$$
  $$\sigma_{i,\ell,p}(b) = \frac{\sigma^{up}_{i,\ell,p}(b) + \sigma^{dw}_{i,\ell,p}(b)}{2},$$

  where $T^{up}_{i,\ell,p}$ is the template built for the up variation of the systematic uncertainty and $T^{dw}_{i,\ell,p}$ is the template for the down variation;
- $\Delta_{i,\ell,p}(b)$ is the shift introduced to symmetrize the error, defined as

  $$\Delta_{i,\ell,p}(b) = \frac{T^{up}_{i,\ell,p}(b) + T^{dw}_{i,\ell,p}(b)}{2} - T_{\ell,p}(b);$$
- α_i is a free parameter constrained to vary following a gaussian distribution with mean 0 and width 1.

In analogy with the effective templates, the expected background yields $N^{exp}_{\ell,p}$ are substituted with the effective expected yields:

$$N'_{\ell,p} = N^{exp}_{\ell,p} + \sum_{i=0}^{nsyst} \left[\alpha_i \sigma_{i,\ell,p} + \Delta_{i,\ell,p}\right].$$

The systematics that affect only the expected yields are:

- luminosity: the uncertainty $\sigma_{\text{lum},\ell,p}$ is taken equal to 1% for all processes and lepton flavours, as explained in Section 4.5.10;

- GEC efficiency: as explained in section 5.4 a data/MC correction factor for GEC efficiency is applied to MC events. Figure 5.14 compares the nSPD distributions for $V + H^0$, $t\bar{t}$, single-t and $W + c\bar{c}$ MC events, before applying the GEC cut. Since the GEC efficiencies calculated separately in these MC samples agree within a level of 2%, $\sigma_{\text{GEC},\ell,p}$ is set equal to 2% for all processes and lepton flavours;

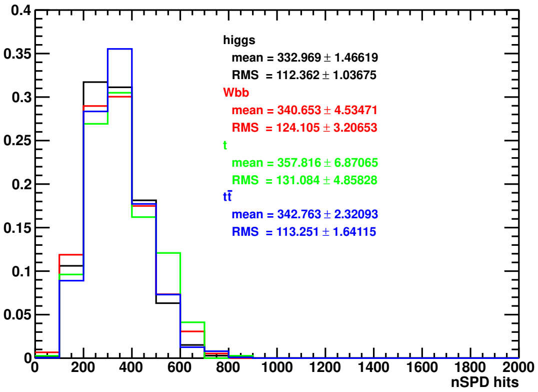

Figure 5.14: nSPD distributions for the relevant physics processes after full selection except the nSPD trigger requirement.

- muon selection efficiency: muon track requirements and muon identification efficiencies are not found to be significantly different between simulation and data, as measured in [87]. Since this agreement is verified at a level of 1.0%, $\sigma_{\mu\text{eff},\ell,p}$ is set equal to 1.0% in all processes and $\ell = \mu^+, \mu^-$;

- electron selection efficiency: electron track requirements, identification and p_T threshold efficiencies are not found to be significantly different between simulation and data, as measured in [87]. Since this agreement is verified at a level of 1.9%, $\sigma_{\mu\text{eff},\ell,p}$ is set equal to 1.9% in all processes and $\ell = e^+, e^-$;

- theoretical cross sections: the uncertainty on the theoretical cross section used to calculated the expected yields are provided by the software MCFM [89]; these uncertainties are propagated to the expected yields and used as $\sigma_{\text{th},\ell,p}$. In the fit this nuisance parameter is not applied to the expected $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ yields, since the σ_s^{exp} are normalization factor for the measured cross sections.

- QCD yield: the uncertainty on the expected QCD yield is computed in Section 5.6.

The systematics that affect both the templates and the expected yields are:

- Jet Energy Correction: up and down templates are built by varying the JEC in MC according to its up and down errors, explained in Section 3.2.5; the variation of the expected yield is computed similarly, by varying the JEC before applying the selection to MC events in the efficiency determination;
- jet tagging efficiency: up and down templates are built by varying the correction applied to MC events for the tagging efficiency according to its up and down errors, explained in Section 3.3.3; the error on the efficiency is also propagated to the expected yields;
- jet energy resolution: only an up variation of the template is considered, obtained by smearing the jet p_T by 10% in MC before the selection, as explained is Section 4.5.4; the uncertainty the expected yield is computed similarly, by smearing the jets p_T before the selection in the efficiency determination;
- jet identification efficiency: up and down templates are built by varying the correction applied to MC events for the jet identification efficiency according to its up and down error, described in Section 3.2.6; the uncertainty on the efficiency is also propagated to the expected yields;
- muon trigger efficiency: up and down templates are built by varying the correction applied to MC events for the muon trigger efficiency according to its up and down error measured in [87]; the uncertainty on the efficiency is also propagated to the expected yields, the average variation of those is of about 1.7%;
- electron trigger efficiency: up and down templates are built by varying the correction applied to MC events for the electron trigger efficiency according to its up and down error measured in [88]; the uncertainty on the efficiency is also propagated to the expected yields, the avarage variation of those is of about 5%.

The MC statistics limitation affects the templates used in this analysis, since in each bin a poissonian fluctuation on the number of MC events is possible. This systematic source is taken into account in the fit by employing the Beeston-Barlow method [92] available in the *RooFit* package.

5.11 $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ cross sections measurements

The fit described in Section 5.9 with nuisance parameters applied as described in Section 5.10 is performed to extract $K(t\bar{t})$, $K(W^+ + b\bar{b})$, $K(W^- + b\bar{b})$, $K(W^+ + c\bar{c})$ and $K(W^- + c\bar{c})$. Figures 5.15, 5.16, 5.17 and 5.18 show data with the result of the fit superimposed. The results on the free parameters, with statistical and systematics uncertainties, are reported in table 5.7.

Table 5.7: Fit results, separating the statistical and systematic uncertainties in each case.

Parameter	Result
$K(t\bar{t})$	$1.17^{+0.35}_{-0.31}(\text{stat.})^{+0.36}_{-0.21}(\text{syst.})$
$K(W^{+}+b\bar{b})$	$1.49^{+0.23}_{-0.22}(\text{stat.})^{+0.36}_{-0.25}(\text{syst.})$
$K(W^{-}+b\bar{b})$	$1.67^{+0.33}_{-0.30}(\text{stat.})^{+0.42}_{-0.29}(\text{syst.})$
$K(W^{+}+c\bar{c})$	$1.92^{+0.68}_{-0.58}(\text{stat.})^{+0.62}_{-0.31}(\text{syst.})$
$K(W^{-}+c\bar{c})$	$1.58^{+0.87}_{-0.73}(\text{stat.})^{+0.59}_{-0.26}(\text{syst.})$

Table 5.8: Fit significances obtained with Wilk's theorem. Note that the results do not include the effect of the MC statistics template variations.

Sample	Significance
$t\bar{t}$	4.9
$W^{+}+b\bar{b}$	7.1
$W^{-}+b\bar{b}$	5.6
$W^{+}+c\bar{c}$	4.7
$W^{-}+c\bar{c}$	2.5

The significances for the $t\bar{t}$, $W^{+}+b\bar{b}$, $W^{-}+b\bar{b}$, $W^{+}+c\bar{c}$ and $W^{-}+c\bar{c}$ observation are computed with the Wilk's theorem, using the formula

$$S=\sqrt{2\Delta LL_{min}},$$

where ΔLL_{min} is the difference between the minimum of the log-likelihood when one of the sample is removed from the fitting model and the minimum of the log-likelihood when the sample is included. The significances are shown in table 5.8.

The $t\bar{t}$, $W^{+}+b\bar{b}$, $W^{-}+b\bar{b}$, $W^{+}+c\bar{c}$ and $W^{-}+c\bar{c}$ cross sections are obtained with

$$\sigma_s=K_s\cdot\sigma_s^{exp}\cdot A,$$

where the index s indicates the process, σ_s^{exp} is the expected cross section, computed using MCFM as explained in Section 5.5, and A is the acceptance factor. The acceptance is defined by requirements in table 5.9 and A is computed using PYTHIA 8.

The observed and expected cross-sections are presented in table 5.10 and represented in figure 5.19. The measurements are compatible with the theoretical prediction within the uncertainties. From these result we can conclude that the Standard Model describes correctly the dijet+lepton data sample collected by LHCb in pp collisions at 8 GeV.

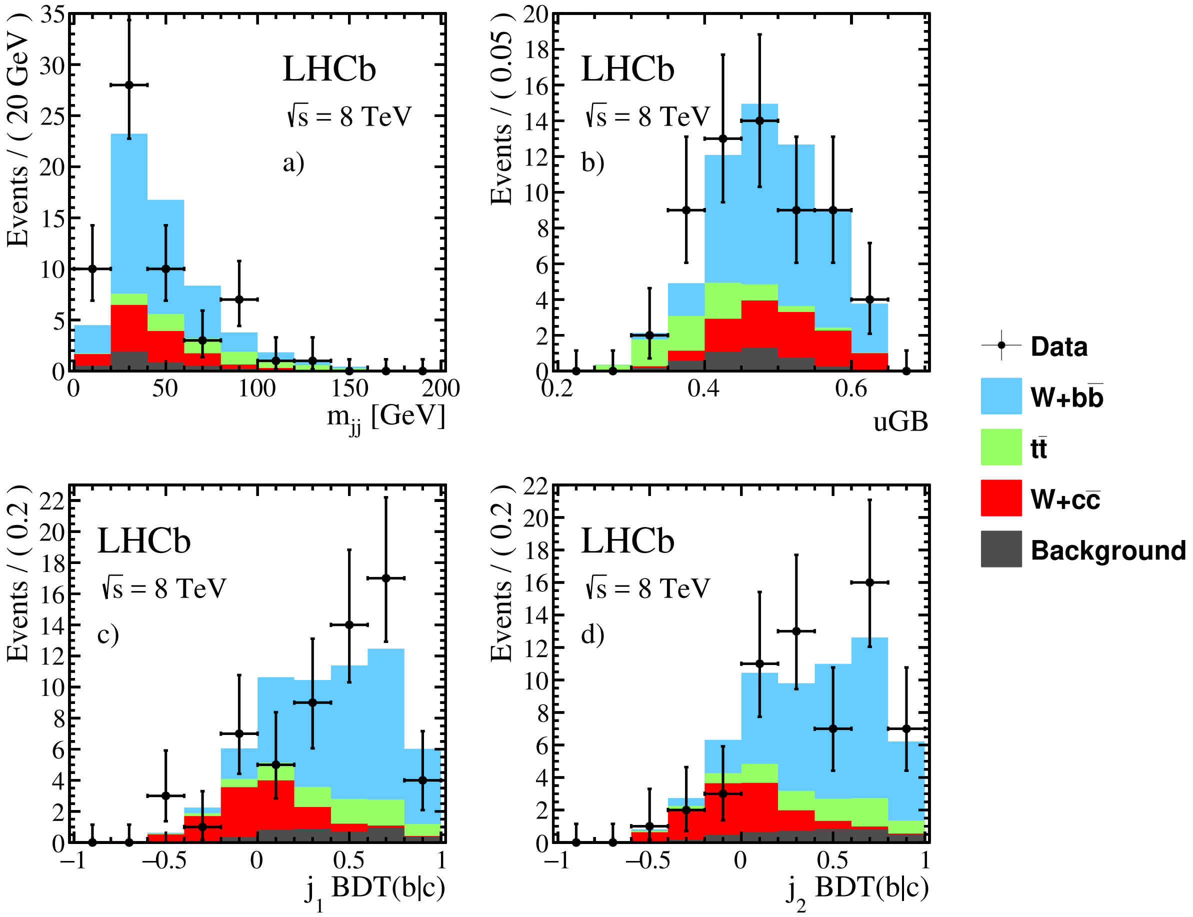

Figure 5.15: Projection of the simultaneous data 4D-fit result on the μ^+ sample: a) the dijet mass; b) the *uGB* response; c) the $\mathrm{BDT}_{b|c}$ of the first jet; and d) the $\mathrm{BDT}_{b|c}$ of the second jet.

