UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Search for high-mass resonances decaying to heavy flavour jets

Thesis supervisor

Prof. Donatella Lucchesi

Thesis co-supervisor

Dr. Lorenzo Sestini

Candidate

Davide Zuliani

Academic Year 2018/19

Contents

Introduction

The main purpose of this thesis is to search for high-mass resonances decaying to a pair of heavy flavour (b or c) jets. Jets originated by b and c quarks are reconstructed in the forward region of the LHCb experiment at LHC collider at CERN. LHCb offers the possibility to study a different phase space respect to ATLAS and CMS experiments enhancing the LHC sensitivity to New Physics (NP) beyond Standard Model (SM). The main challenge of these measurements is the determination of the Quantum Chromodynamics (QCD) multi-jet background. The Monte Carlo (MC) simulation is not able to reproduce with enough precision the heavy flavour di-jets characteristics, due to the large theoretical uncertainties on QCD sub-processes, therefore data is used. LHCb has collected less data respect to ATLAS and CMS due to its data acquisition system, but the lower number of events is compensated by the powerful silicon vertex detector (Vertex Locator, VELO) that allows to reduce the QCD background and have high efficiency in the heavy flavor jets identification. In this thesis the inclusive processes $pp \to H \to b\bar{b}$ and $pp \to H \to c\bar{c}$ are studied and limits on the production cross sections are set. The decays $H \to b\bar{b}$ has been studied by ATLAS and CMS mainly in association with a vector boson $V = Z, W^{\pm}$. Both collaborations have presented results in agreement with the SM expectations [1]. Preliminary results on $H \to c\bar{c}$ have been presented by CMS [2]: the observed limit on the cross section for the associated production with Z and W bosons is 70 times the SM value at 95% confidence level. In this thesis model-independent limits on $b\bar{b}$ and $c\bar{c}$ dijet resonances cross sections with a mass different from that of the SM Higgs boson are also set. A similar study has already been done for $b\bar{b}$ dijet resonances by CMS [3]. The inclusive search for $c\bar{c}$ resonances is done for the first time.

The details of the analysis and the results obtained with LHCb data, are described in thesis which is structured as follow chapter by chapter.

1. Chapter 1: the theoretical framework, i.e. the Electroweak theory and the Brout-Englert-Higgs mechanism, is presented. Since its discovery in July 2012, the Higgs boson H has been studied to precisely measure its couplings with fermions and bosons, in order to test SM predictions and to search for possible NP. In this chapter an overview of the theoretical framework is given together with the state of the art of experimental results. Finally, an overview of some extensions of SM is given, connecting these NP theories with the search for di-jets high mass resonances at LHCb.

2. Chapter 2: the experimental instrument, LHCb, used to perform the aforementioned searches is presented. LHCb is a forward spectrometer initially designed to study heavy flavour hadron physics in a complementary phase space region with respect to ATLAS and CMS (which have a cylindrical structure) where the production cross section is higher than the central one. In this chapter the LHC accelerator facility is briefly described, then a description of the LHCb detector is given, mainly focusing on its sub-systems used in this measurements and their performances.

3. Chapter 3: in this chapter the main algorithms used for jet reconstruction and heavy flavour identification at LHCb during Run II are presented, focusing on their functioning and performances. The *Boosted Decision Tree* (BDT) algorithm is described in detail. It is the essential instrument to separate b and c initiated jets from light quarks one.

4. Chapter 4: this chapter is dedicated to the study of the backgrounds. A data-driven method to model the QCD heavy flavor di-jets background is presented. Starting from a MC simulation of

the QCD processes and a small portion of 2016 Run II data two sub-samples are created, the *Control Region* (CR), where no signal is expected but the characteristics are similar to the *Signal Region* (SR), where the signal is expected. In this way it is possible to build a function, called Transfer Function (TF), that allows to describe the background in the SR starting from the CR data. The $Z \to b\bar{b}$ and $Z \to c\bar{c}$ contributions to the SR are studied and evaluated. Finally, the computation of the systematic uncertainties on the background number of events is described.

5. Chapter 5: here, the results on the reconstruction of $H \to b\bar{b}$ and $H \to c\bar{c}$ are presented. Since there is not enough sensitivity to these processes an upper limit on the production cross sections are set by means of the CL_s *method*. The same procedure is then applied to the search for heavy flavor di-jets high-mass resonances (45, 60, 80, 100 and 175 GeV) in order to set limits on the production cross sections on new particles.

6. Chapter 6: in this chapter the upgrades on the LHCb detector relevant for these searches are described. On a parallel path, new analysis techniques are under development, using new computational techniques as *Deep Neural Networks* (DNN), in order to build new tools that will be able to process the increasing amount of data collected by the experiment. A new particle-based DNN architecture is described in detail, because it is a possible algorithm for heavy flavour tagging. To conclude, the expected LHCb sensitivity to heavy flavor di-jets high-mass resonances for Run III, Run IV and High-Luminosity LHC (HL-LHC) phase are presented.

Chapter 1

Theoretical introduction: Higgs boson and New Physics

1.1 Electroweak theory

In presenting the theoretical framework, the first element to consider is the Electroweak (EW) theory [4–6], which unifies the electromagnetic and the weak interactions. The EW theory is invariant under the following group transformation:

$$G_{EW} = U_Y(1) \times SU_L(2) \tag{1.1}$$

where $U_Y(1)$ represents the local phase invariance of the electromagnetic interaction while $SU_L(2)$ is the isospin invariance of the weak interaction. It is known that the weak current couples only to left-handed fermions, thus reflecting the non-conservation of parity in weak interactions. Using Dirac notation, we can define respectively left-handed and right-handed fermions ψ_L and ψ_R as

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

where ψ is the spinor, P_L and P_R are chirality operators and γ^5 is defined as the product of Dirac matrices

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 \tag{1.3}$$

where γ^i with $i = 0, \ldots, 3$ are Dirac matrices. In this way in the EW lagrangian left-handed fermions transform as doublets while right-handed as singlets; quarks are written as

$$q_L^i = \begin{bmatrix} u_L \\ d_L \end{bmatrix}, \begin{bmatrix} s_L \\ c_L \end{bmatrix}, \begin{bmatrix} t_L \\ b_L \end{bmatrix} \tag{1.4}$$

$$\begin{aligned} u_R^i &= u_R,\ c_R,\ t_R \\ d_R^i &= d_R,\ s_R,\ b_R \end{aligned} \tag{1.5}$$

and similarly for the leptons

$$l_L^i = \begin{bmatrix} e_L \\ \nu_{e,L} \end{bmatrix}, \begin{bmatrix} \mu_L \\ \nu_{\mu,L} \end{bmatrix}, \begin{bmatrix} \tau_L \\ \nu_{\tau,L} \end{bmatrix} \tag{1.6}$$

$$l_R^i = e_R,\ \mu_R,\ \tau_R \tag{1.7}$$

where the index i is related to the three families of quarks or leptons. Since in the SM the neutrino is considered massless it is only left-handed.
The complete EW Lagrangian of the SM can be written in the following way

$$\mathcal{L} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_\phi + \mathcal{L}_Y \tag{1.8}$$

where $\mathcal{L}_{gauge}$ ($\mathcal{L}_f$) describes the gauge (fermion) sector, $\mathcal{L}_\phi$ introduces a field ϕ related to symmetry breaking and $\mathcal{L}_Y$ describes the interaction between the ϕ field and fermions. The EW gauge Lagrangian is expressed by

$$\mathcal{L}_{gauge} = -\frac{1}{4} W^i_{\mu\nu} W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.9}$$

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

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial\nu B_\mu \\ W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon_{ijk} W^j_\mu W^k_\nu \end{aligned} \tag{1.10}$$

where B_μ and W^i_μ (with $i = 1, 2, 3$) are fields, g is the weak coupling constant and ϵ_{ijk} is the Levi-Civita tensor; it is interesting to spot the non-abelian structure of $W^i_{\mu\nu}$ (the third term) typical of the $SU(2)$ symmetry group.
The fermion part of the Lagrangian is written as

$$\mathcal{L}_f = \bar{q}^i_L i \not{D} q_{i,L} + \bar{l}^i_L i \not{D} l_{i,L} + \bar{u}_R i \not{D} u_R + \bar{d}_R i \not{D} d_R + \bar{e}_R i \not{D} e_R \tag{1.11}$$

where $\not{D} = D_\mu \gamma^\mu$ and D_μ is the covariant derivative defined as:

$$D_\mu = \partial_\mu + ig\frac{\tau_i}{2} W^i_\mu + i\frac{g'}{2} B_\mu \tag{1.12}$$

where g' is the coupling constant related to $U_Y(1)$ transformations and τ_i are the Pauli matrices. In principle the EW theory would require four force carriers which are massless, two electrically charged and two neutrally. Since experimentally weak interactions are proved to be short ranged, the force carriers must have mass. This is why there is a mechanism that breaks the EW symmetry in order to create massive vector bosons: this is the so called Electroweak Symmetry Breaking (EWSB), which creates two massive bosons with electric charge (W^+ and W^-), one massive boson with neutral charge (Z) and the photon, which stays massless. The three massive force carriers are exchanged in weak interactions.
To conclude this section, the last two terms of $\mathcal{L}_{EW}$ are related to EWSB by describing the field responsible of the breaking ($\mathcal{L}_\phi$) and how it interacts with fermions (the *Yukawa Lagrangian* $\mathcal{L}_Y$); these will be described in Sec.1.2 and 1.3.