Table 5.9: Acceptance definition for $t\bar{t}$, $W^+ + b\bar{b}$, $W^- + b\bar{b}$, $W^+ + c\bar{c}$ and $W^- + c\bar{c}$ cross sections measurements.

object	requirement
muon	$p_\mathrm{T} > 20\,\mathrm{GeV}, 2 < \eta < 4.5$
electron	$p_\mathrm{T} > 20\,\mathrm{GeV}, 2 < \eta < 4.25$
jets	$12.5\,\mathrm{GeV} < p_\mathrm{T} < 100\,\mathrm{GeV}, 2.2 < \eta < 4.2, \Delta R(j_1, j_2) > 0.5$
lepton, jets	$\Delta R(\mu, j) > 0.5$
lepton, jets	$p_\mathrm{T}(\mu + j_1 + j_2) < 15\,\mathrm{GeV}$

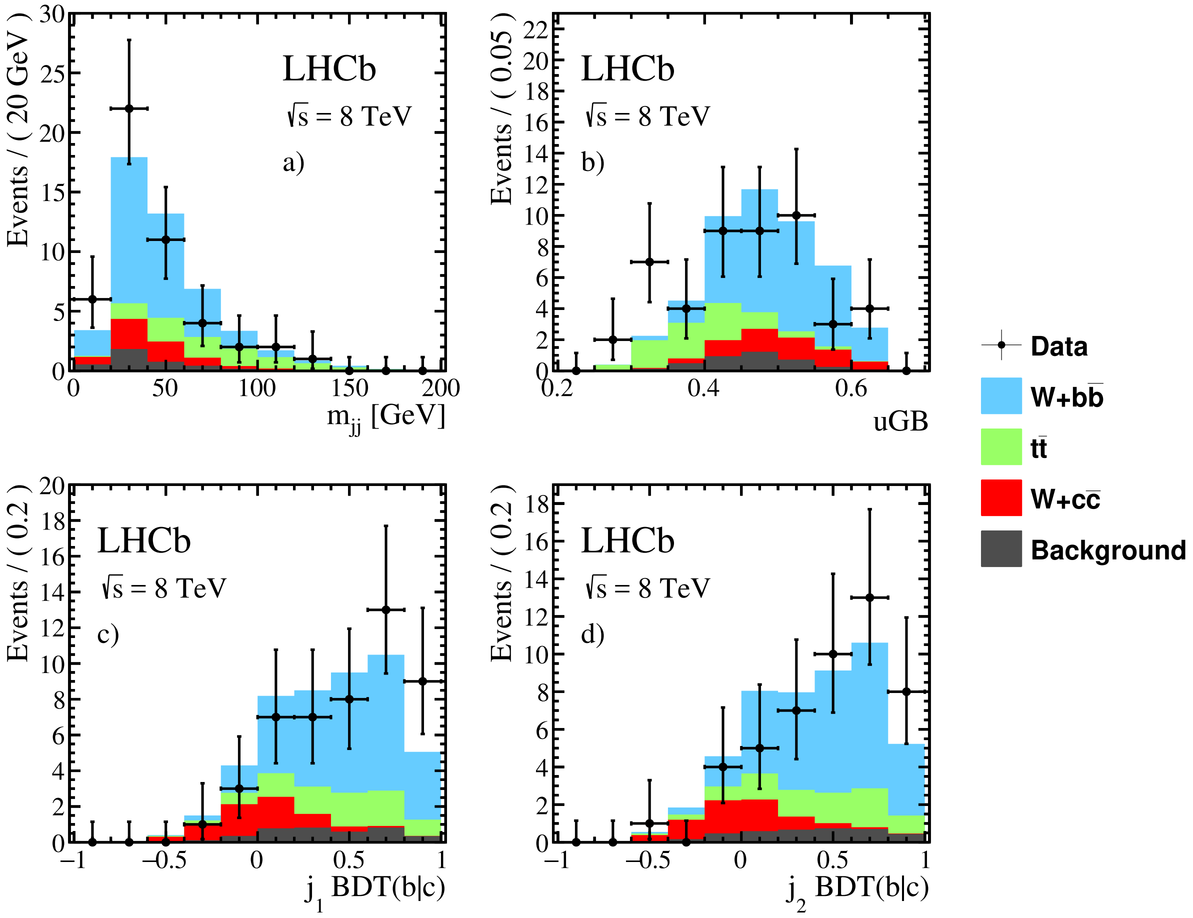

Figure 5.16: Projection of the simultaneous data 4D-fit result on the μ^- sample: a) the dijet mass; b) the *uGB* response; c) the $\mathrm{BDT}_{b|c}$ of the first jet; and d) the $\mathrm{BDT}_{b|c}$ of the second jet.

Table 5.10: Cross-sections in pb.

Process	Expected	Observed
$W^+ + b\bar{b}$	$0.081^{+0.022}_{-0.013}$ (scale) $^{+0.040}_{-0.018}$ (total)	$0.121^{+0.019}_{-0.018}$(stat.)$^{+0.029}_{-0.020}$(syst.)
$W^- + b\bar{b}$	$0.056^{+0.014}_{-0.010}$ scale) $^{+0.018}_{-0.013}$ (total)	$0.093^{+0.018}_{-0.017}$(stat.)$^{+0.023}_{-0.016}$(syst.)
$W^+ + c\bar{c}$	$0.123^{+0.034}_{-0.020}$ scale) $^{+0.060}_{-0.027}$ (total)	$0.24^{+0.08}_{-0.07}$(stat.)$^{+0.08}_{-0.04}$(syst.)
$W^- + c\bar{c}$	$0.084^{+0.021}_{-0.015}$ scale) $^{+0.027}_{-0.020}$ (total)	$0.133^{+0.073}_{-0.062}$(stat.)$^{+0.050}_{-0.022}$(syst.)
$t\bar{t}$	$0.045^{+0.008}_{-0.007}$scale) $^{+0.012}_{-0.010}$(total)	$0.05^{+0.02}_{-0.01}$(stat.)$^{+0.02}_{-0.01}$(syst.)

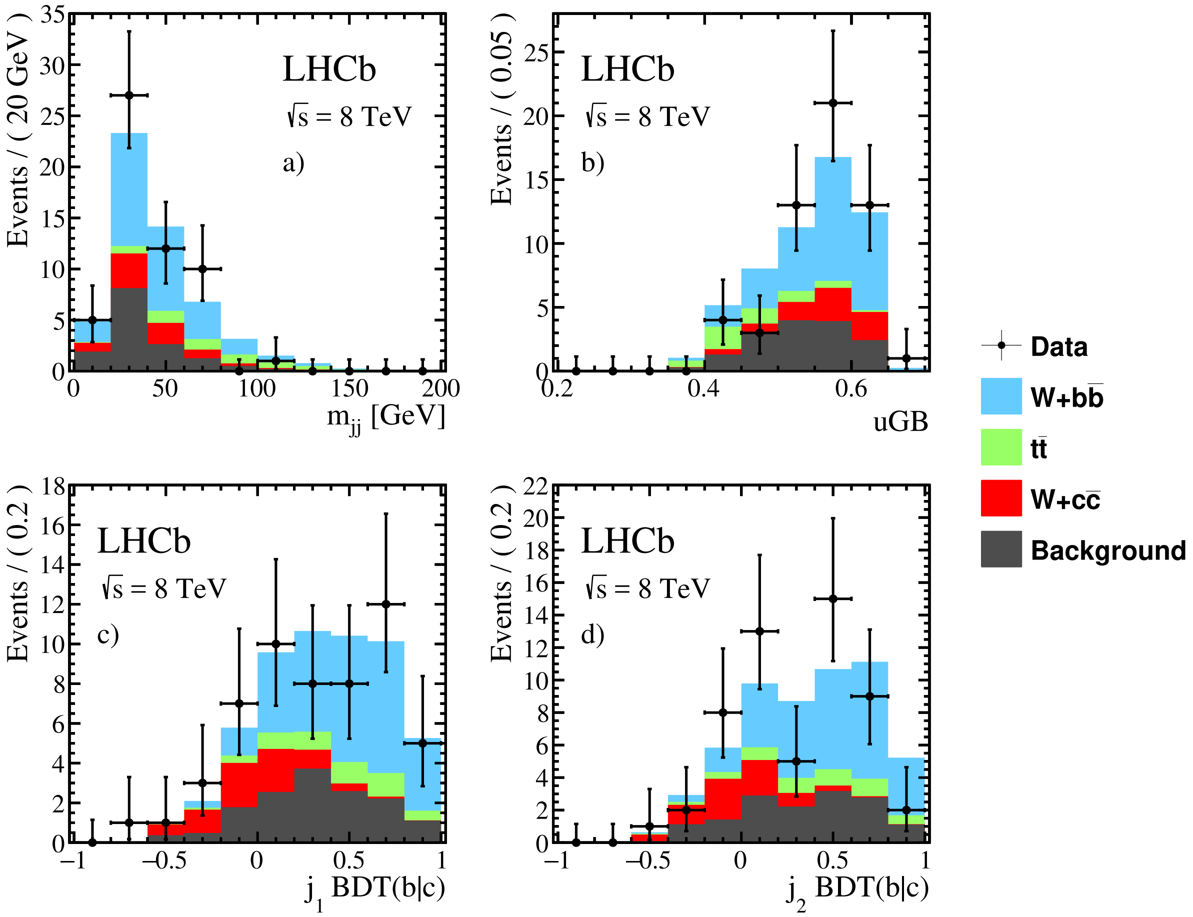

Figure 5.17: Projection of the simultaneous data 4D-fit result on the e^+ sample: a) the dijet mass; b) the uGB response; c) the $\mathrm{BDT}_{b|c}$ of the first jet; and d) the $\mathrm{BDT}_{b|c}$ of the second jet.

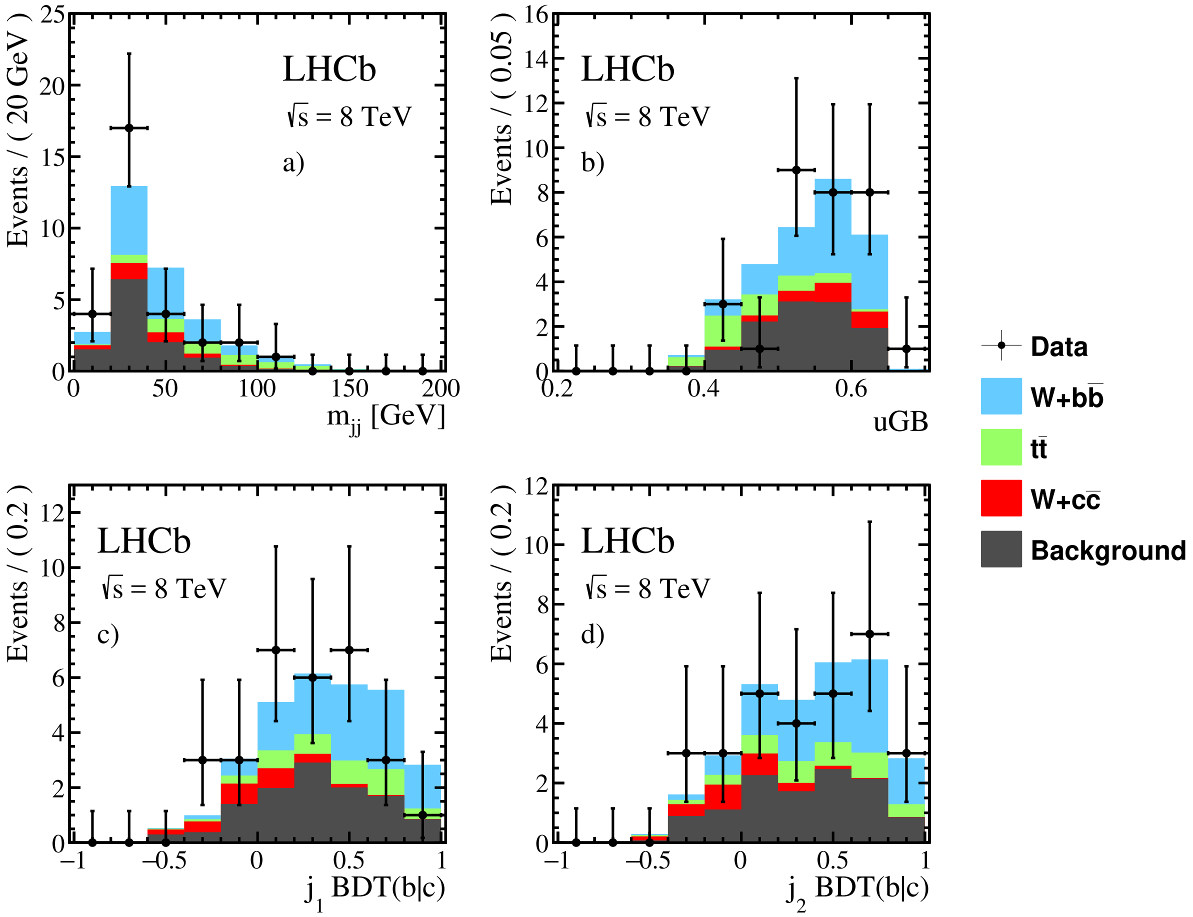

Figure 5.18: Projection of the simultaneous data 4D-fit result on the e^- sample: a) the dijet mass; b) the *uGB* response; c) the $\mathrm{BDT}_{b|c}$ of the first jet; and d) the $\mathrm{BDT}_{b|c}$ of the second jet.

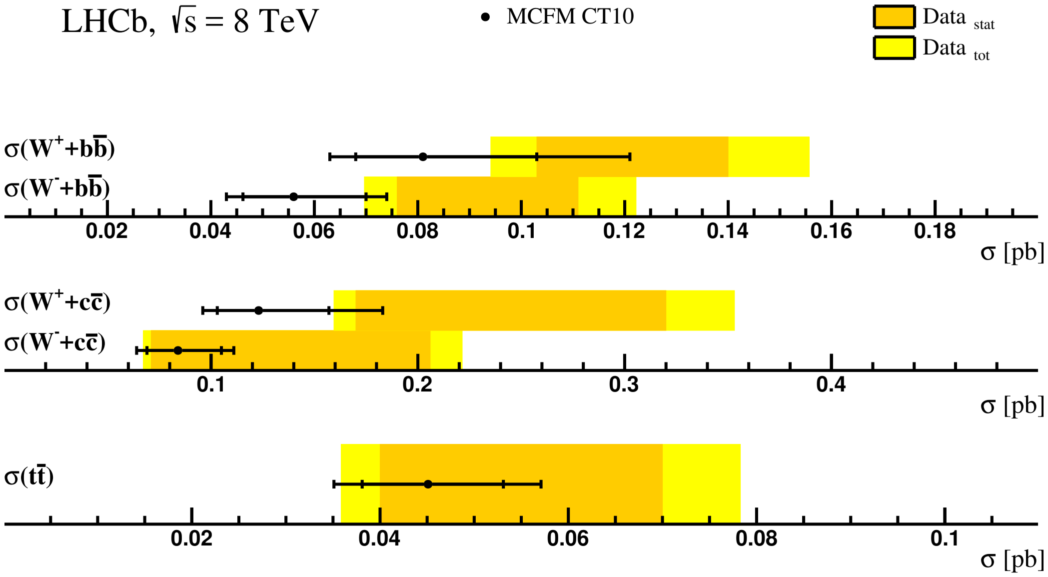

Figure 5.19: Observed cross-sections in pb for all the measured processes compared to expected ones from MCFM CT10, where the inner error bars are due to the scale errors.

5.12 CL_s method

The CL_s method [93] is used in particle physics to exclude the presence of a signal in data with a certain Confidence Level (CL).

In a search the null hypothesis is that the signal does not exist, while the alternate hypothesis is that the signal exists. To distinguish between the two hypothesis a test-statistics Q can be defined. Several choices are possible for Q but, according to the Neyman-Pearson lemma, the most powerful test-statistics is the likelihood ratio:

$$Q = \frac{L(H_{s+b})}{L(H_b)},$$

where $L(H_{s+b})$ is the likelihood for the null hypothesis while $L(H_b)$ is the likelihood for the alternate hypothesis. The test-statistics Q is a random observables, function of the experiment outcome, which distribution can depend from one or more signal model parameters, like the Higgs mass or cross section. From the result of the experiment the observed test-statistics can be computed (Q_{obs}). The probability to obtain a Q less or equal of Q_{obs} in the signal plus background hypothesis is

$$P_{s+b}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{s+b}}{dQ} dQ,$$

where $\frac{dP_{s+b}}{dQ}$ is the probabilty distribution function (*pdf*) for Q in the signal plus background hypothesis. In the practice $P_{s+b}(Q \leq Q_{obs})$ is computed using pseudo-experiments, starting from the signal and background expected models. Then CL_{s+b} is defined as

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}).$$

In an analogous way CL_b is defined as

$$CL_b = P_b(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_b}{dQ} dQ,$$

where $\frac{dP_b}{dQ}$ is the *pdf* for Q in the signal plus background. Finally the CL_s value can be computed with the following re-normalisation:

$$CL_s = \frac{CL_{s+b}}{CL_s}.$$

Given a certain Confidence Level, the signal hypothesis is excluded if

$$1 - CL_s \leq CL.$$

For the case of an upper limit on the production cross section, one compute the CL_s for different cross section hypotheses. The upper limit is the cross section that satisfies the equivalence:

$$1 - CL_s = CL.$$

5.13 Upper limits on the $V{+}H^0(\rightarrow b\bar{b})$ and $V{+}H^0(\rightarrow c\bar{c})$ productions

Since the measured cross sections of $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$ are compatible with the SM prediction and no enhanced Higgs production is expected, upper limits are set on the $(W/Z)H^0 \rightarrow b\bar{b}$ and $(W/Z)H^0 \rightarrow c\bar{c}$ productions. The limit are calculated assuming a H^0 with a mass of 125 GeV and with SM properties.

The search for $(W/Z)H^0 \rightarrow b\bar{b}$ is performed using the dataset selected as described in Section 5.3. To search for $(W/Z)H^0 \rightarrow c\bar{c}$, an additional requirement on the SV-tagger algorithm variable that separates b jets from c jets ($\mathrm{BDT}_{b|c}$) is applied for both jets ($\mathrm{BDT}_{b|c} < 0.2$). This requirement removes about 90% of $(W/Z)H^0 \rightarrow b\bar{b}$ events while retaining 62% of $(W/Z)H^0 \rightarrow c\bar{c}$ events. The simulated $(W/Z)H^0 \rightarrow b\bar{b}$ and $(W/Z)H^0 \rightarrow c\bar{c}$ distributions of $\mathrm{BDT}_{b|c}$ for both jets are shown in figure. 5.20.

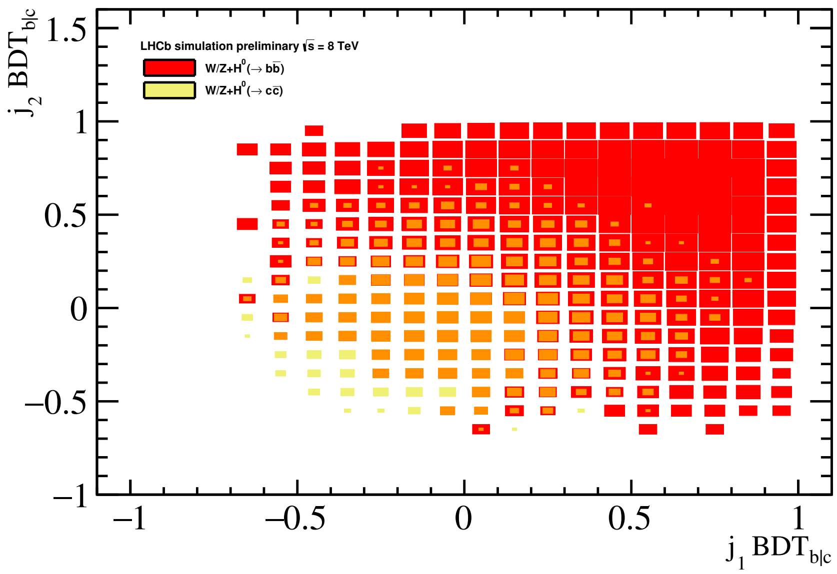

Figure 5.20: SV-tagger algorithm $\mathrm{BDT}_{b|c}$ for the highest jet p_{T} versus $\mathrm{BDT}_{b|c}$ for second highest jet p_{T} distributions. Distributions are obtained from $(W/Z)H^0 \rightarrow b\bar{b}$ (blue) and $(W/Z)H^0 \rightarrow c\bar{c}$ (red) simulation. The areas of the rectangles are proportional to the number of entries which is equal to the SM expectation after the selection cuts.

Four observables are used in the limit computation:

- the dijet invariant mass m_{jj};
- $uGB(W/Z + H^0 \text{ vs } W + b\bar{b})$;
- $uGB(W/Z + H^0 \text{ vs } t\bar{t})$;

- the lepton flavour, μ or e.

The observed and expected distributions of m_{jj}, $uGB(W/Z + H^0 \text{ vs } W + b\bar{b})$ and $uGB(W/Z + H^0 \text{ vs } t\bar{t})$ are shown in figures 5.21 and 5.22 for muon+dijet and electron+dijet events used in the $(W/Z)H^0 \to b\bar{b}$ search. The data is compatible with the background expectation within the statistical uncertainties. The expected distributions of events with the selection used in the $(W/Z)H^0 \to c\bar{c}$ search are shown in Figs. 5.23 and 5.24. In this case no data events are observed.

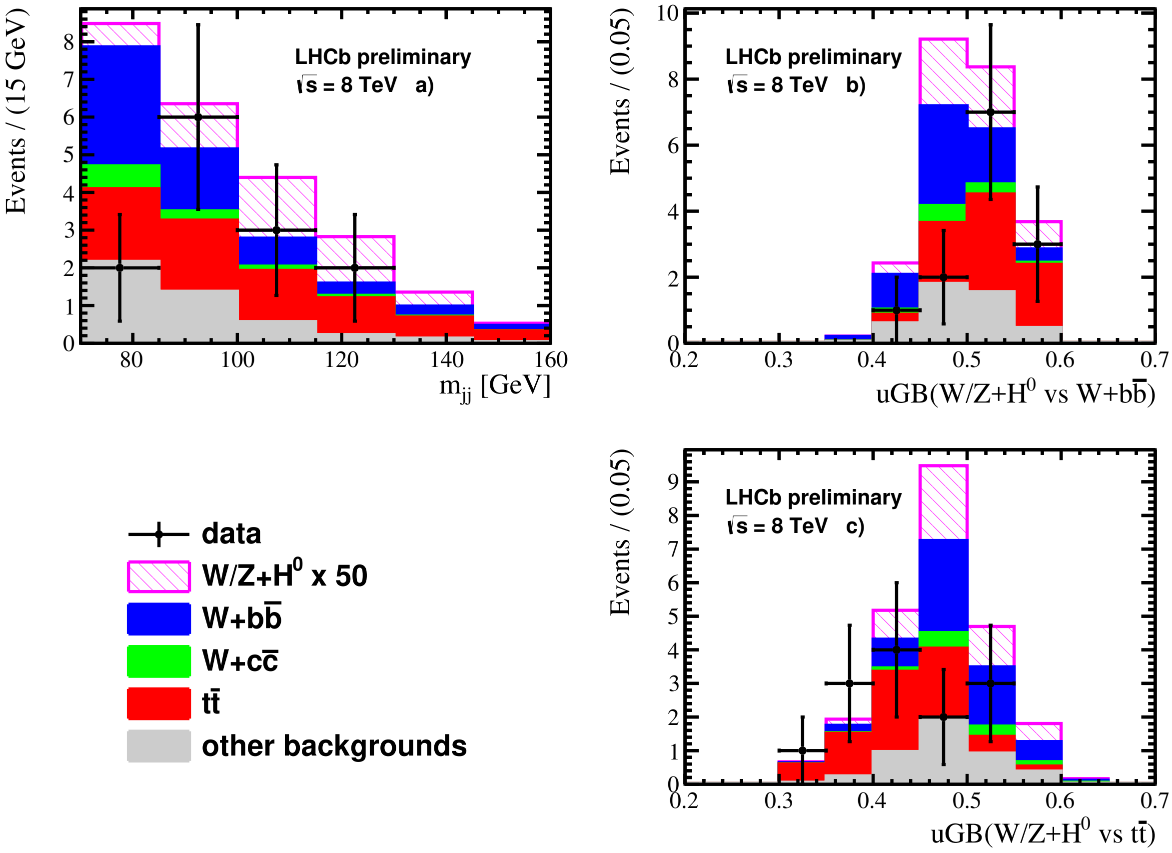

Figure 5.21: a) Dijet mass, b) $uGB(W/Z + H^0 \text{ vs } W + b\bar{b})$ and c) $uGB(W/Z + H^0 \text{ vs } t\bar{t})$ distributions of selected data events with dijet mass greater than 70 GeV for the muon sample used in the $H^0 \to b\bar{b}$ limit. The SM signal and background prediction is also shown. The $H^0 \to b\bar{b}$ yield (in magenta) is multiplied by a factor 50. The compatibility between the observed data and the expected theory yield is below 2 Gaussian standard deviations.

In order to improve the sensitivity of the search, a transformation is applied to the uGB two dimensional plane $[uGB(W/Z + H^0 \text{ vs } W + b\bar{b}), uGB(W/Z + H^0 \text{ vs } t\bar{t})]$. The goal of this transformation is to make the Higgs signal uniformly distributed in the uGB plane. The procedure followed is similar to that described in [94] and it is explained next:

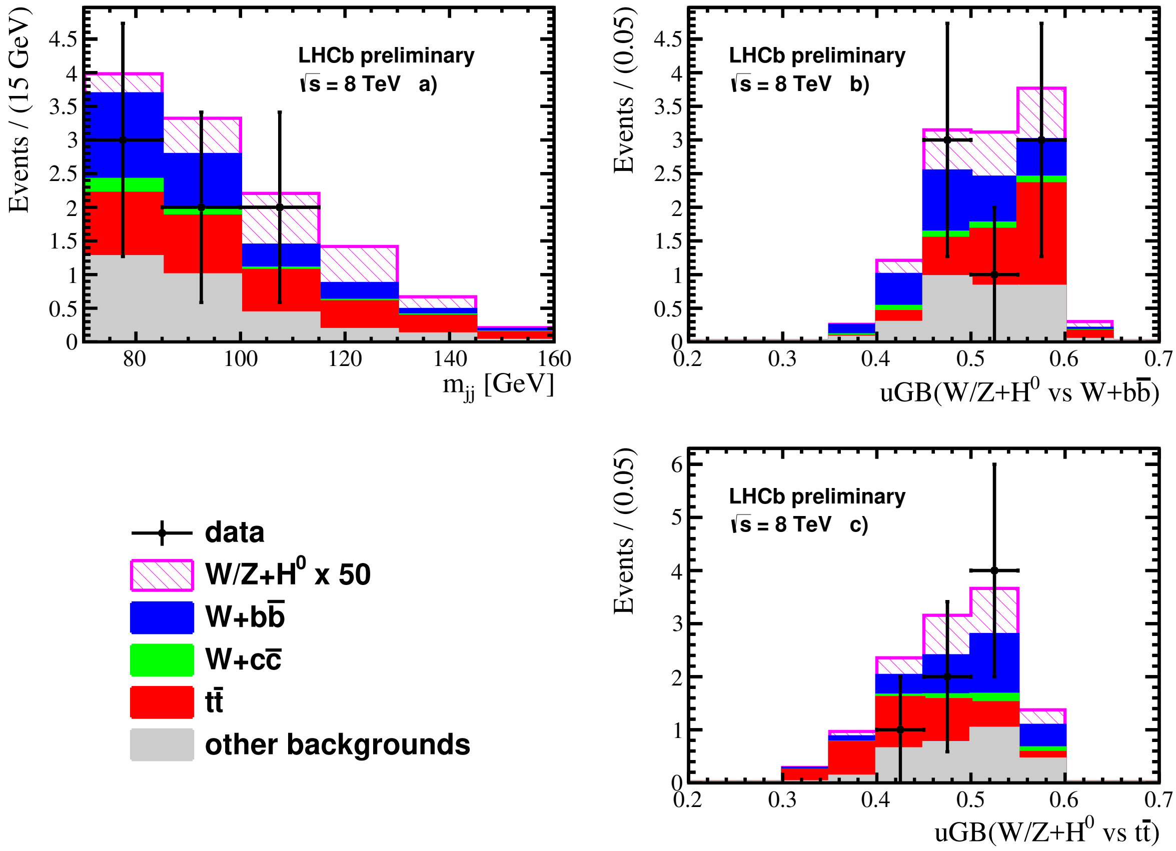

Figure 5.22: a) Dijet mass, b) $uGB(W/Z + H^0$ vs $W + b\bar{b})$ and c) $uGB(W/Z + H^0$ vs $t\bar{t})$ distributions of selected data events with dijet mass greater than 70 GeV for the electron sample used in the $H^0 \to b\bar{b}$ limit. The SM signal and background prediction is also shown. The $H^0 \to b\bar{b}$ yield (in magenta) is multiplied by a factor 50. The data is compatible with the background expectation.