1.2 Brout-Englert-Higgs mechanism

The $\mathcal{L}_\phi$ Lagrangian responsible for the EWSB can be written as:

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger D^\mu \phi + V(\phi) \tag{1.13}$$

where ϕ is a scalar field doublet defined as

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{bmatrix} \tag{1.14}$$

where ϕ_i (with $i = 1, 2, 3, 4$) are real scalar fields. $V(\phi)$ is the field potential defined as

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \tag{1.15}$$

and μ and λ are real constants. For $\lambda > 0$ the "Mexican hat" potential is obtained: the energy of the system (thus the potential) is minimized by a continuum infinite set of possible field configurations, thus by choosing arbitrarily one of them (*spontaneous symmetry breaking*) we can assume

$$\begin{aligned} \langle 0|\phi_i|0\rangle &= 0 \text{ for } i = 1, 2, 4 \\ \langle 0|\phi_3|0\rangle &= \nu \end{aligned} \tag{1.16}$$

where ν is a constant called *vacuum expectation value*. It is now possible to expand the ϕ_3 component around ν with linear perturbations

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ \nu + H \end{bmatrix} \tag{1.17}$$

where H is a real scalar field called *Higgs fields*. Before inserting this ϕ field inside $\mathcal{L}_\phi$ it is convenient to rotate the four vector bosons by using the *Weinberg angle* θ_W

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp W_\mu^2 \right) \\ Z_\mu &= \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \\ A_\mu &= \sin(\theta_W) W_\mu^3 + \sin(\theta_W) B_\mu \end{aligned} \tag{1.18}$$

and the Weinberg angles is defined such that

$$\begin{aligned} \sin(\theta_W) &= \frac{g}{\sqrt{g^2 + g'^2}} \\ \cos(\theta_W) &= \frac{g'}{\sqrt{g^2 + g'^2}} \end{aligned} \tag{1.19}$$

Inserting Eq.1.18 and Eq.1.19 in Eq.1.13 and considering up to the fourth order in H the Higgs Lagrangian is obtained

$$\mathcal{L}_h = \frac{1}{2} \partial_\mu H \partial^\mu H + (\nu + H)^2 \left[\frac{g^2}{4} W_\mu^\dagger W^\mu + \frac{1}{8} \left(g^2 + g'^2 \right) Z_\mu^\dagger Z^\mu \right] - \lambda \nu^2 H^2 - \lambda \nu H^3 - \frac{1}{4} H^4 \tag{1.20}$$

After the EWSB the W and Z bosons acquire mass given by the coefficients of the $W_\mu^\dagger W^\mu$ and $Z_\mu^\dagger Z^\mu$ terms, while the photon A stays massless:

$$\begin{aligned} m_W &= \frac{1}{2} \nu g \\ m_Z &= \frac{1}{2} \nu \sqrt{g^2 + g'^2} \end{aligned} \tag{1.21}$$

The Higgs fields is quantized and it generates the Higgs boson, whose mass is given by the coefficient of the H^2 term

$$m_H = \sqrt{2 \lambda \nu^2} \tag{1.22}$$

The mechanism described in this section that generates the EWSB and gives mass to vector bosons is known as *Brout-Englert-Higgs mechanism* [7–12]. Moreover by knowing the value of the Fermi constant G_F (measured in muon decays [13]) it is possible to get the vacuum expectation value ν by means of the first equation in Eq.1.21

$$\nu = \frac{2 m_W}{g} \simeq \left(\sqrt{2} G_F \right)^{-\frac{1}{2}} \simeq 246 \text{ GeV} \tag{1.23}$$

1.3 Yukawa Couplings

The Yukawa Lagrangian $\mathcal{L}_Y$ is devoted to describe the interaction between the field ϕ and fermions. In order to understand its structure, let's first consider the mass term of fermions according to Dirac equation

$$-m \bar{\psi} \psi = -m \left(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R \right) \tag{1.24}$$

Since the $SU_L(2)$ group acts only on the left-handed fermions, the mass term of the Dirac field is not $SU_L(2)$ invariant, thus it is not possible to insert Eq.1.24 inside $\mathcal{L}_{EW}$, in contrast with the fact that fermions are massive. In order to solve this inconsistency it is necessary to introduce a *Yukawa coupling* (and thus $\mathcal{L}_Y$) by assuming interactions between fermions and the scalar field ϕ; this Lagrangian has to be invariant under G_{EW}. For leptons the Yukawa Lagrangian has the following form

$$\mathcal{L}^i_{e,Y} = -g^i \left(\bar{l}^i_L \phi e^i_R + \bar{e}^i_R \phi^\dagger l^i_L \right) \tag{1.25}$$

where g^i is a constant related to the flavour i, and ϕ is the scalar doublet in Eq.1.14. Again neutrinos are not considered since they are massless in the SM. After the EWSB if the Higgs fields acquires a vacuum expectation value $\nu \neq 0$ then Eq.1.25 becomes

$$\mathcal{L}^i_{e,Y} = -\frac{g^i \nu}{\sqrt{2}} \left(\bar{e}^i_L e^i_R + \bar{e}^i_R e^i_L \right) - \frac{g^i}{\sqrt{2}} \left(\bar{e}^i_L e^i_R + \bar{e}^i_R e^i_L \right) h = \frac{g^i \nu}{\sqrt{2}} \bar{e}^i e^i - \frac{g^i}{\sqrt{2}} \bar{e}^i e^i H \tag{1.26}$$

It is clear now that the coefficient of the term $\bar{e}^i e^i$ is the mass of the lepton, which is then given by

$$m^i = \frac{g^i \nu}{\sqrt{2}} \tag{1.27}$$

while the term $\bar{e}^i e^i h$ describes the interaction between leptons and the Higgs field. The same reasoning can be done for quarks, whose Yukawa Lagrangian before EWSB is

$$\mathcal{L}_{q,Y} = -Y^u_{ij} \bar{q}^i_L \phi u^j_R - Y^d_{ij} \bar{q}^i_L \phi_C d^j_R + \ h.c. \tag{1.28}$$

where i and j indicate quark flavour, $\phi_C = i\tau_2 \phi$ is the conjugate Higgs field and Y^u and Y^d are 3×3 matrices called *Yukawa matrices* which have to be determined by experiments. In order to obtain quark masses after EWSB it is necessary to diagonalize Y^u and Y^d by means of a unitary transformation V

$$\begin{aligned} M^u &= \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{bmatrix} = \frac{\nu}{\sqrt{2}} V^u_L Y^u V^{u\dagger}_R \\ M^d &= \begin{bmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{bmatrix} = \frac{\nu}{\sqrt{2}} V^d_L Y^d V^{d\dagger}_R \end{aligned} \tag{1.29}$$

It is possible to understand that the interaction eigenstates are different from the mass eigenstates, but they are connected by the so called *Cabibbo-Kobayashi-Maskawa* (CKM) matrix [14, 15], given by

$$V_{CKM} = V^d_L V^{u\dagger}_L \tag{1.30}$$

To conclude this section it is interesting to notice that a very important consequence of the Yukawa Lagrangian is that the coupling between fermions and the Higgs field is proportional to the mass of the fermion, by simply rewriting Eq.1.27

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

1.4 Higgs boson at LHC

Once it is clear the theoretical framework behind EW theory and the Higgs mechanism, it is interesting to study the physics of the Higgs boson for particle physics experiments such as the Large Hadron Collider (LHC) at CERN.

1.4.1 Production mechanism

The main process at LHC is proton-proton collision. A general process where a Higgs boson is produced after a proton-proton collision has the following cross section [16]

$$\sigma_{pp\to H+X} = \sum_{a,b} \int f_{a/A}\left(x_a, Q^2\right) f_{b/B}\left(x_b, Q^2\right) \sigma_{ab\to H+X}\left(Q^2\right)\, dx_a dx_b \tag{1.32}$$

where $a(b)$ indicates a parton inside the proton $A(B)$, $f_{a/A}$ ($f_{b/B}$) is the so called *Parton Distribution Function* which describes the probability for the parton $a(b)$ to have a fraction of the proton momentum $x_a(x_b)$ and a transferred momentum squared Q^2. $\sigma_{ab\to H+X}$ is the partonic cross section that can be easily obtained from the matrix element of the considered process. Considering proton-proton collisions there are several leading process $ab \to H + X$:

- $gg \to H$: a Higgs boson is produced via gluon-gluon fusion, after a quark loop whose main contribution comes from the top quark (since it is the most massive) corresponding to diagram (a) in Fig.1.1;
- $q\bar{q}' \to W/ZH$: this process is the so called Higgs-strahlung, which after fusing a quark and an anti-quark it produces an excited vector boson (Z^*/W^*) which afterward de-excites producing a vector boson and a Higgs boson. This process is pictured by diagram (b) in Fig.1.1;
- $qq' \to qq'H$: a Higgs boson is produced via vector-boson fusion (VBF), where two vector bosons (W^+W^- or ZZ) fuse together to produce a Higgs boson, corresponding to diagram (c) in Fig.1.1;
- $gg \to t\bar{t}H$: in this last process starting from two gluons a $t\bar{t}$ couple is produced together with a Higgs boson, a process corresponding to diagram (d) in Fig.1.1.