1. the following transformation is applied to each of the uGB variables:

   $$G_i(x_i) = \sqrt{2} \cdot erf^{-1}(2 \cdot U(x_i) - 1),$$

   where $x_i =$$uGB(W/Z + H^0$ vs $W + b\bar{b})$ or $uGB(W/Z + H^0$ vs $t\bar{t})$, $erf^{-1}(x)$ is the inverse of the error function and

   $$U(x_i) = \frac{\int_{-\infty}^{x_i} \rho_i(x_i')dx_i'}{\int_{-\infty}^{+\infty} \rho_i(x_i')dx_i'}$$

   where $\rho(x_i)$ is the x_i 1-dimensional distribution in the $V + H^0$ MC; the consequence of this transformation is that the G_1 and G_2 variables are distributed as gaussians in the $V + H^0$ MC;

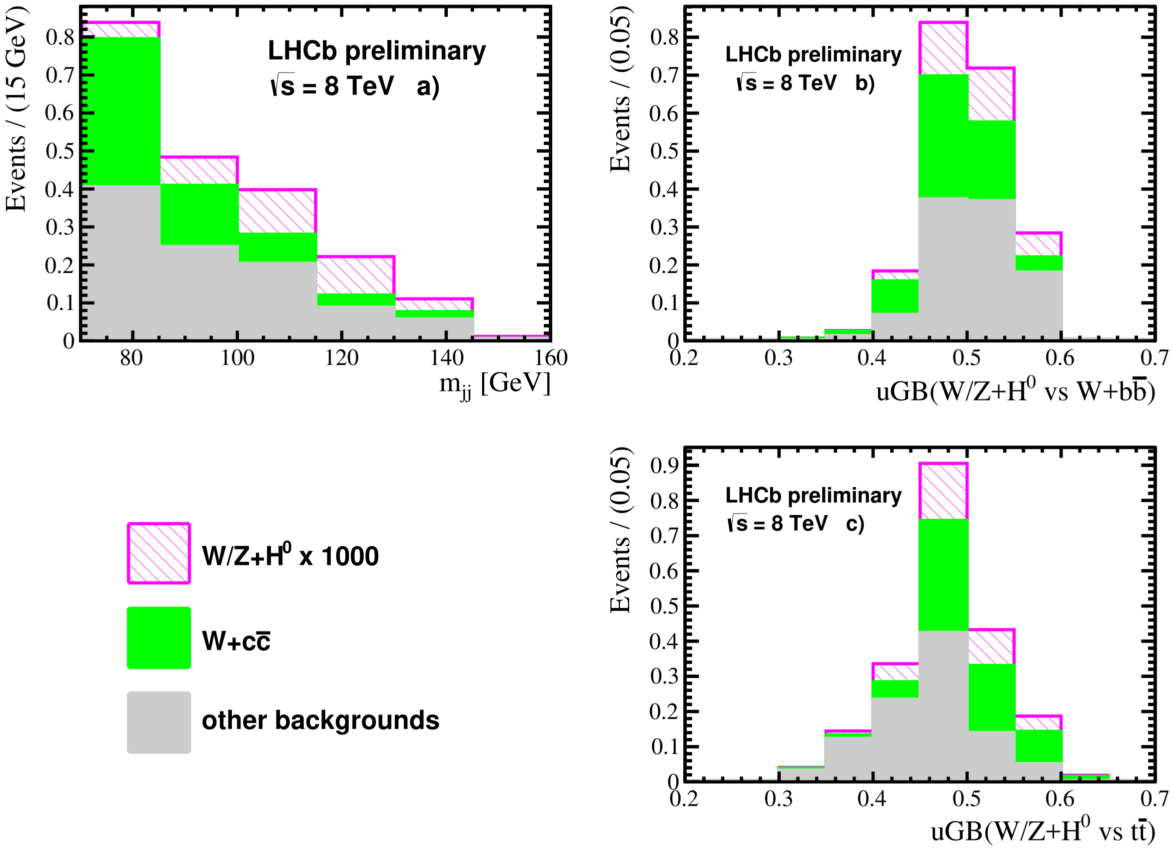

Figure 5.23: SM prediction of a) Dijet mass, b) $uGB(W/Z + H^0$ vs $W+b\bar{b})$ and c) $uGB(W/Z + H^0$ vs $t\bar{t})$ distributions of events with dijet mass greater than 70 GeV for the muon sample used in the $H^0 \to c\bar{c}$ limit. The $H^0 \to c\bar{c}$ yield (in magenta) is multiplied by a factor of 1000. There are no data events passing the selection criteria for this sample.

2. the following transformation is applied to (G_1, G_2):

$$q_1 = \frac{1}{\sigma_1\sqrt{2}}(G_1 + G_2),$$

$$q_2 = \frac{1}{\sigma_2\sqrt{2}}(G_1 - G_2),$$

   where σ_i are the widths of the G_i gaussian distribution in the $V + H^0$ MC. If $uGB(W/Z + H^0$ vs $W + b\bar{b})$ and $uGB(W/Z + H^0$ vs $t\bar{t})$ are perfectly uncorrelated in MC then (q_1, q_2) is distributed as an uncorrelated 2-dimensional gaussian, with mean equal to 0 and width equal to 1 in the two dimensions;

3. to obtain a uniform distribution in the $V + H^0$ MC the following transformation is

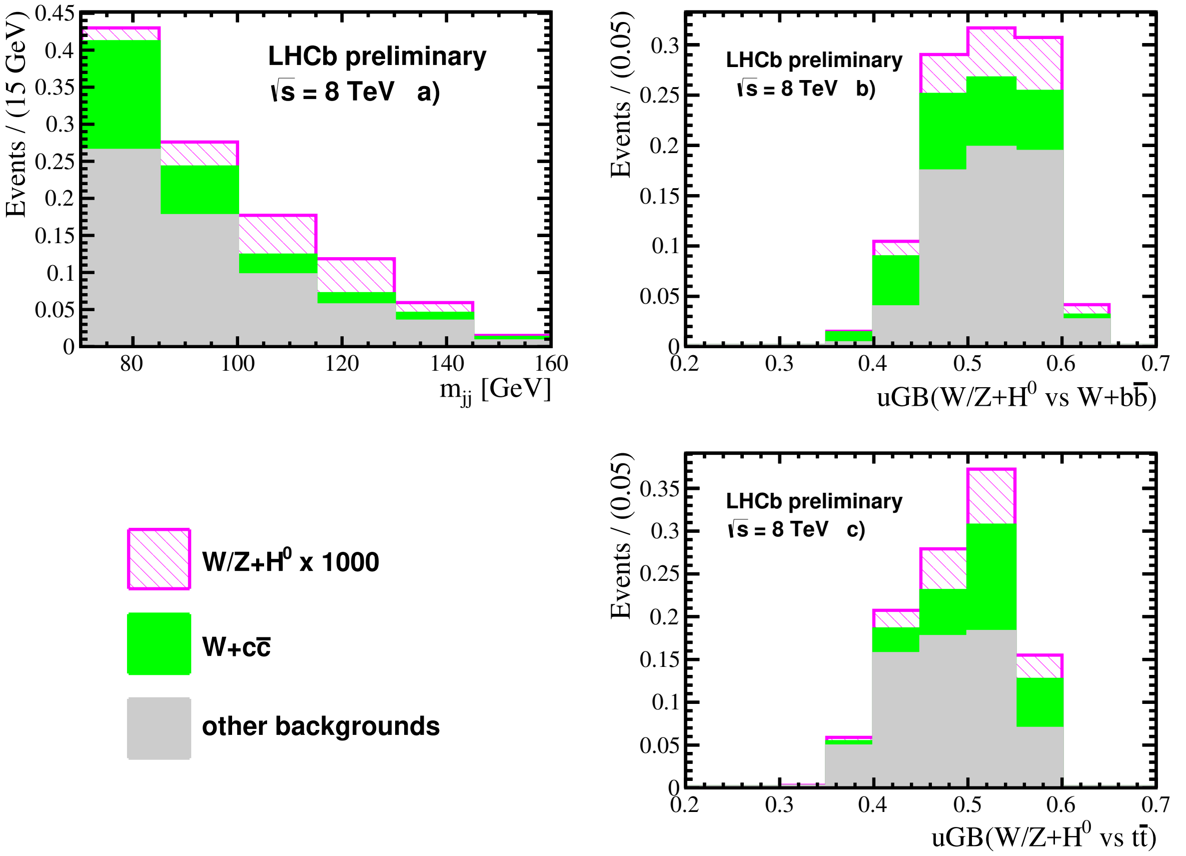

Figure 5.24: SM prediction of a) Dijet mass, b) $uGB(W/Z + H^0$ vs $W + b\bar{b})$ and c) $uGB(W/Z + H^0$ vs $t\bar{t})$ distributions of events with dijet mass greater than 70 GeV for the electron sample used in the $H^0 \to c\bar{c}$ limit. The $H^0 \to c\bar{c}$ yield (in magenta) is multiplied by a factor of 1000. There are no data events passing the selection criteria for this sample.

applied:

$$uGB_i = \frac{1}{2}\left[1 - erf\left(\frac{q_i}{\sqrt{2}}\right)\right].$$

The distributions of the $(W/Z)H^0 \to b\bar{b}$, $W + b\bar{b}$ and $t\bar{t}$ simulated events in the transformed space (uGB_1, uGB_2) are shown in figure 5.25. It can be seen that $(W/Z)H^0 \to b\bar{b}$ is uniform but $W + b\bar{b}$ and $t\bar{t}$ accumulates to the corner.

The upper limits on the production cross section of the Higgs boson are set using the CL_s method. To compute the limit a 4-dimensional $(m_{jj}, uGB_1, uGB_2, \ell)$ histogram is filled with data events. The following binning scheme is used:

- five bins of m_{jj}: one for events with $m_{jj} < 70$ GeV, three m_{jj} bins of equal width in the range [70,115] GeV, and one *bin* for events with $m_{jj} > 115$ GeV;

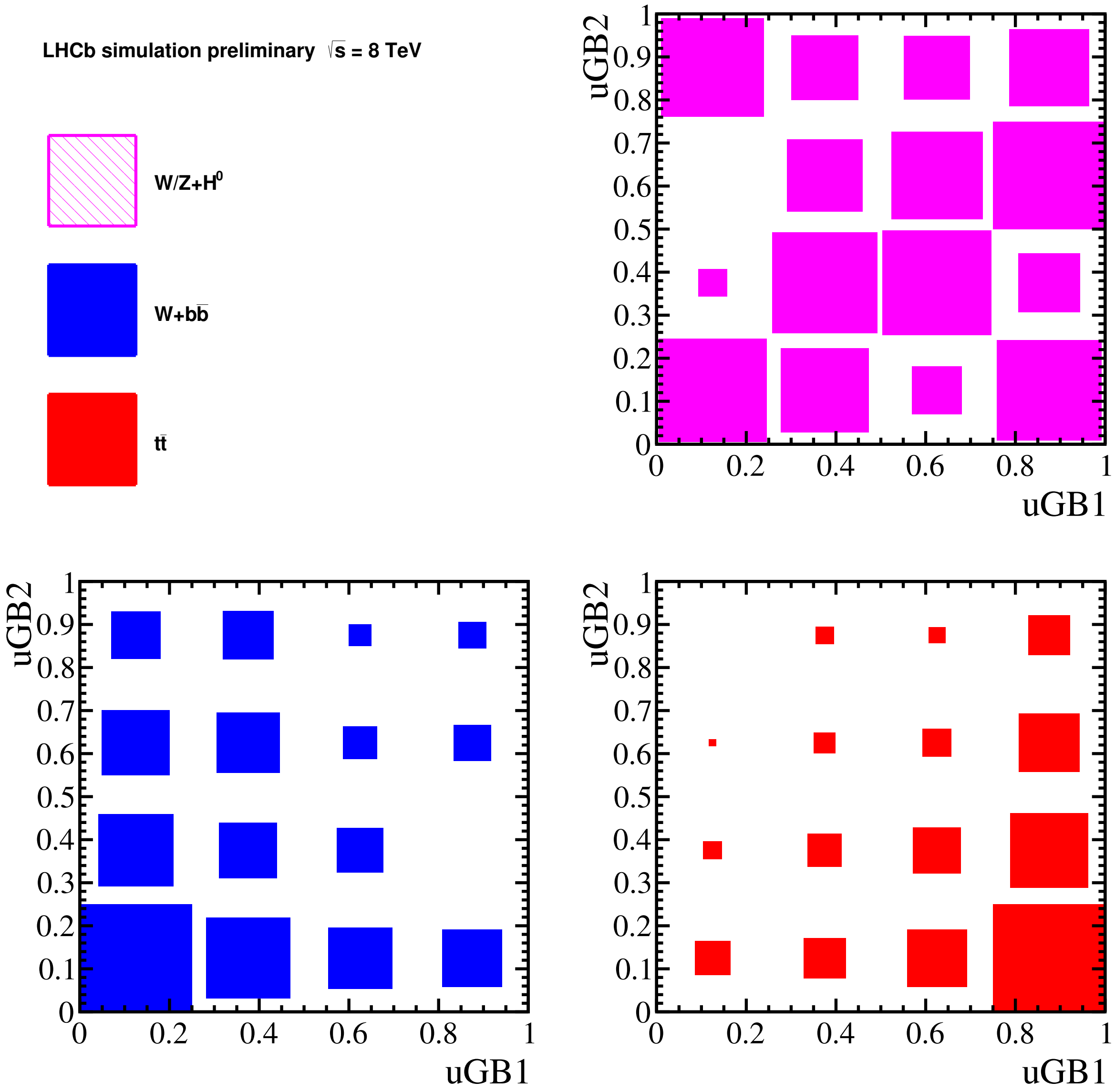

Figure 5.25: Distributions of $(W/Z)H^0 \to b\bar{b}$, $Wb\bar{b}$ and $t\bar{t}$ simulated events in the transformed extscuGB space.