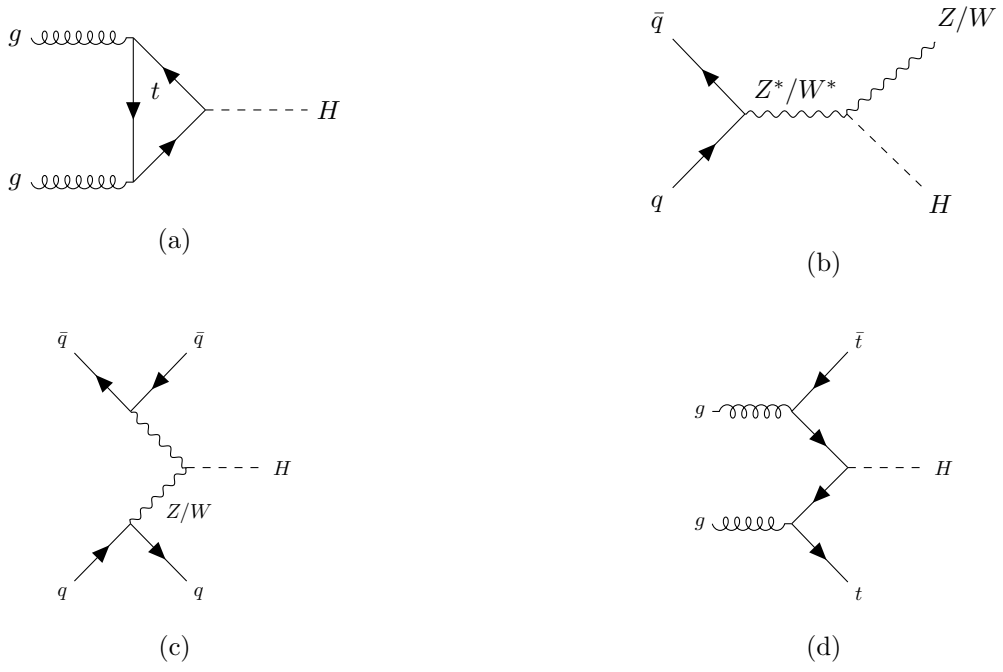

Figure 1.1: Feynman diagrams for different Higgs production processes in proton-proton collisions: gluon-gluon fusion (a), Higgs-strahlung (b), VBF (c) and $pp \to t\bar{t}H$ (d).

It is also interesting to recall the production cross section for a Higgs with a mass of 125.09 GeV/c^2 in proton-proton collisions at a CoM energy $\sqrt{s}$ of 13 TeV, reported in Table 1.1. For gluon-gluon fusion QCD corrections are computed at next-to-next-to-next-to-leading order[1] (N^{3}LO) while EW theory

[1]When expanding a theory through a perturbative expansion of the coupling constant it is possible to consider different levels of expansions: at *leading order* no loop are present; at *next-to-leading order* 1 loop is present, and so on. In this way when introducing loops divergences arise and the theory has to be renormalized at every level of expansion: if this happens the theory is called *renormalizable*.

is computed at next-to-leading order (NLO), also considering NLO contribution for mixed QCD-EW and finite quark masses; for VBF and Higgs-strahlung QCD is computed at NNLO while EW at NLO; finally for $pp \to t\bar{t}H$ cross section is computed at NLO both for QCD and QCD+EW [17].

Production process	cross section (pb)
gluon-gluon fusion	48.61
VBF	3.766
Higgs-strahlung (ZH)	0.880
Higgs-strahlung (WH)	1.358
$pp \to t\bar{t}H$	0.5122

Table 1.1: Cross sections for different Higgs production processes at $\sqrt{s} = 13$ TeV [17].

In Fig.1.2 production cross sections are plotted with respect to the CoM energy $\sqrt{s}$, for a Higgs mass $m_H = 125$ GeV/c^2 [18]. It is evident that the gluon-gluon fusion production mechanism is the most relevant (blue line). The band associated to each curve represents the error.

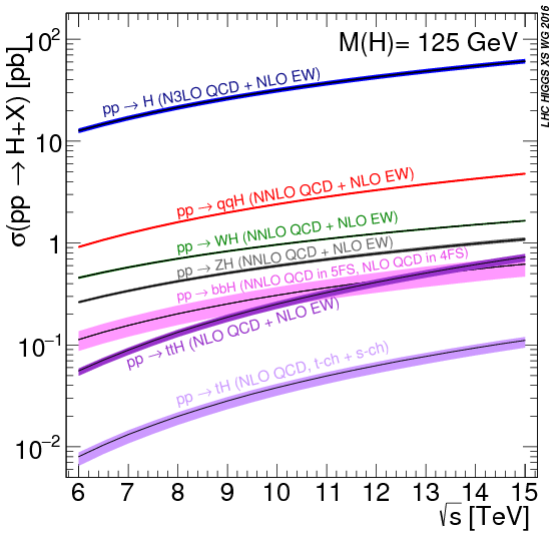

Figure 1.2: Higgs production cross sections for different production mechanisms as functions of the CoM energy $\sqrt{s}$ [18].

1.4.2 Higgs decay mode

The Higgs boson decays are summarised in Table 1.2 with the relative branching ratios [17].

Decay channel	$\mathcal{BR}$
$H \to b\bar{b}$	5.809×10^{-1}
$H \to WW$	2.152×10^{-1}
$H \to \tau\bar{\tau}$	6.256×10^{-2}
$H \to c\bar{c}$	2.884×10^{-2}
$H \to ZZ$	2.641×10^{-2}
$H \to \gamma\gamma$	2.270×10^{-3}

Table 1.2: Branching ratios for different Higgs decay channels [17].

It is interesting to notice that the channel $H \to b\bar{b}$ is dominant one and it involves the analysis of jet physics. In Fig.1.3 $\mathcal{BR}$ for different Higgs decays are plotted as a function of the Higgs boson mass around 125 GeV/c^2 [18]: since the Higgs mass m_H is known with a certain error it is interesting to

understand how the $\mathcal{BR}$s change with respect to m_H. The width of each curve represents the error. In the following subsection the most recent results on Higgs physics are recalled.

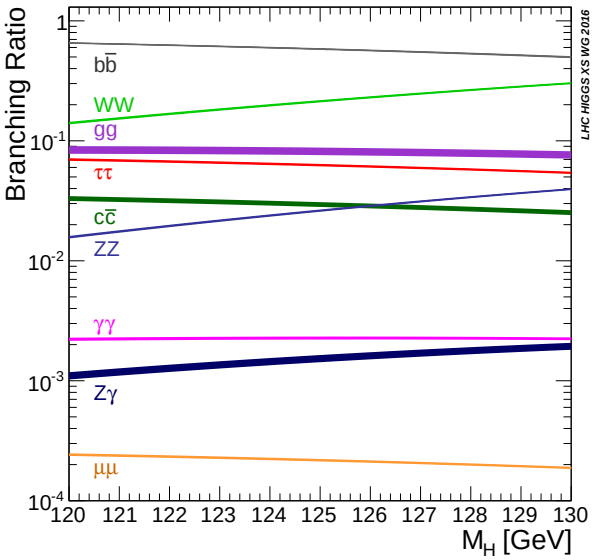

Figure 1.3: Higgs decay branching ratios for different processes as functions of the Higgs boson mass [18].