- four bins of uGB_1 with equal width in the [0,1] range;
- four bins of uGB_2 with equal width in the [0,1] range;
- two bins for each lepton flavour, e or μ.

The total number of bins is $5 \times 4 \times 4 \times 2 = 160$. Three histograms with this binning scheme are used in input to the CL_s computation:

- the data histogram;
- the background histogram: it is obtained using MC templates and the expected yields for the background processes described in Section 5.2. The QCD background template and yield is obtained as explained in Section 5.6.
- the signal histogram: obtained as template using $(W/Z)H^0 \to b\bar{b}$ (or $(W/Z)H^0 \to c\bar{c}$) MC events.

The CL_s are computed for different $V + H^0$ cross section hypotheses and upper limits are set at 95% Confidence Level (CL).

The systematic uncertainties discussed in Section 4.5 are introduced in the CL_s computation as nuisance parameters. The nuisance parameters may affect:

- the templates and the yields used to build the background histogram;
- the signal histogram;
- the expected number of signal events for different cross section hypotheses.

The expected and observed CL_s limits as a function of the tested cross sections normalised to the SM prediction for $(W/Z)H^0 \to b\bar{b}$ and $(W/Z)H^0 \to c\bar{c}$ are shown in figure 5.26 and 5.27, respectively. The expected upper limit on the $(W/Z)H^0 \to b\bar{b}$ production at

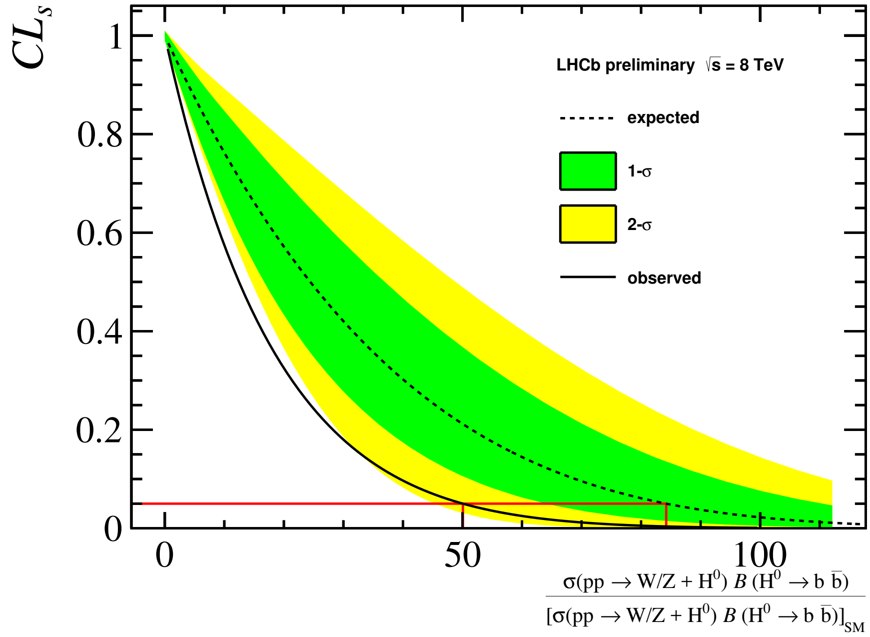

Figure 5.26: Observed and expected CL_s and 95% CL upper limit for the $V + H^0(\to b\bar{b})$. The 0.05 CL_s level is indicated by the red horizontal line.

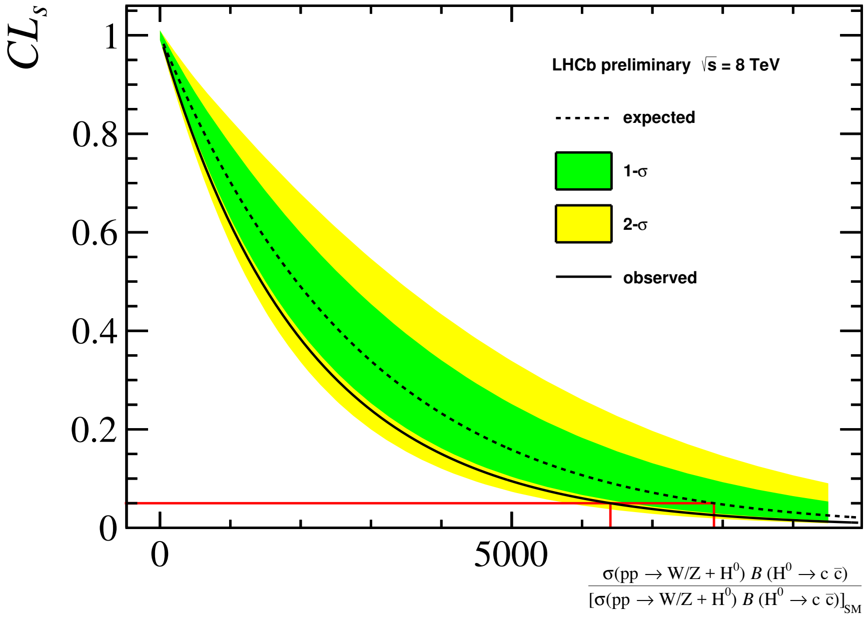

Figure 5.27: Observed and expected CL_s and 95% CL upper limit for the $V + H^0(\to c\bar{c})$. The 0.05 CL_s level is indicated by the red horizontal line.

95 (90) % CL is 84 (69) times the SM expectation, while the observed upper limit is 50 (40) times the SM expectation. Therefore the observed limit on $\sigma(pp \to W/Z + H^0)\mathcal{B}(H^0 \to b\bar{b})$ in the LHCb acceptance (two b quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$), is:

$$\sigma(pp \to W/Z + H^0)\mathcal{B}(H^0 \to b\bar{b}) < 1.6\ (1.3)\ \text{pb},\ \text{at 95 (90)\% CL and at 8 TeV}.$$

The expected upper limit on the $(W/Z)H^0 \to c\bar{c}$ production at 95% (90%) CL is 7900 (6200) times the SM expectation, while the observed upper limit is 6400 (4900) times the SM expectation. Therefore the observed limit on $\sigma(pp \to W/Z + H^0)\mathcal{B}(H^0 \to c\bar{c})$ in the LHCb acceptance (two c quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$) is

$$\sigma(pp \to W/Z + H^0)\mathcal{B}(H^0 \to c\bar{c}) < 9.4\ (7.2)\ \text{pb},\ \text{at 95 (90)\% CL and at 8 TeV}.$$

Chapter 6

Search for the inclusive Higgs production and future prospects

6.1 Upper limit on $\sigma(pp \rightarrow X + H^0)BR(H^0 \rightarrow b\bar{b})$

The LHCb sensitivity to the inclusive SM $H^0 \rightarrow b\bar{b}$ observation is studied in this Section. The dataset collected in *pp* collisions at a centre-of-mass energy of 8 TeV is used in this search.

The same *b*-dijet selection requirements described in Section 4.2.2 for the $Z \rightarrow b\bar{b}$ cross section measurements are applied. In this case the recoil jet selection is not applied, since it would reduce the $H^0 \rightarrow b\bar{b}$ significance. Moreover the mass window where the Higgs is searched is [60,180] GeV[1] instead of the [45,165] GeV window used in the $Z \rightarrow b\bar{b}$ measurement.

A gluon-gluon fusion $H \rightarrow b\bar{b}$ MC sample and a vector boson fusion $H^0 \rightarrow b\bar{b}$ MC sample generated with PYTHIA 8 are used to study the signal properties. In these samples the Higgs has a mass of 125 GeV and SM properties. The other Higgs production mechanisms are considered negligible. Figure 6.1 shows the dijet invariant mass distribution of selected data event superimposed with the mass distribution in $H^0 \rightarrow b\bar{b}$ MC events. Even in this case the dijet invariant mass is the best variable to discriminate the signal from the background. The expected yields of $H \rightarrow b\bar{b}$ and backgrounds are computed as explained in Section 4.2.3 and they are shown in table 6.2. The number of selected data events is also reported.

Since the $H^0 \rightarrow b\bar{b}$ observation is not expected, an upper limit on the inclusive production cross section is set. The strategy used is the following:

1. a BDT is trained with the *uGB* method (for uncorrelation with the dijet invariant mass), to obtain a discriminator optimized for Higgs-background separation;
2. a cut is applied to the *uGB* output, optimized for the signal significance;

---

[1] in this Chapter natural units where $\hbar = c = 1$ are used.

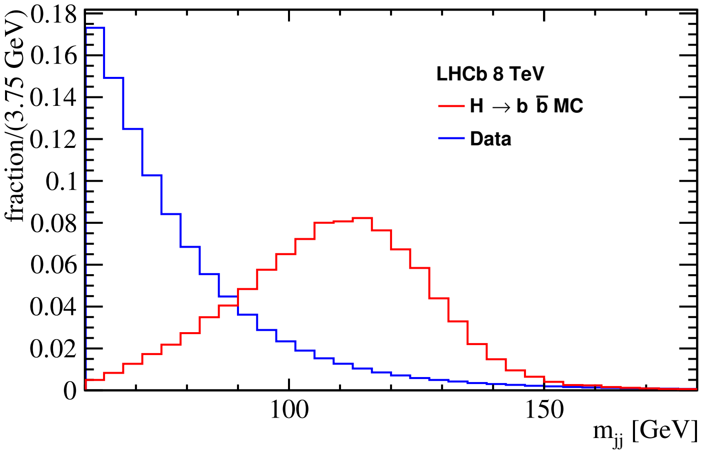

Figure 6.1: Distribution of the dijet invariant mass for data and $H \to b\bar{b}$ MC, both normalized to 1.

Table 6.1: Yield predictions for the studied processes.

Process	Expected yield
$Z \to b\bar{b}$	102263
$Z \to c\bar{c}$	3986
$W \to q\bar{q}'$	1897
$t\bar{t}$	158
$gg \to H^0 \to b\bar{b}$	61.6
VBF $H^0 \to b\bar{b}$	2.26
data	2900878

3. the background distribution is obtained with a fit to the dijet invariant mass distribution in the [60,80] GeV and [150,180] GeV mass sidebands. Then it is extrapolated in the [80,150] region;

4. the limit is set using the CL_s method.

In analogy with the $Z \to b\bar{b}$ analysis strategy, the data sample is considered representative of the background, since the expected Higgs yield is very low. QCD events are the main contributions to the background, followed by $Z \to b\bar{b}$ events. The observables used to discriminate the Higgs signal from the background are the following:

- the absolute b-jets pseudorapidity difference, $|\Delta\eta_{12}|$;
- the elicity angle θ_{el}, defined as the angle between the b-dijet momentum in the

laboratory frame and the momentum of the b-jet with the highest p_T in the couple in the b-dijet rest frame. It is connected to the spin of the resonance;

- the event sphericity S defined in Section 4.3.3;
- the event thrust T defined in Section 4.3.3.

The distributions of these observables in the $gg \to H^0 \to b\bar{b}$ MC are compared with data in figure 6.2. In these plots only events in the [100,130] GeV invariant mass region are shown. They show that the observables provide good discrimination.

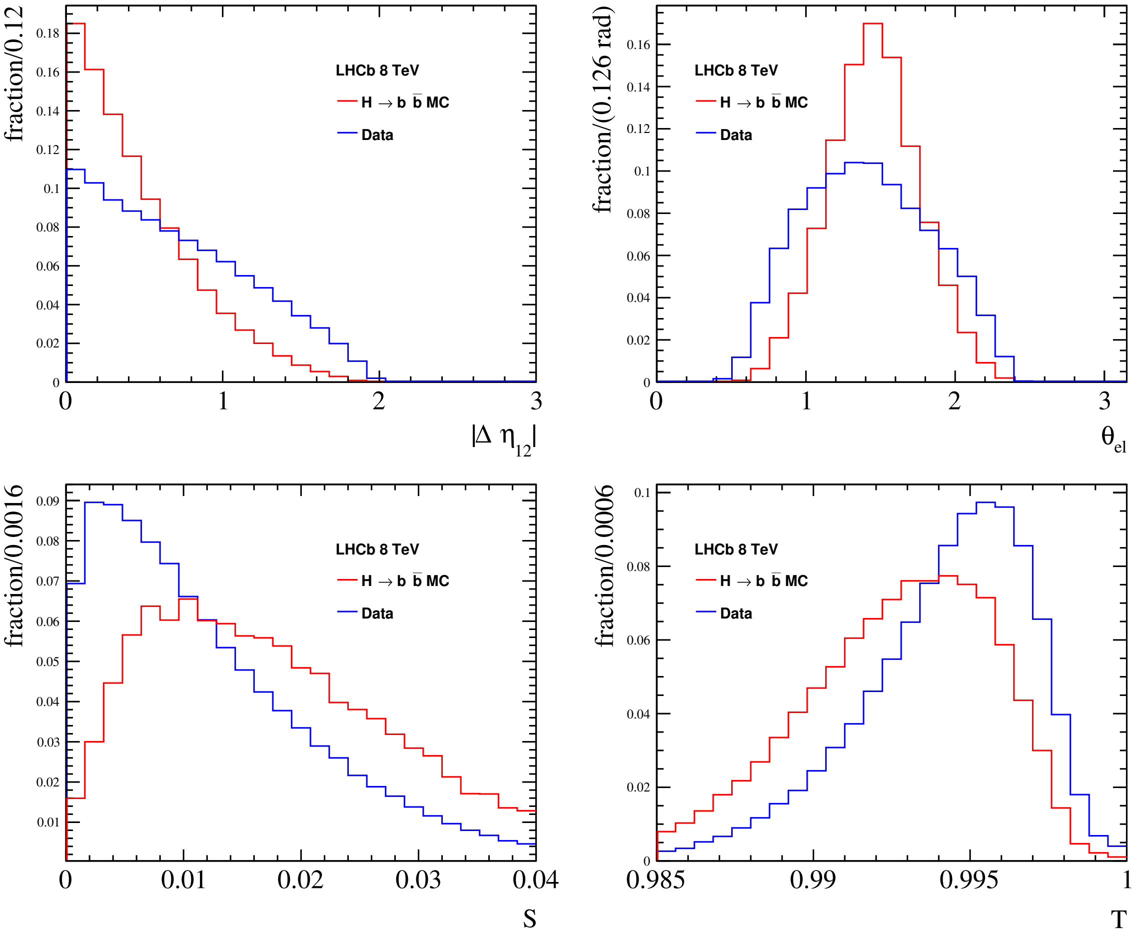

Figure 6.2: Distribution of variables used to discriminate the signal from the background, for data and MC events in the dijet invariant mass region [100,130] GeV.

The observables described in the latter are used in input to a BDT. The BDT is trained using the uGB method described in section 4.3.5, to obtain uncorrelation with respect to the dijet mass. A small fraction of data events ($f = 1\%$) is randomly extracted and used as background sample in the training, then it is removed from the search. The $gg \to H^0 \to b\bar{b}$ MC is used as signal sample. The parameters α and p are set to 10 and 1.5 respectively, as in the $Z \to b\bar{b}$ analysis. A fraction of the training events (20%) is used as test sample to avoid overtraining. The distribution of the discriminator output, called uGB, in data and $gg \to H \to b\bar{b}$ MC is shown in figure 6.3. It provides good discrimination. The signal significance as a function of the uGB cut is computed as

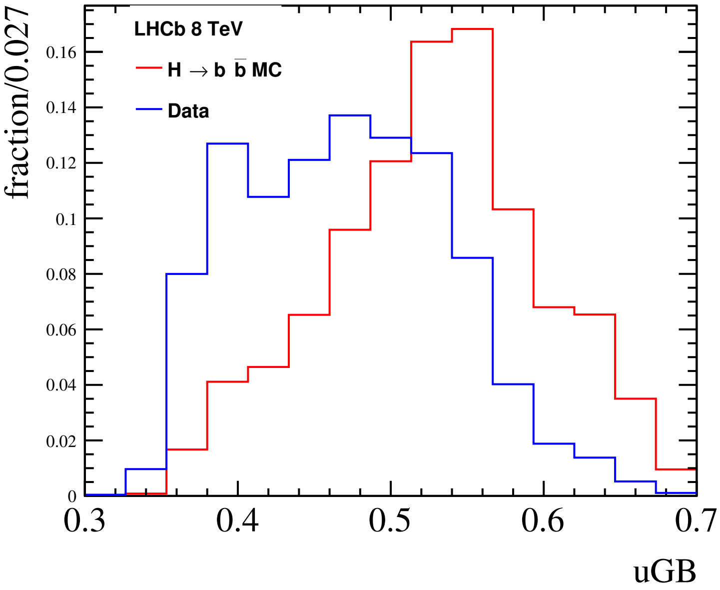

Figure 6.3: Distribution of the uGB output for data and $H \to b\bar{b}$ MC.

$$S(x) = \frac{N^H_{exp}(uGB > x)}{\sqrt{N_{data}(uGB > x)}},$$

where $N^H_{exp}(uGB > x)$ is the number of expected Higgs events that pass the $uGB > x$ cut and $N_{data}(uGB > x)$ is the number of selected data events for $uGB > x$. S as a function of x is shown in figure 6.4. The x value that maximizes S is chosen to define the signal region ($uGB > x_s$ with $x_s = 0.506$), where the limit is computed. The expected yields of $H \to b\bar{b}$ and backgrounds in the signal region are reported in table 6.2.

The dijet invariant mass distribution of events in the $uGB > x_s$ region is used in the limit computation. The analytical description of the background is obtained with a Unbinned Maximum Likelihood fit to the dijet mass distribution of data events in the [60,80] GeV and [150,180] GeV sidebands. The model used in the fit is composed by a combinatorial QCD *pdf* plus a Z *pdf* which includes the contributions from $Z \to b\bar{b}$ and $Z \to c\bar{c}$ processes. Other backgrounds are considered negligible. The Z model is fixed, and its parameters are obtained from a fit to the $Z \to b\bar{b}$ MC dijet mass distribution, using the triple gaussian model described in 4.4.2. The k_{JES} factor is fixed to the measurement

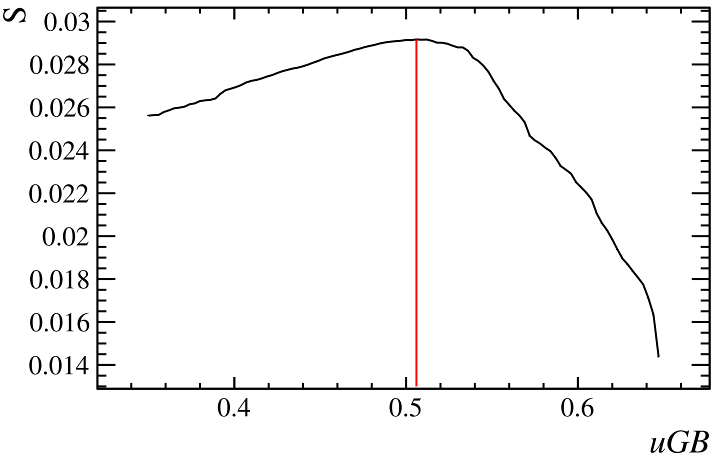

Figure 6.4: Signal significance as a function of the $uGB > x$ cut.

Table 6.2: Yield predictions for the studied processes in the $uGB > x_s$ region.

Process	Signal region exp. yield
$Z \rightarrow b\bar{b}$	36868
$Z \rightarrow c\bar{c}$	1337
$W \rightarrow q\bar{q}'$	490
$t\bar{t}$	134
$gg \rightarrow H^0 \rightarrow b\bar{b}$	39.3
VBF $H^0 \rightarrow b\bar{b}$	1.45
data	969245

performed in Chapter 4 and the Z yield is fixed to that obtained using the $Z \rightarrow b\bar{b}$ cross section measured in Chapter 4. The QCD model is the Pearson IV *pdf* described in Section 4.4.3. The free parameters in the fit are the QCD background coefficients and the QCD yield. The fit result is shown in figure 6.5, where the extrapolation in the [80,150] region is shown. The χ^2/ndof is 2.56 and the QCD background parameters obtained from the fit are reported in table 6.3.

The signal invariant mass model is obtained from a fit to the $gg \rightarrow H^0 \rightarrow b\bar{b}$ MC. MC events are weighted for data/MC corrections as explained in Section 4.2.3. The signal model is a triple gaussian *pdf*. The fit result is shown in figure 6.6 and the fitted parameters are reported in table 6.4. The mass shape is corrected for the k_{JES} factor measured in in Chapter 4.

The inclusive $H^0 \rightarrow b\bar{b}$ upper limit is set using the CL_s technique described in Section 5.13. 32 bins of dijet mass with equal width from 80 to 150 GeV are used. Events in

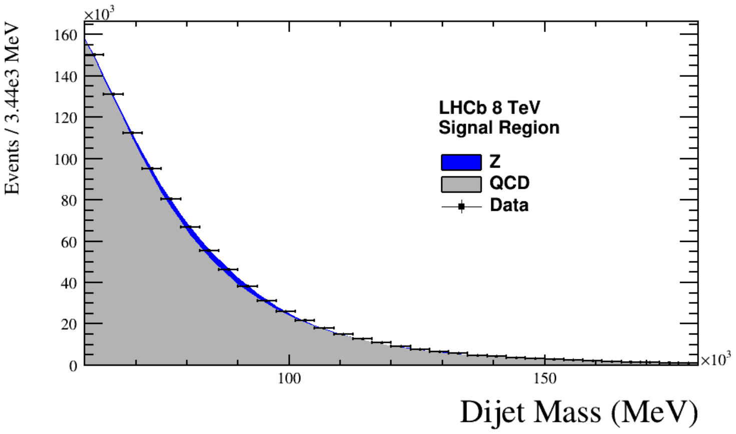

Figure 6.5: Data events superimposed with the background distribution obtained from the fit to the [60,80] GeV and [150,180] GeV dijet invariant mass sidebands.

Table 6.3: QCD mass distribution parameters obtained from the fit to the [60,80] GeV and [150,180] GeV dijet invariant mass sidebands.

Parameter	Result	Stat. uncertainty
a_1	24.39 GeV	4.98 GeV
a_2	3.63 GeV	0.58 GeV
a_3	3.01	0.81
a_4	49.11	2.48

the [60,80] GeV and [150,180] GeV invariant mass sidebands are excluded from the limit computation, since they have been used in the background determination. The following nuisance parameters are considered in the computation:

- Z yield: in the Z yield determination the cross section is varied between the total uncertainty of its measurement;
- b tagging efficiency: the weights of data/MC corrections for b-tagging are varied within their uncertainty in the signal template determination and in the signal efficiency estimation (the latter varies the number of expected signal events for a given cross-section);
- QCD background: a template is obtained by fitting the [60,70] GeV and [160,180] GeV dijet invariant mass sidebands instead of the [60,80] GeV and [150,180] GeV sidebands; in the limit computation the expected QCD background is varied between this one and the nominal one.

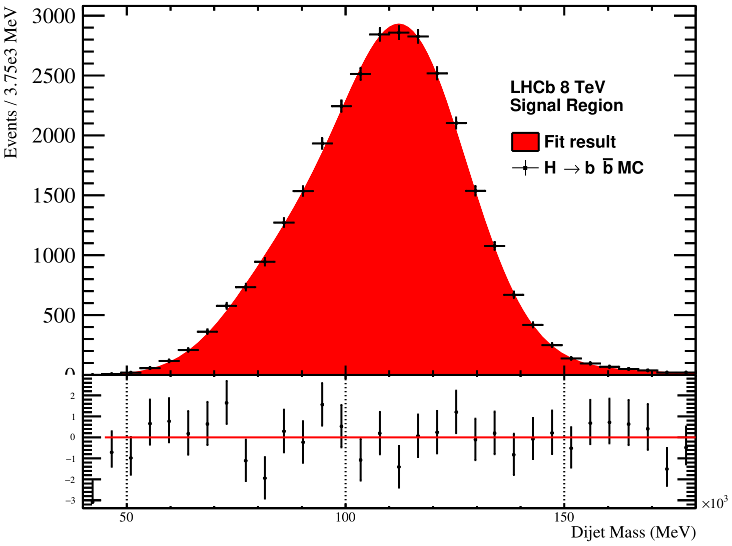

Figure 6.6: Fit to the dijet invariant mass distribution of $gg \to H^0 \to b\bar{b}$ MC.

Table 6.4: Fitted parameters and statistical uncertainties obtained from the $H^0 \to b\bar{b}$ dijet invariant mass fit in the signal region.