1.4.3 Current experimental results

ATLAS and CMS after the Higgs discovery have measured its properties. LHCb is entering the game exploiting a complementary region. The latest results from Run II on the various Higgs decay modes are now presented:

- $H \to \gamma\,\gamma$: this channel has been observed and its branching ratio measured, obtaining a value for the production cross section times $\mathcal{BR}_{H\to\gamma\gamma}$ of 64.0 ± 9.6 pb for CMS [19] and $47.9^{+9.1}_{-8.6}$ for ATLAS [20], both compatible with SM expectations;
- $H \to ZZ$: as for the previous decay results for the cross section times the $H \to ZZ$ $\mathcal{BR}$ have been measured by ATLAS [20] and CMS [19], with values respectively of 68^{+11}_{-10} pb and 58.2 ± 9.8 pb, still in agreement with SM expectations;
- $H \to WW$: at CMS [21] an observed (expected) significance of 9.1 (7.1) standard deviations has been measured. The observed cross section times $\mathcal{BR}$ is $1.28^{+0.18}_{-0.17}$ times the SM prediction for a Higgs boson with mass $m_H = 125.09$ GeV/c^2. At ATLAS [22] compatible results for the $H \to WW$ have been obtained considering different production mechanisms: for gluon-gluon fusion $\sigma(gg \to H) \times \mathcal{BR}_{H\to WW} = 11.4^{+1.2}_{-1.1}$ (stat) $^{+1.8}_{-1.7}$ (syst) pb and for VBF $\sigma(pp \to qqH) \times \mathcal{BR}_{H\to WW} = 0.50^{+0.24}_{-0.22}$ (stat) ± 0.17 (syst) pb;
- $H \to \tau\bar{\tau}$: ATLAS [23] measured this process with an observed (expected) significance of 6.4 (5.4) standard deviations. Combining data from Run I (at 7 and 8 TeV) and Run II the measured cross section times $\mathcal{BR}$ is $3.77^{+0.60}_{-0.59}$(stat) $^{+0.87}_{-0.74}$ (syst) pb;
- $t\bar{t}H$: the most important results has been achieved by the CMS [24] experiment. In this work results are characterised by an observed $t\bar{t}H$ signal strength relative to the SM cross section, $\mu = \sigma/\sigma_{SM}$, under the assumption of a Higgs boson mass $m_H = 125$ GeV/c^2. A combined fit of multivariate discriminant distributions in all categories results in an observed (expected) upper limit on μ of 1.5 (0.9) at 95% confidence level.

In Fig.1.4 combined Run I results from ATLAS and CMS [25] for Higgs production cross sections and branching ratios are shown (combined Run II results are not available yet).

From the measured $\mathcal{BR}$ it is possible to extract informations on the Yukawa couplings between the Higgs boson and some particles. It is possible to introduce a set of coupling modifiers $\vec{\kappa}$ to parametrize

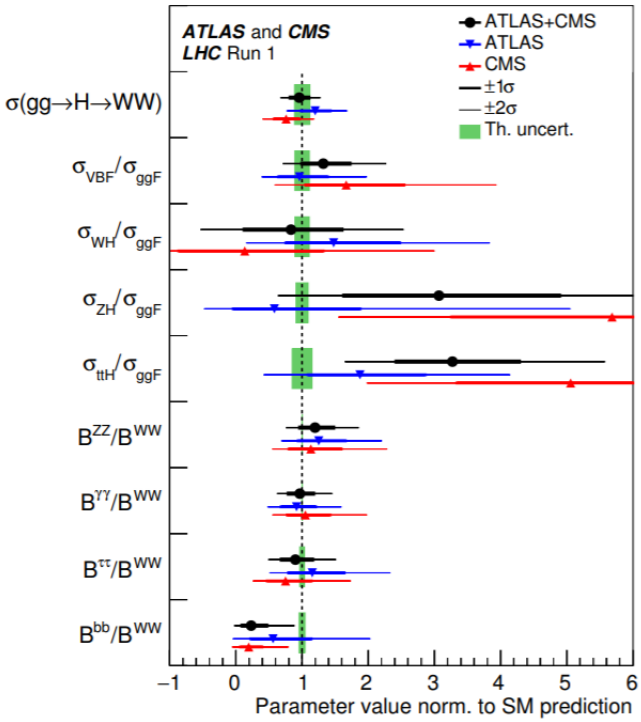

Figure 1.4: ATLAS and CMS combined results (for Run I) for Higgs production cross sections and branching ratios with respect to SM expectations [25].

possible deviations from the SM predictions of the Higgs boson couplings to SM bosons and fermions. For a given production mechanism or decay mode (denoted by "i") a coupling modifier κ_i is defined as

$$\kappa_i^2 = \frac{\sigma^i}{\sigma^i_{SM}} \qquad \kappa_i^2 = \frac{\mathcal{BR}^i}{\mathcal{BR}^i_{SM}} \tag{1.33}$$

where σ^i and $\mathcal{BR}^i$ are respectively the measured production cross section and branching ratio for the i process, while σ^i_{SM} and $\mathcal{BR}^i_{SM}$ are the SM predictions. Possible deviations from the SM can be checked if κ_i are not close to one. In Fig.1.5 combined results from ATLAS and CMS Run I campaign are shown: on the $y-$axis couplings for fermions (weak vector bosons) are defined as $\kappa_F m_F/\nu$ ($\sqrt{\kappa_V} m_V/\nu$) where $\nu = 246$ GeV is the vacuum expectation value of the Higgs field. The dashed (blue) line indicates the predicted dependence for the couplings on the particle mass in the case of the SM Higgs boson. The solid (red) line indicates the best fit result to the $[M, \varepsilon]$ phenomenological model. It is evident the agreement with the SM expectations.

1.4.4 Experimental results for $H \to b\bar{b}$ and $H \to c\bar{c}$

ATLAS has observed the decay $pp \to V(H \to b\bar{b})$ in associated production with a vector boson V, measuring an observed (expected) significance of 5.4 (5.5) standard deviations, while CMS has obtained the best limit so far on the $pp \to V(H \to c\,\bar{c})$ branching ratio considering an Higgs-strahlung production process, with an expected (observed) limit of 70 (37) with respect to SM prediction. By now the inclusive decays $H \to b\bar{b}$ and $H \to c\bar{c}$ have never been observed, therefore limits on the production cross section times branching ratio are computed. In this context analysis performed at LHCb could be really interesting: differently from ATLAS and CMS, LHCb analysis mostly rely on its Vertex Locator (VELO) which allows a good reconstruction of vertices inside the detector. Another interesting feature is the flavour tagging algorithm exploited by LHCb: jets inside the detector are "tagged" with a specific flavour, so that it is possible to separate b and c jets from the light quarks jets. Latest results at LHCb for $pp \to VH(\to b\bar{b})$ and $pp \to VH(\to c\bar{c})$ inclusive processes are presented

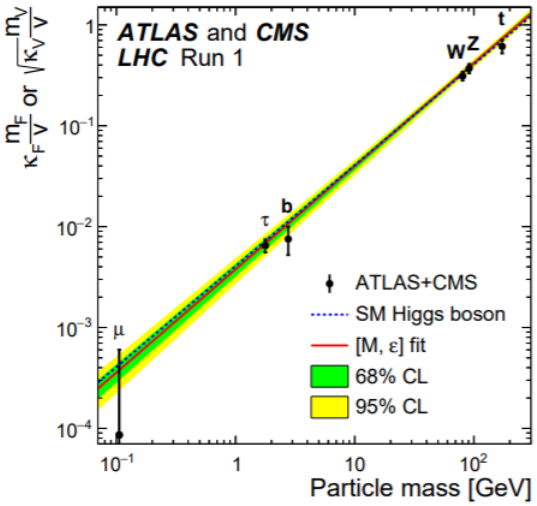

Figure 1.5: Combined ATLAS and CMS Run I results showing the linear relation between the Yukawa couplings with different particles and their masses, compared to SM predictions. The dashed (blue) line indicates the predicted dependence on the particle mass in the case of the SM Higgs boson. The solid (red) line indicates the best fit result to the $[M, \varepsilon]$ phenomenological model [25].

here [26].

1.5 New Physics: theoretical models and results

In this thesis a search for $b\bar{b}$ and $c\bar{c}$ dijet resonances at different masses is performed using LHCb data. Although the approach used in the analysis is model independent, thus results are completely general, these searches can be seen in the light of New Physics (NP) extensions of the SM. There are several theoretical (the cosmological constant Λ, the smallness of the Higgs mass) and experimental elements (neutrino oscillations, Dark Matter) that could not be explained by the SM, thus leading physicists to not consider the SM as the ultimate theory. In this way there are a lot of theoretical extensions of the SM, for which experimentalist are trying to find proofs in order to validate them. In the following two aspects of possible SM extensions that can predict high mass heavy flavour resonances are discussed: SUper SYmmetry and Dark Matter.

1.5.1 SUper SYmmetry

In *SUper SYmmetry* (SUSY) theories [27], every boson is allowed to have a fermionic partner and vice-versa. Among several SUSY theories, the so called *Minimal Supersymmetric Standard Model* (MSSM) [28] is an extension of the SM where to each SM particle a "superpartner" with different spin properties is given; in this way all the bosons (fermions) have a heavy fermionic (bosonic) partner, thus doubling the number of particles. In order to achieve this extension the SM lagrangian needs to be extended by introducing a SUSY-lagrangian, where fermions and scalars are promoted to so called *chiral superfields* while vectors become *real superfields*; by also introducing an appropriate mechanism of SUSY breaking, the superpartners can become really heavy. Finally to avoid some processes impossible for the SM (such as proton decay and lepton number violation) the concept of R-parity is considered: SM particles and their supersymmetric partners have different R-parities, thus a supersymmetric particle can never decay into SM particles. SUSY can be studied at colliders such as LHC: after a proton-proton collision a supersymmetric particle can be produced and by looking at its product decay and measuring the missing transverse momentum it is possible to establish its properties, first of all its mass. Right now no evidences of SUSY particles have been found.

1.5.2 Dark Matter

Only the 4% of our Universe is made by ordinary matter, while almost the 25% is made by the so called *Dark Matter* (DM), a mysterious kind of matter that has mass (since it interacts gravitationally), it is neutral to electromagnetic radiation and it interacts very weakly with ordinary matter; moreover it has been discovered that its behaviour is non-relativistic, so that DM is usually defined as *cold* [29]. Over the past years several approaches have been followed, but the most interesting (but still unlucky) one is the description of DM as a particle χ called *Weakly Interactive Massive Particle* (WIMP): starting from several assumptions DM is assumed to behave as a particle (and in recent models there are several DM particles creating a so called *dark sector*) that can interact with ordinary matter by means of some (pseudo-)scalar or (axial-)vector force carriers, as shown in Fig.1.6. These interactions are called *simplified models* which are an attempt to create more specific models than the ones used by *Effective Field Theories* (EFT) [30] but without considering all the possible kinds of extension of the SM.

(a) (b)

Figure 1.6: Feynman diagrams of typical EFTs describing SM particles interacting with DM particles χ and $\bar{\chi}$ via (pseudo-)scalar (a) or (axial-)vector (b) mediators.

In order to include these interactions the SM Lagrangian needs to be extended by adding new terms such as

$$\begin{aligned} \mathcal{L}_V &= -\, g_{DM} Z'_\mu \bar{\chi} \gamma^\mu \chi - g_q \sum_q Z'_\mu \bar{q} \gamma^\mu q \\ \mathcal{L}_A &= -\, g_{DM} Z'_\mu \bar{\chi} \gamma^\mu \gamma_5 \chi - g_q \sum_q Z'_\mu \bar{q} \gamma^\mu \gamma_5 q \end{aligned} \tag{1.34}$$

which describe for example the interactions mediated by a (axial-)vector force carrier Z'_μ. Similar terms for scalar interactions are inserted. WIMPs are expected to have a mass $m_\chi > 1$ GeV/c^2, thus it is possible to create them in colliders (such as LHC) and find resonances in that mass range. Finally it is possible to link the mass of the mediator to a possible choice for the couplings to matter (g_q) and to DM (g_{DM}), thus considering a particular model among the others (which are basically infinite since the choices for the couplings are infinite). Till now no evidences of WIMPs have been found, thus it is only possible to put limits on the masses of mediators and on couplings, as shown in Fig.1.7 [3].

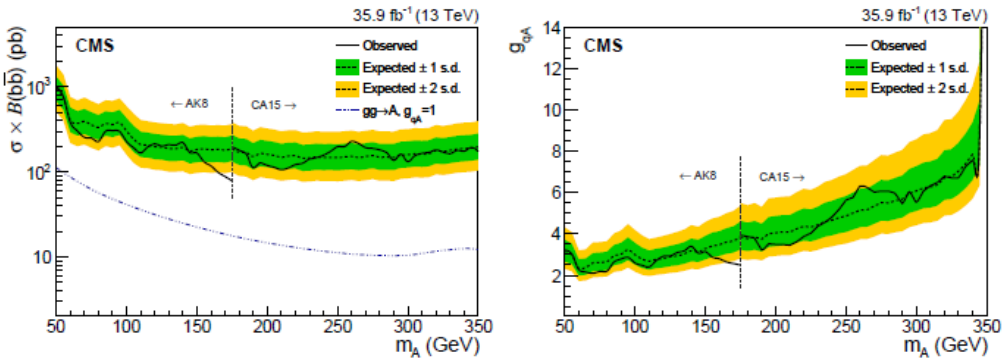

Figure 1.7: CMS results for a DM vector mediator for simplified models: number of events (left) and coupling to quarks (right) with respect to mediator mass; green (yellow) bands show $\pm 1\sigma$ ($\pm 2\sigma$) deviations [3].

Chapter 2

LHCb detector description

2.1 LHC overview

The LHC [31] is a hadron collider that lies 100 m underground, near the border between France and Switzerland, in the vicinity of Geneve, and it has a length of 26.7 km. The basic structure is made by two beam pipes (where hard vacuum is obtained) where two counter-rotating proton or heavy ion beams circulate; these pipes intersect in four so called *interaction points*, where experiments are located. Before entering in LHC, proton are accelerated via other CERN facilities, which provide the initial acceleration stages: proton are produced from a ionizing hydrogen source and accelerated up to 50 MeV by Linac (Linear Accelerator); then they are injected inside the Proton Synchrotron Booster (PSB) where they are accelerated up to 1.4 GeV, and after this they reach a energy of 450 GeV by passing through the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). Finally they reach LHC where they can reach a nominal center of mass (CoM) energy of 14 TeV (for Run II the CoM energy is 13 TeV) by means of 16 radiofrequency (RF) cavities housed in four cylindrical refrigerators; inside these cavities an electromagnetic field oscillates with a frequency of 400 MHz. The beams are not continuous, but they are made by 30-cm-long bunches which have a transverse dimension of the order of the mm. In the nominal configuration the proton beam is made by around 2808 bunches, spaced by 25 ns and made by approximately 11.5×10^{11} protons; moreover these bunches are squeezed to 16 μm at the interaction points. Once inside LHC the proton beams need to be bent to follow the circular shape of the collider; in order to do so 1232 dipolar superconducting magnets are used, as they are cooled to a temperature of 1.9 K using liquid helium, reaching a maximum dipole field of 8.33 T. In addition, 3500 quadrupolar magnets (also held at cryogenic temperatures) are needed to shape and focus the beams. As a general result, the proton beams intersect with a collision frequency of about 40 MHz and with a nominal peak luminosity $\mathcal{L} = 2 \times 10^{34} \mathrm{cm}^2 \mathrm{s}^{-1}$.
The four main experiments are located at the four interaction points: the two General Purpose Detectors have a cylindrical structure and are called ATLAS [32] (A Toroidal LHC ApparatuS) and CMS [33] (Compact Muon Solenoid); then there is ALICE [34] (A Lead Ion Collision Experiment) which is devoted to study a phase of matter where quarks and gluons are free by analyzing heavy ions collisions; finally there is LHCb, designed to study b- and c- hadron physics. In Fig.2.1 a general scheme of the whole LHC and CERN accelerator infrastructures is presented.

2.2 Detector overview

LHCb [35] is a single-arm spectrometer, whose forward angular coverage is approximately from 10 mrad to 300 (250) mrad in the bending (non-bending) plane; this corresponds to a pseudorapidity between 1.8 and 4.9. Considering LHCb standard coordinates system the z axis coincides with the beam direction, while the y axis (vertical direction) is parallel and opposite to gravity acceleration; finally the x axis (horizontal direction) is orthogonal to both axes, forming a right-handed system.

As many other collider experiments also LHCb is made by several sub-detectors, as shown in Fig.2.2,

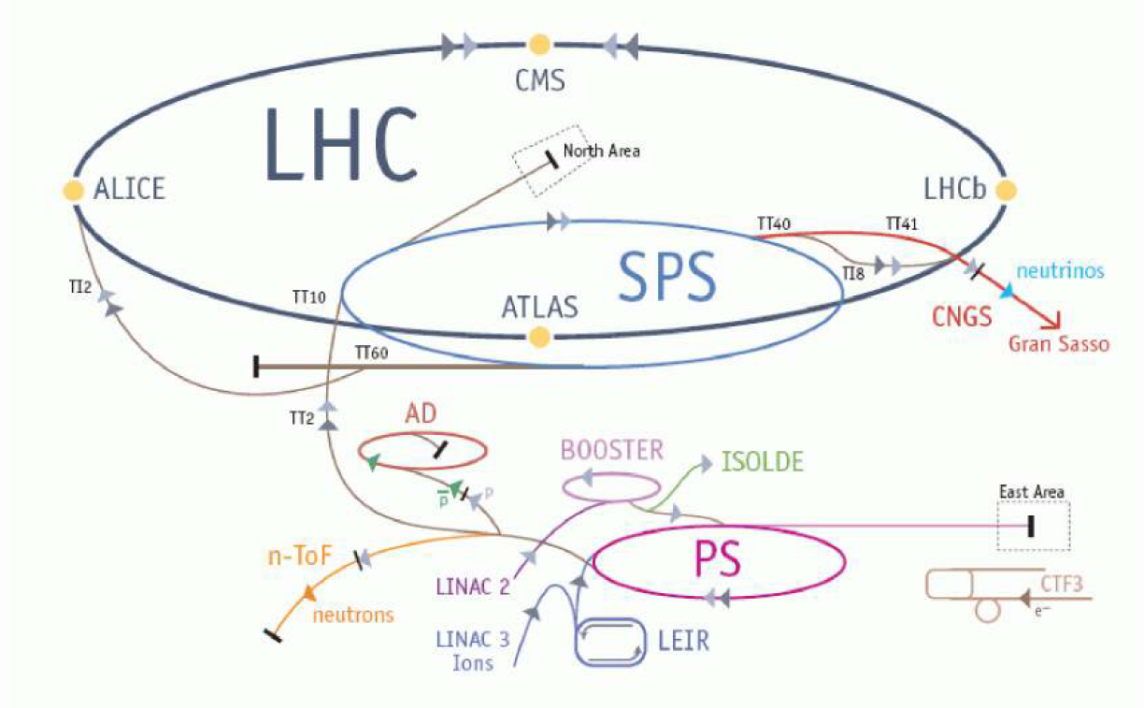

Figure 2.1: Overview of the CERN and LHC infrastructures, with the four main experiments highlighted.