Parameter	Fitted Value	Stat. Uncertainty
f_1	0.182	0.020
μ_1	85.15 GeV	0.96 GeV
σ_1	12.81 GeV	0.38 GeV
f_2	0.784	0.031
μ_2	113.07 GeV	0.31 GeV
σ_2	14.63 GeV	0.40 GeV
μ_3	135.00 GeV	6.38 GeV
σ_3	29.17 GeV	28.4 GeV

The other systematics sources, discussed in Section 4.5, are considered negligible. The observed and expected CL_s are shown in figure 6.7, as a function of $\sigma(pp \to X + H^0)BR(H^0 \to b\bar{b})$ normalized to the standard model expectation. The observed (expected) upper limit is 29.8 (41.7) times the Standard Model expectation. The observed and expected upper limits are compatible within the 2-σ uncertainty bar. The upper limit on the cross section is extrapolated in the fiducial phase space where the two b quarks from the Higgs decay are emitted in the LHCb acceptance ($2 < \eta < 5$):

$$\sigma(pp \to X + H^0)\mathcal{B}(H^0 \to b\bar{b}) < 19.4\ (16.3)\ \text{pb},\ \text{at 95 (90)\% CL and at 8 TeV}.$$

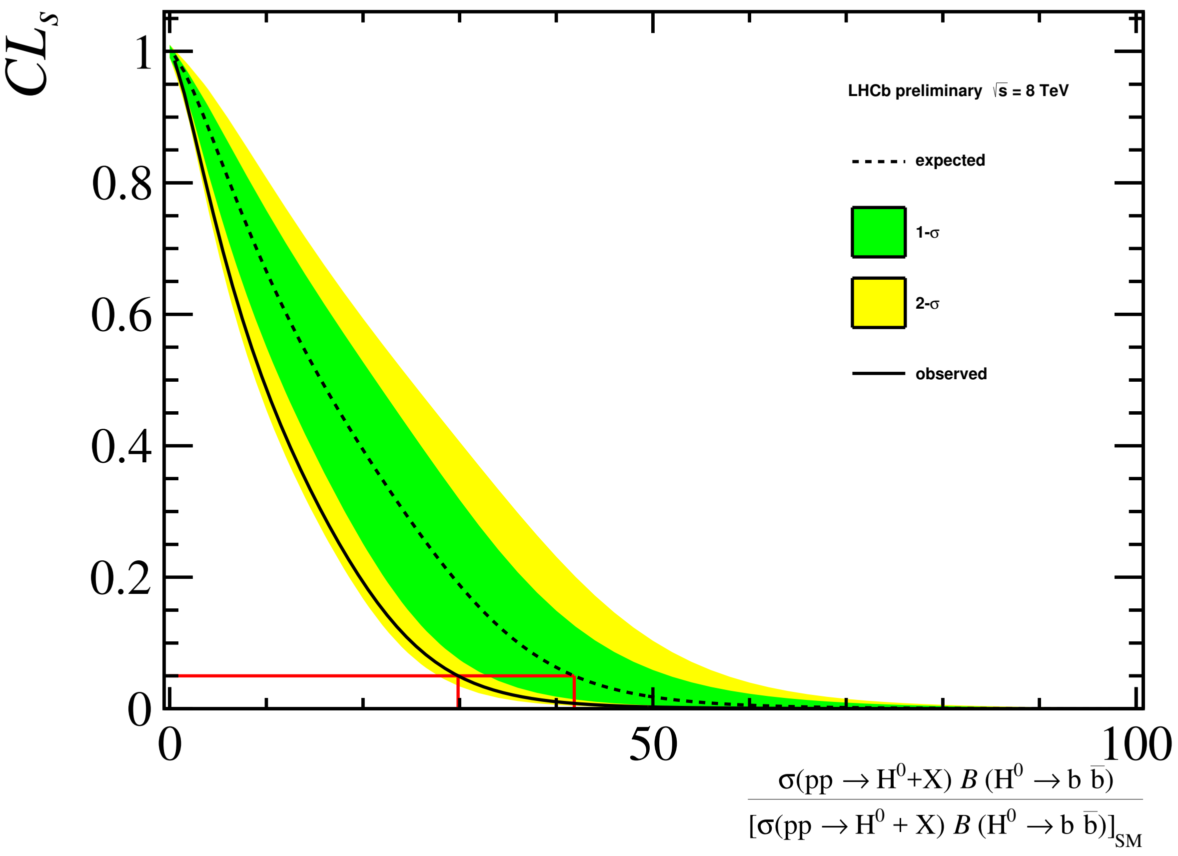

Figure 6.7: Observed and expected CL_s and 95% CL upper limit for the inclusive $H^0 \to b\bar{b}$ production. The 0.05 CL_s level is indicated by the red horizontal line.

6.2 Future prospects

In Section 5.13 and 6.1 direct experimental limits have been set on the $V + H^0(\to b\bar{b})$ production and on the inclusive $H^0 \to b\bar{b}$ production. The upper limits are not competitive with those obtained by ATLAS and CMS by studying the $V + H^0(\to b\bar{b})$ production in [28] and [29]. In this Section prospects on the LHCb sensitivity in the current LHC data taking at 13 TeV in the centre-of-mass (Run II) are discussed.

The upper limit roughly scales with the inverse of the Higgs significance S:

$$S = \frac{N_H}{\sqrt{N_{bb} + N_H}} \simeq \frac{N_H}{\sqrt{N_{bb}}},$$

Where N_H is the number of selected $H^0 \to b\bar{b}$ events and N_{bb} is the number of $b\bar{b}$ events.

Therefore the scale factor between the 13 TeV upper limit and the 8 TeV upper limit is:

$$\begin{aligned} f = \frac{S(8\text{ TeV})}{S(13\text{ TeV})} &= \frac{N_H(8\text{ TeV})}{N_H(13\text{ TeV})} \cdot \sqrt{\frac{N_{bb}(13\text{ TeV})}{N_{bb}(8\text{ TeV})}} \\ &= \frac{\sigma_H(8\text{ TeV})}{\sigma_H(13\text{ TeV})} \cdot \frac{\epsilon_H(8\text{ TeV})}{\epsilon_H(13\text{ TeV})} \cdot \sqrt{\frac{\sigma_{bb}(13\text{ TeV})}{\sigma_{bb}(8\text{ TeV})}} \cdot \sqrt{\frac{\epsilon_{bb}(13\text{ TeV})}{\epsilon_{bb}(8\text{ TeV})}} \cdot \sqrt{\frac{\mathcal{L}(8\text{ TeV})}{\mathcal{L}(13\text{ TeV})}}, \end{aligned}$$

where σ_H and σ_{bb} are the Higgs and $b\bar{b}$ cross sections, ϵ_H and ϵ_{bb} are the Higgs and $b\bar{b}$ selection efficiencies and $\mathcal{L}$ is the integrated luminosity. One can assume approximately the same selection efficiencies for 13 TeV and 8 TeV, with the exception of the kinematical acceptance, which is calculated with PYTHIA 8. Therefore:

$$\frac{\epsilon_H(13\text{ TeV})}{\epsilon_H(8\text{ TeV})} \simeq 1.43,$$
$$\frac{\epsilon_{bb}(13\text{ TeV})}{\epsilon_{bb}(8\text{ TeV})} \simeq 1.06.$$

The $\frac{\sigma_{bb}(13\text{ TeV})}{\sigma_{bb}(8\text{ TeV})}$ ratio can be taken from [21]. $\frac{\sigma_H(8\text{ TeV})}{\sigma_H(13\text{ TeV})}$ can be taken from [25]: the values for the inclusive production and for the associated production are compatible within the theoretical uncertainty and they are approximately equal to 2.0. With these assumptions the scale factor is

$$f = 0.46 \cdot \sqrt{\frac{\mathcal{L}(8\text{ TeV})}{\mathcal{L}(13\text{ TeV})}}.$$

The integrated luminosity at 8 TeV is 2 fb^{-1} and in the Run II LHCb plans to collect 5 fb^{-1}. Therefore the scale factor results equal to 0.29 and the inclusive $H^0 \to b\bar{b}$ limit would be reduced from 30 times the SM cross section to 8.7 times the SM cross section at the end of Run II. The $V + H^0(\to b\bar{b})$ limit would be reduced from 50 times the SM cross section to 15 times the SM cross section at the end of Run II. The integrated luminosity that LHCb should collect at 13 TeV to reduce the inclusive limit to 1 time the SM cross section, when the $H^0 \to b\bar{b}$ observation would be probably possible, is 381 fb^{-1}. Starting from 2018 an upgraded version of the LHCb detector will be available and the experiment plans to collect more integrated luminosity at 14 TeV. Neverthless efforts should be made to improve the $H^0 \to b\bar{b}$ sensitivity. Several improvements are possible:

- jet energy resolution: the jet reconstruction algorithm can be improved for a better jet energy resolution. This would lead to a better invariant mass peak resolution for $b\bar{b}$ resonances, increasing the sensitivity to their observation;
- jet b tagging: the b tagging efficiency can be improved by studying new tagging algorithms;
- b-jet trigger: LHCb had no trigger lines dedicated to b jets in Run I. This has been one of the main limitation to the LHCb capability in b-jets physics. Implementing

such trigger lines in the LHCb trigger system would surely lead to an improvement in the selection efficiency. Some lines optimized for this purpose are already running in the Run II data taking;

- analysis techniques: new techniques for the reconstruction of $b\bar{b}$ resonances can be tested at LHCb. An example is the regression technique used by CMS in to improve the Higgs resolution [29].

Conclusions

In this thesis techniques to identify and reconstruct $b\bar{b}$ resonances with the LHCb detector have been developed. They have been used in the search for the SM Higgs decaying to a pair of b quarks in the forward region of pp collisions.

First the jet reconstruction have been validated studying a b-dijet data sample for the measurement of the $Z \rightarrow b\bar{b}$ cross section and of the Jet Energy Scale. The measured $Z \rightarrow b\bar{b}$ cross section in the forward region (two b quarks with $2 < \eta < 5$) of pp collisions at 8 TeV is

$$\sigma(pp \rightarrow Z)\mathcal{B}(Z \rightarrow b\bar{b}) = 712 \pm 108(\text{stat.}) \pm 146(\text{syst.}) \pm 7(\text{lum.}),$$

which is compatible with the SM theoretical prediction within its uncertainty. The measured Jet Energy Scale factor, which has to be applied to the jet four-momentum in simulation to match the real distribution, is

$$k_{JES} = 1.019 \pm 0.018(\text{stat.}) \pm 0.014(\text{syst.}).$$

The heavy flavour dijet+lepton data sample has been studied to set an upper limit on the SM Higgs production in association with a vector boson. First the cross sections of $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$, which are the main background processes in the $V + H^0$ search, have been measured and they have been found to be compatible with the SM expectations within their uncertainties. The experimental upper limit on the $V + H^0(\rightarrow b\bar{b})$ cross section in the forward region (two b quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$) is

$$\sigma(pp \rightarrow W/Z + H^0)\mathcal{B}(H^0 \rightarrow b\bar{b}) < 1.6 \text{ pb}, \text{ at 95\% CL and at } 8\,\text{TeV},$$

which corresponds to 50 times the SM expectation. The experimental upper limit on the $V + H^0(\rightarrow c\bar{c})$ cross section in the forward region (two c quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$) is

$$\sigma(pp \rightarrow W/Z + H^0)\mathcal{B}(H^0 \rightarrow c\bar{c}) < 9.4 \text{ pb}, \text{ at 95\% CL and at } 8\,\text{TeV},$$

which corresponds to 6400 times the SM expectation.

The b-dijet data sample have been studied to set an upper limit on the inclusive SM $H^0 \rightarrow b\bar{b}$ production in the forward region. This has found to be

$$\sigma(pp \rightarrow X + H^0)\mathcal{B}(H^0 \rightarrow b\bar{b}) < 19.4 \text{ pb}, \text{ at 95\% CL and at } 8\,\text{TeV}.$$

which corresponds to 30 times the SM expectation.

In conclusions more effort should be done to increase the LHCb sensitivity to the $H^0 \rightarrow b\bar{b}$ observation studying the data collected during the Run II. Improvements from jet reconstruction algorithms, b-jet tagging and b-jet trigger are expected.

Bibliography

[1] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[2] CMS, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[3] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[4] ATLAS, CMS collaborations, G. Aad *et al.*, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, `arXiv:1606.02266`.

[5] ATLAS collaboration, G. Aad *et al.*, *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. **D91** (2015) 012006, `arXiv:1408.5191`.

[6] CMS collaboration, S. Chatrchyan *et al.*, *Measurement of the properties of a Higgs boson in the four-lepton final state*, Phys. Rev. **D89** (2014) 092007, `arXiv:1312.5353`.

[7] LHCb, A. A. Alves, Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[8] U. Baur, M. Spira, and P. M. Zerwas, *Excited Quark and Lepton Production at Hadron Colliders*, Phys. Rev. **D42** (1990) 815.

[9] P. H. Frampton and S. L. Glashow, *Chiral Color: An Alternative to the Standard Model*, Phys. Lett. **B190** (1987) 157.

[10] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[11] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, Nuovo Cim. **11** (1959) 568.

[12] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[13] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction*, 1989.

[14] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413.

[15] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[16] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[17] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585.

[19] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145** (1966) 1156.

[20] T. W. B. Kibble, *Symmetry breaking in nonAbelian gauge theories*, Phys. Rev. **155** (1967) 1554.

[21] Particle Data Group, J. Beringer *et al.*, *Review of Particle Physics (RPP)*, Phys. Rev. **D86** (2012) 010001.

[22] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531, [,648(1963)].

[23] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[24] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, `arXiv:hep-ph/0409313`.

[25] LHC Higgs Cross Section Working Group, J. R. Andersen *et al.*, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, `arXiv:1307.1347`.

[26] ATLAS collaboration, G. Aad *et al.*, *Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector*, Phys. Rev. **D90** (2014) 112015, `arXiv:1408.7084`.

[27] CMS collaboration, V. Khachatryan *et al.*, *Observation of the diphoton decay of the Higgs boson and measurement of its properties*, Eur. Phys. J. **C74** (2014), no. 10 3076, `arXiv:1407.0558`.

[28] ATLAS collaboration, G. Aad *et al.*, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) 069, `arXiv:1409.6212`.

[29] CMS collaboration, S. Chatrchyan *et al.*, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys. Rev. **D89** (2014) 012003, `arXiv:1310.3687`.

[30] ATLAS, CMS collaborations, G. Aad *et al.*, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, `arXiv:1606.02266`.

[31] CDF collaboration, T. Aaltonen *et al.*, *Search for the standard model Higgs boson decaying to a bb pair in events with one charged lepton and large missing transverse energy using the full CDF data set*, Phys. Rev. Lett. **109** (2012) 111804, `arXiv:1207.1703`.