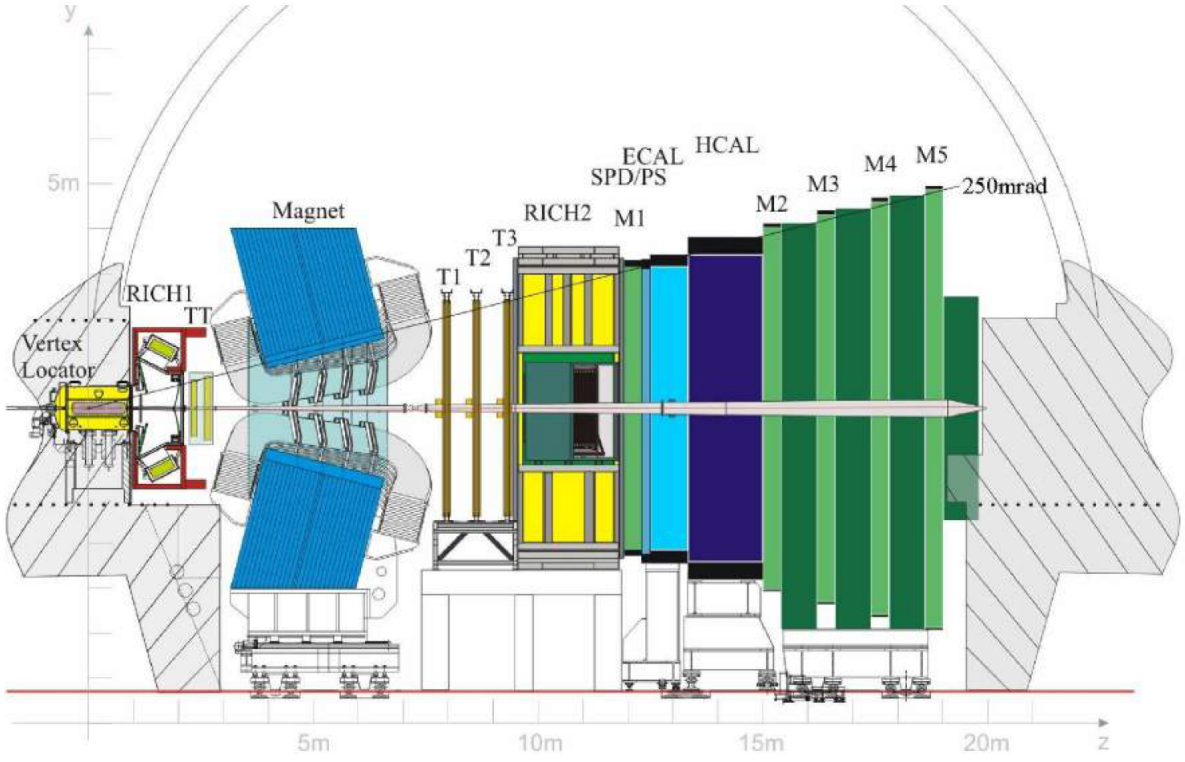

Figure 2.2: LHCb structure and its sub-detectors [35].

which are listed here (starting from the interaction points):

- a vertex locator (VELO) system located near the interaction point;
- a tracking system made by four stations, one called Trigger Tracker (TT) placed upstream of the spectrometer magnet, and the other three (T1, T2 and T3) located downstream of the magnet;
- two RICH imaging Cherenkov detector (RICH1 upstream of the magnet and RICH2 downstream) devoted to separate pions from kaons in the momentum range from 2 to 100 GeV/c;
- a calorimeter system made by a Scintillator Pad Detector and a Preshower system (SPD/PS) and two calorimeters, one electromagnetic (ECAL) and the other one hadronic (HCAL);
- finally there is a muon detection system as the outer part of the detector.

Concerning luminosity it is necessary to highlight LHCb particular feature: in order to achieve great tracking performances near the interaction point the optimal collision rate is of maximum 2.5 proton-proton interactions per bunch crossing. LHCb luminosity has to be reduced via a *luminosity levelling* technique [36]. In this way the LHCb luminosity is kept constant to a value of $4.5 \times 10^{32}\text{cm}^2\text{s}^{-1}$, 2 orders of magnitude lower than LHC nominal luminosity. In Table2.1 CoM energy and integrated luminosity values for the Run II campaign are shown.

year	$\sqrt{s}$ [TeV]	$\mathcal{L}_{int}$ [fb^{-1}]
2015	13	0.33
2016	13	1.67
2017	13	1.81
2018	13	2.19

Table 2.1: LHCb CoM energy and integrated luminosity during Run II campaign.

2.2.1 Vertex Locator

The first sub-detector near the interaction point is the VErtex LOcator (VELO) [37], a system devoted to measure tracks near the interaction point in order to identify secondary vertices, which could be a signature of b and c hadrons decay. The VELO is made by two different types of semicircular silicon sensor: the R sensor is measuring the radial coordinate r (from 8 mm to 42 mm) while the Φ sensor is devoted to the measurement of the azimuthal coordinate (from 15 mrad to 390 mrad). Both the two sensors have a diameter of 84 mm and a thickness of 300 μm, and they have different geometries, as shown in Fig.2.3.

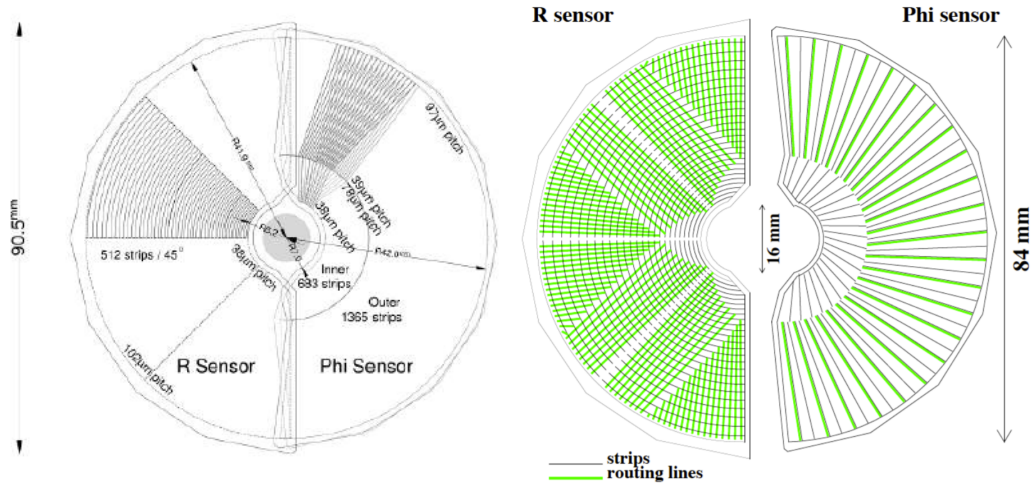

Figure 2.3: VELO's sensor structure, highlighting the $r\phi$ geometry and the different structures between R and Φ sensors [37].

The R sensor strips are arranged into four approximately 45° segments and have routing lines perpendicular to the strips. The Φ sensor has two zones with inner and outer strips. The routing lines of the inner strips are orientated parallel to the outer strips. The inter-strip pitch varies from approximately 40 to 100 μm across the sensor. The VELO is made by 21 of these sensors, placed along a distance of 1 m and parallel to the beam axis, as shown in Fig.2.4.

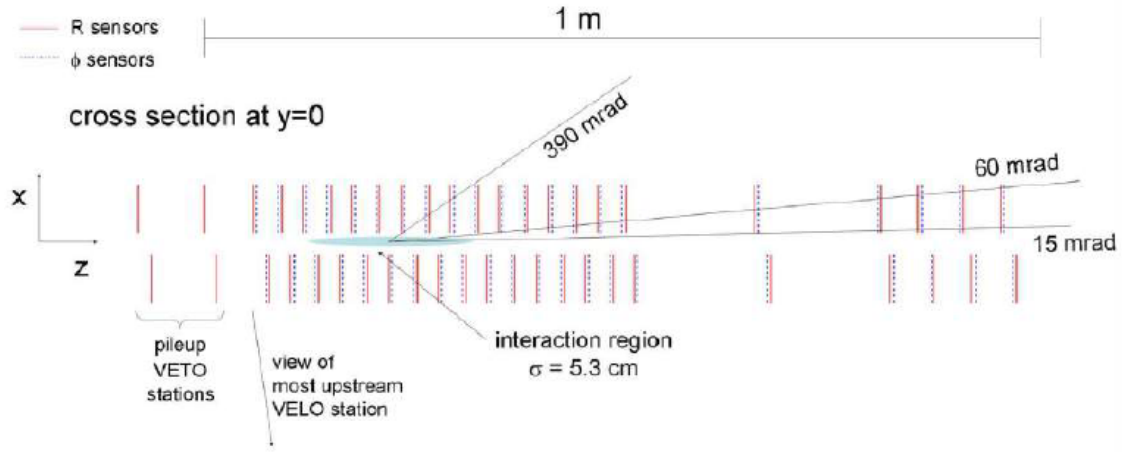

Figure 2.4: VELO structure and sensors position [37].

The nominal distance between the sensors and the beam axis is 8 mm, and since the beam site is greater during the injection phase, the sensors in an open configuration as shown in Fig.2.5, till the beam is stable. The individual hit resolutions of the sensors have been measured during test beams, and the best raw resolution obtained is around 7 μm.

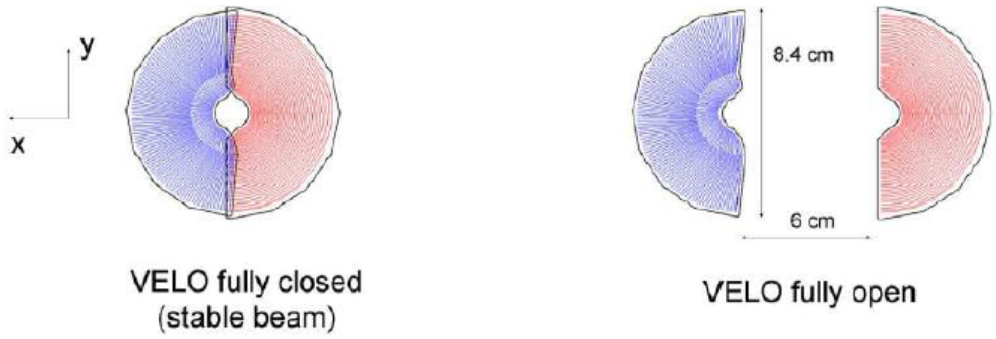

Figure 2.5: VELO sensor closed configuration (left) and open configuration (right) [37].

2.2.2 Magnet

Between the TT and the other three T1-T3 tracking stations, a dipole magnet is placed in order to measure charged particles momentum. Differently from the magnets inside LHC, this magnet is warm (non-superconductive) and it is made by two aluminium conical coils which are placed opposite to each other inside the magnet yoke, as shown in Fig.2.6. Moreover the magnetic field is parallel to the y axis (the xz plane is the bending plane) and the momentum measurement cover the full acceptance of the detector; the direction of the magnetic field can be inverted by reverting the electric current that generates it, thus allowing measurements with two opposite polarities. The integrated magnetic field for tracks 10 m long is about 4 Tm, while there is still a residual magnetic field of 2 Tm inside the two RICHs.

2.2.3 Tracking system

The tracking system is made by four stations with a planar structure perpendicular to the beam axis: the Trigger Tracker (TT) placed between the first RICH detector and the magnet, and the

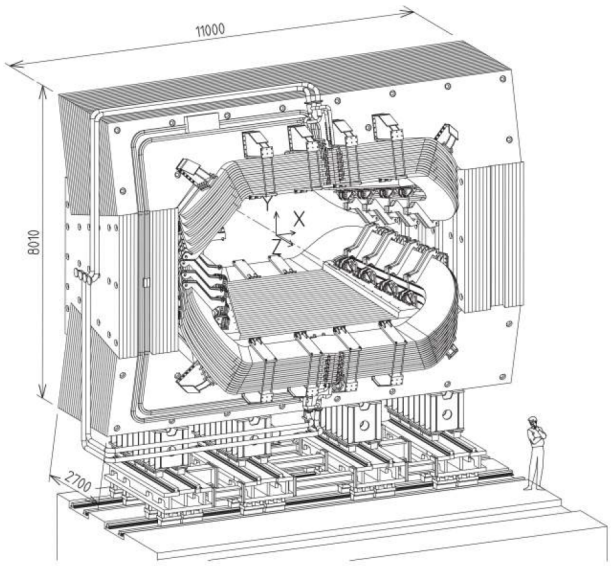

Figure 2.6: A representation of the magnet inside LHCb (units expressed in mm).

three stations T1, T2 and T3 placed between the magnet and the second RICH detector. The TT station and the inner part of the T1-T3 stations, called Inner Tracker (IT), are composed by 200 μm wide silicon-microstrip sensors, displaced in four layers of three different types depending on the angle with respect to the y axis: the $x-$layer has vertical strips, while the $u-$ and $v-$ layers have strips respectively rotated to $-5°$ and $+5°$ with respect to the y axis; the overall structure (from the interaction point) is $x-u-v-x$. In Fig.2.7 a typical $v-$layer of the TT station (left) and a typical $x-$layer of an IT station (right) are shown.

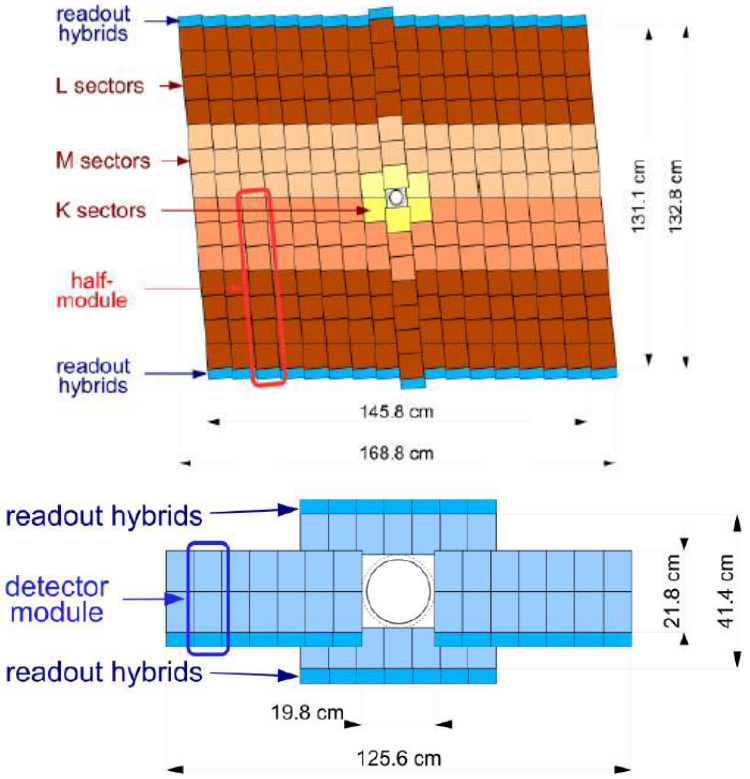

Figure 2.7: Representation of a $v-$layer of the TT station (up) and of a $x-$layer of an IT station (down).

The TT station is 150 cm and 130 cm high, thus covering the full acceptance of the experiment. The IT inside the other three stations is instead 120 cm wide and 40 cm high, covering approximately 1.3%

of LHCb acceptance; since in this region the particle flux is pretty high the silicon detector must have radiation protection properties. For both the TT and IT the single spatial distribution is around 50 μm.
The external region of the T1-T3 stations is called Outer Tracker (OT) and it is a drift-time detector with an acceptance of 300 mrad (250) mrad in the horizontal (vertical) plane; its structure is an array of straw-tube modules, each one of them formed by two layers of drift-tubes with an inner diameter of 4.9 mm. The tubes contain a gas mixture of Ar (70%), CO_2 (28.5%) and O_2 (1.5%), chosen such that the drift distance resolution is about 200 μm. In Fig.2.8 a scheme of the tracking system is shown.