[32] D0 collaboration, V. M. Abazov *et al.*, *Search for the standard model Higgs boson in $\ell\nu$ + jets final states in 9.7 fb^{-1} of $p\bar{p}$ collisions with the D0 detector*, Phys. Rev. **D88** (2013), no. 5 052008, `arXiv:1301.6122`.

[33] D0 collaboration, V. M. Abazov *et al.*, *Search for the standard model Higgs boson in associated WH production in 9.7 fb^{-1} of $p\bar{p}$ collisions with the D0 detector*, Phys. Rev. Lett. **109** (2012) 121804, `arXiv:1208.0653`.

[34] D0 collaboration, V. M. Abazov *et al.*, *Search for $ZH \to \ell^+\ell^-b\bar{b}$ production in 9.7 fb^{-1} of $p\bar{p}$ collisions with the D0 detector*, Phys. Rev. **D88** (2013), no. 5 052010, `arXiv:1303.3276`.

[35] D0 collaboration, V. M. Abazov *et al.*, *Search for the standard model Higgs boson in $ZH \to \ell^+\ell^-b\bar{b}$ production with the D0 detector in 9.7 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. Lett. **109** (2012) 121803, `arXiv:1207.5819`.

[36] CDF, D0, T. Aaltonen *et al.*, *Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron*, Phys. Rev. Lett. **109** (2012) 071804, `arXiv:1207.6436`.

[37] ATLAS, A. B. Kowalewska, *Higgs τ-lepton Yukawa coupling measurement and the τ embedding method for background estimation*, Acta Phys. Polon. **B47** (2016) 1691.

[38] CMS collaboration, S. Chatrchyan *et al.*, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons*, JHEP **05** (2014) 104, `arXiv:1401.5041`.

[39] LHCb collaboration, R. Aaij *et al.*, *Limits on neutral Higgs boson production in the forward region in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **05** (2013) 132, `arXiv:1304.2591`.

[40] CDF, D0, T. Aaltonen *et al.*, *Higgs Boson Studies at the Tevatron*, Phys. Rev. **D88** (2013), no. 5 052014, `arXiv:1303.6346`.

[41] ATLAS, CMS collaborations, G. Aad *et al.*, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s} = 7$ *and 8 TeV*, JHEP **08** (2016) 045, `arXiv:1606.02266`.

[42] ATLAS, CMS collaborations, G. Aad *et al.*, *Combined measurement of the Higgs boson mass in pp collisions at* $\sqrt{s} = 7$ *and 8 TeV with the ATLAS and CMS experiments*, Phys. Rev. Lett. **114** (2015) 191803, `arXiv:1503.07589`.

[43] C. Delaunay, T. Golling, G. Perez, and Y. Soreq, *Enhanced Higgs boson coupling to charm pairs*, Phys. Rev. **D89** (2014) 033014, `arXiv:1310.7029`.

[44] ATLAS collaboration, G. Aad *et al.*, *Search for Higgs and Z boson decays to* $J/\psi\gamma$ *and* $\Upsilon(nS)\gamma$ *with the ATLAS detector*, Phys. Rev. Lett. **114** (2015), no. 12 121801, `arXiv:1501.03276`.

[45] CMS collaboration, V. Khachatryan *et al.*, *Search for a Higgs boson decaying into* $\gamma^*\gamma \to \ell\ell\gamma$ *with low dilepton mass in pp collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Lett. **B753** (2016) 341, `arXiv:1507.03031`.

[46] A. Falkowski, F. Riva, and A. Urbano, *Higgs at last*, JHEP **11** (2013) 111, `arXiv:1303.1812`.

[47] G. Perez, Y. Soreq, E. Stamou, and K. Tobioka, *Constraining the charm Yukawa and Higgs-quark coupling universality*, Phys. Rev. **D92** (2015) 033016, `arXiv:1503.00290`.

[48] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[49] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[50] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015), no. 07 1530022, `arXiv:1412.6352`.

[51] M. Kucharczyk *et al.*, *Primary Vertex Reconstruction at LHCb*, .

[52] F. Sauli, *GEM: A new concept for electron amplification in gas detectors*, Nucl. Instrum. Meth. **A386** (1997) 531.

[53] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[54] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[55] LHCb, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[56] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`.

[57] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, `arXiv:0709.2092`.

[58] M. L. Mangano *et al.*, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, `arXiv:hep-ph/0206293`.

[59] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[60] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[61] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[62] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[63] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, `arXiv:physics/0703039`.

[64] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063, `arXiv:0802.1189`.

[65] LHCb, R. Aaij *et al.*, *Study of forward Z + jet production in pp collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **01** (2014) 033, `arXiv:1310.8197`.

[66] LHCb, R. Aaij *et al.*, *Identification of beauty and charm quark jets at LHCb*, JINST **10** (2015), no. 06 P06013, `arXiv:1504.07670`.

[67] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[68] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, `arXiv:physics/0408124`.

[69] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119.

[70] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, `arXiv:1210.6861`.

[71] J. Donini *et al.*, *Energy Calibration of* b^{-} *Quark Jets with* $Z \rightarrow b\bar{b}$ *Decays at the Tevatron Collider*, Nucl. Instrum. Meth. **A596** (2008) 354, `arXiv:0801.3906`.

[72] ATLAS, G. Aad *et al.*, *Measurement of the cross section of high transverse momentum $Z \to b\bar{b}$ production in proton–proton collisions at $\sqrt{s} = 8TeV$ with the ATLAS Detector*, Phys. Lett. **B738** (2014) 25, `arXiv:1404.7042`.

[73] LHCb, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014), no. 12 P12005, `arXiv:1410.0149`.

[74] CMS, S. Chatrchyan *et al.*, *Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Rev. Lett. **112** (2014) 191802, `arXiv:1402.0923`.

[75] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, `arXiv:1112.5675`.

[76] A. Banfi, G. P. Salam, and G. Zanderighi, *Phenomenology of event shapes at hadron colliders*, JHEP **06** (2010) 038, `arXiv:1001.4082`.

[77] LHCb, R. Aaij *et al.*, *Measurement of forward W and Z boson production in association with jets in proton-proton collisions at $\sqrt{s} = 8$ TeV*, JHEP **05** (2016) 131, `arXiv:1605.00951`.

[78] A. Rogozhnikov *et al.*, *New approaches for boosting to uniformity*, JINST **10** (2015), no. 03 T03002, `arXiv:1410.4140`.

[79] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, `arXiv:physics/0306116`, [,186(2003)].

[80] E. L. Berger, C. B. Jackson, S. Quackenbush, and G. Shaughnessy, *Calculation of $W b\bar{b}$ Production via double parton scattering at the LHC*, Phys. Rev. **D84** (2011) 074021, `arXiv:1107.3150`.

[81] R. Frederix *et al.*, *W and Z/γ^* boson production in association with a bottom-antibottom pair*, JHEP **09** (2011) 061, `arXiv:1106.6019`.

[82] C. Oleari and L. Reina, *$W^\pm b\bar{b}$ production in POWHEG*, JHEP **08** (2011) 061, `arXiv:1105.4488`, [Erratum: JHEP **11** (2011) 040].

[83] ATLAS collaboration, G. Aad *et al.*, *Measurement of the cross-section for W boson production in association with b-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **06** (2013) 084, `arXiv:1302.2929`.

[84] CMS collaboration, S. Chatrchyan *et al.*, *Measurement of the production cross section for a W boson and two b jets in pp collisions at $\sqrt{s}$=7 TeV*, Phys. Lett. **B735** (2014) 204, `arXiv:1312.6608`.

[85] CMS collaboration, V. Khachatryan *et al.*, *Measurement of the production cross section of the W boson in association with two b jets in pp collisions at $\sqrt{s} = 8$ TeV*, `arXiv:1608.07561`.

[86] R. Gauld, *Feasibility of top quark measurements at LHCb and constraints on the large-x gluon PDF*, JHEP **02** (2014) 126, `arXiv:1311.1810`.

[87] LHCb, R. Aaij *et al.*, *Measurement of forward W and Z boson production in pp collisions at $\sqrt{s} = 8$ TeV*, JHEP **01** (2016) 155, `arXiv:1511.08039`.

[88] LHCb, R. Aaij *et al.*, *Measurement of forward* $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ *production at* $\sqrt{s} = 8$ *TeV*, JHEP **05** (2015) 109, `arXiv:1503.00963`.

[89] J. M. Campbell and R. K. Ellis, *Radiative corrections to Z b anti-b production*, Phys. Rev. **D62** (2000) 114012, `arXiv:hep-ph/0006304`.

[90] H.-L. Lai *et al.*, *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, `arXiv:1007.2241`.

[91] LHCb, R. Aaij *et al.*, *Measurement of forward* $W \rightarrow e\nu$ *production in pp collisions at* $\sqrt{s} = 8$ *TeV*, JHEP **10** (2016) 030, `arXiv:1608.01484`.

[92] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.

[93] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. **A434** (1999) 435, `arXiv:hep-ex/9902006`.

[94] LHCb collaboration, R. Aaij *et al.*, *Search for the rare decays* $B_s^0 \rightarrow \mu^+\mu^-$ *and* $B^0 \rightarrow \mu^+\mu^-$, Phys. Lett. **B699** (2011) 330, `arXiv:1103.2465`.

Thesis summary

Title: Search for a Higgs boson decaying to a pair of b quarks in the forward region of pp collisions with the LHCb detector.
Phd student: Lorenzo Sestini
Supervisor: Prof. Donatella Lucchesi

LHCb is an experiment initially designed for heavy flavour physics, located at the Large Hadron Collider (LHC). It is a forward spectrometer (pseudorapidity coverage $2 < \eta < 5$) which primary goal has been to look for evidence of new physics in CP violation and rare decays of beauty and charm hadrons in proton-proton collisions. In the last years, the development of jets reconstruction and heavy flavour tagging algorithms optimized for LHCb has qualified the experiment as a General Forward Detector. Thanks to its unique features LHCb is able to perform electroweak and jets measurements in a complementary phase space with respect to the General Purpose Detectors (GPD) at LHC, ATLAS and CMS.

In this thesis techniques to identify and reconstruct $b\bar{b}$ resonances with the LHCb detector are developed. In particular the goal of this work is to set experimental limit on the Standard Model (SM) $H^0 \to b\bar{b}$ production in the forward region.

First an overview of the Electroweak theory and of the Brout-Englert-Higgs mechanism is given. The Higgs production at the LHC is explained and the experimental results on the Higgs are discussed.

After a brief introduction of the LHC accelerator facility, the sub-systems forming the LHCb detector are described and their performance are discussed. Moreover the jet reconstruction and heavy flavour identification algorithms used at LHCb the Run I data taking are presented with the measurements of their performance.

The data collected by LHCb during the Run I data taking (pp collisions with a centre-of-mass energy of 8 TeV corresponding to an integrated luminosity of 2 fb^{-1}) are analyzed to identify the $Z \to b\bar{b}$ decay and to measure its cross section. The measured cros section is

$$\sigma(pp \to Z)\mathcal{B}(Z \to b\bar{b}) = 712 \pm 108(\text{stat.}) \pm 146(\text{syst.}) \pm 7(\text{lum.}),$$

where the first uncertainty is statistical, the second systematic and the third is related to the luminosity measurement. The measured cross section is compatible with the SM theoretical prediction within its uncertainty. The Jet Energy Scale (JES) is also measured by studying the $Z \to b\bar{b}$ sample, which is fundamental to validate the LHCb b-jet reconstruction techniques. The JES is a correction factor that should be applied to the jets energy in the LHCb simulation to match the real distribution. It has been found to be

$$k_{JES} = 1.019 \pm 0.018(\text{stat.}) \pm 0.014(\text{syst.}).$$

The heavy flavour dijet+lepton data sample collected during the Run I has been studied to set an upper limit on the SM Higgs production in association with a vector boson (W or Z). First the cross sections of $W + b\bar{b}$, $W + c\bar{c}$ and $t\bar{t}$, which are the main background processes in the $V + H^0$ search, have been measured and they have been found to be compatible with the SM expectations within their uncertainties. The experimental upper limit on the $V + H^0(\rightarrow b\bar{b})$ cross section in the forward region (two b quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$) at 95% Confidence Level (CL) is

$$\sigma(pp \rightarrow W/Z + H^0)\mathcal{B}(H^0 \rightarrow b\bar{b}) < 1.6 \text{ pb}, \text{ at } 95\% \text{ CL and at } 8\,\text{TeV},$$

which corresponds to 50 times the SM expectation. The experimental upper limit on the $V + H^0(\rightarrow c\bar{c})$ cross section in the forward region (two c quarks from H^0 and one lepton from W/Z with $2 < \eta < 5$) is

$$\sigma(pp \rightarrow W/Z + H^0)\mathcal{B}(H^0 \rightarrow c\bar{c}) < 9.4 \text{ pb}, \text{ at } 95\% \text{ CL and at } 8\,\text{TeV},$$

which corresponds to 6400 times the SM expectation.

The b-dijet data sample colected during the Run I have been studied to set an upper limit on the inclusive SM $H^0 \rightarrow b\bar{b}$ production in the forward region. This has found to be

$$\sigma(pp \rightarrow X + H^0)\mathcal{B}(H^0 \rightarrow b\bar{b}) < 19.4 \text{ pb}, \text{ at } 95\% \text{ CL and at } 8\,\text{TeV}.$$

which corresponds to 30 times the SM expectation.

In conclusions more effort should be done to increase the LHCb sensitivity to the $H^0 \rightarrow b\bar{b}$ observation studying the data collected during the Run II. Improvements from jet reconstruction algorithms, b-jet tagging and b-jet trigger are expected.