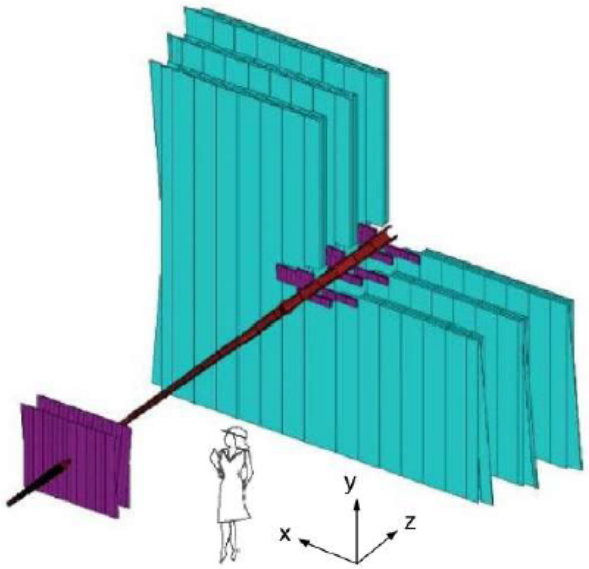

Figure 2.8: Schematic view of the tracking system, with the IT (violet) and the OT (light-blue);

2.2.4 Ring Imaging CHerenkov detectors (RICH)

One of the main feature of LHCb is the particle identification, usually called PID. The two Ring Imaging CHerenkov detectors RICH1 and RICH2 (placed respectively before the TT and after the T3) can identify charged hadrons (π, K, p); the fact that two RICHs are needed is to cover the full momentum range, quite impossible with just one detector.

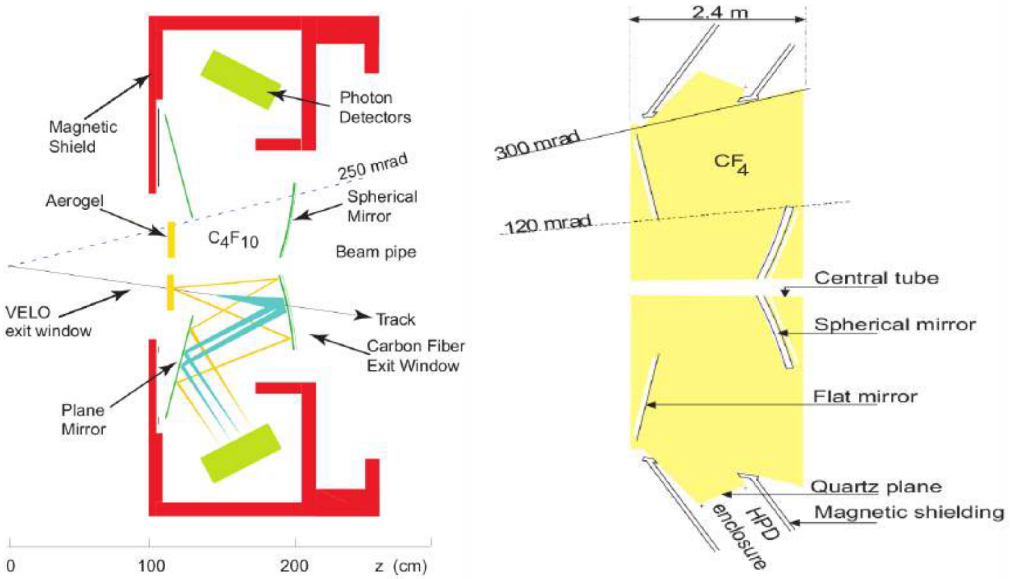

Figure 2.9: Schematic view of the two RICHs: RICH1 (on the left) and RICH2 (on the right).

The two detectors basically share the same features, but still have some differences, mainly related to their position inside the detector: a schematic view of their structure can be seen in Fig.2.9. The RICH1 covers a momentum range between 1 to 60 GeV/c and it is located between the VELO and the TT station, allowing the coverage of almost all the LHCb acceptance, from ±25 mrad to ±300 (±250) mrad along the horizontal (vertical) direction; it is surrounded by a magnetic shield in order to cope with the residual magnetic field. The RICH1 contains both a solid aerogel and a fluorobutane (C_4F_10) gas radiator, and spherical and plane mirrors reflect the Cherenkov light to Hybrid Photo Detectors (HPDs) where photons are converted into electrons.

The RICH2 is located between the last tracking station (T3) and the first muon station (M1), and besides its limited angular coverage (from ±15 mrad to ±120 (±100) mrad along the horizontal (vertical) direction) it covers the region where particles with high momentum are produced (in a range from 15 GeV/c to 100 GeV/c). Inside RICH2 there is a CF_4 gas radiator, and as for RICH1 the Cherenkov light is reflected by both spherical and flat planes and collected by HPDs.
To conclude, from the two RICHs is also possible to extract informations on charged leptons (e,μ) by considering also results obtained from the calorimeters and the muon system.

2.2.5 Calorimeters

The calorimeters system is composed by a Scintillating Pad Detector (SPD) and a Preshower detector (PS), together with an electromagnetic (ECAL) and a hadronic calorimeter (HCAL). The system is located between the M1 and the M2 stations of the muon system. The calorimeter system has several purposes:

- first of all it selects and identifies hadrons, electrons and photons;
- it also measures the energy of neutral particles (photons, π^0, neutral hadrons), which is fundamental in order to reconstruct jets.

The first part is the SPD/PS system, which is devoted to the identification of electrons and photons. With a covered angular acceptance from ±30 mrad to ±300 (±250) in the horizontal (vertical) plane, this system is essentially made by a 15 mm lead converter in between two almost identical scintillator pads of high granularity, respectively called SPD (the first) and PS (the last). The light produced after scintillation is then collected by wavelength-shifthing (WLS) fibres which are then connected to photomultiplier tubes (PMT). The identification of photons and electrons is possible since in the SPD photons do not produce any signal, while electrons do; moreover after passing the lead separting the SPD from the PS, photons do produce a shower in the PS, thus allowing the separation of photons from electrons; finally, since hadrons do not produce showers in the SPD/PS system, it is possible to separate them from electrons.
In the ECAL electrons and photons are absorbed and they deposit their energy. The ECAL is placed at 12.5 m from the interaction point, and it has the same acceptance than the SPD/PS system; it is basically composed by cells made by alternating 66 layers of 2 mm thick lead tiles and 4 mm scintillator tiles, for a total length of 42 cm corresponding to almost 25 radiation lengths. Again the scintillation light is collected by WLS fibres and measured by PMTs. The granularity of the ECAL allows a good resolution given by

$$\left(\frac{\sigma_E}{E}\right)_{ECAL} = \frac{10\%}{\sqrt{E}} \oplus 1\% \tag{2.1}$$

(where E is the energy in GeV and $\oplus$ indicates the sum in quadrature) and it guarantees a one-to-one correspondence with the SPD/PS pads.
Finally in the HCAL charged and neutral hadrons deposit their remaining amount of energy, as they are absorbed by 1 cm iron layers interspersed with scintillating tiles, for a total length of almost 5.6 radiation lengths. Thus the structure is identical to the ECAL, again with the scintillation light collected by WLS fibres and measured by PMTs. The HCAL resolution is given by

$$\left(\frac{\sigma_E}{E}\right)_{HCAL} = \frac{69\%}{\sqrt{E}} \oplus 9\% \tag{2.2}$$

with E the energy expressed in GeV. The HCAL granularity is different from the ECAL, giving in general a worse resolution. In Fig.2.10 the granularity of ECAL and HCAL is shown.

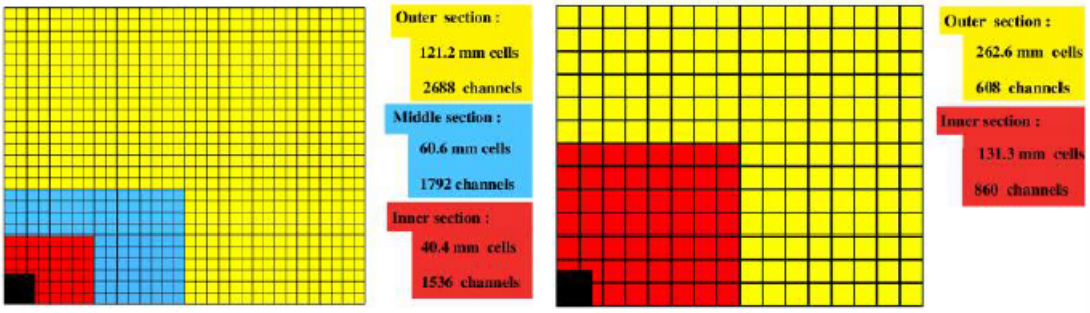

Figure 2.10: Segmentation of SPD/PS and ECAL (left) and HCAL (right).

2.2.6 Muon system

The outer part of LHCb is mainly devoted to the detection of muons. The muon system is composed by five stations (M1-M5) of rectangular shape, perpendicular to the beam axis as shown in Fig.2.11 (left); it has an angular coverage from ±20 (±16) mrad to ±306 (±258) mrad in the horizontal (vertical) plane. The five station are not placed together: the M1 station is located right in between the RICH2 and the SPD/PS system (in this way it is possible to obtain information on the muons transverse momentum for the trigger) while the M2-M5 stations are placed downstream of the calorimeters. These last stations are interspersed with 80 cm thick iron absorbers in order to select penetrating muons; in this way the total absorbing length is around 20 interaction lengths, corresponding to muons having a momentum greater than 6 GeV/c.

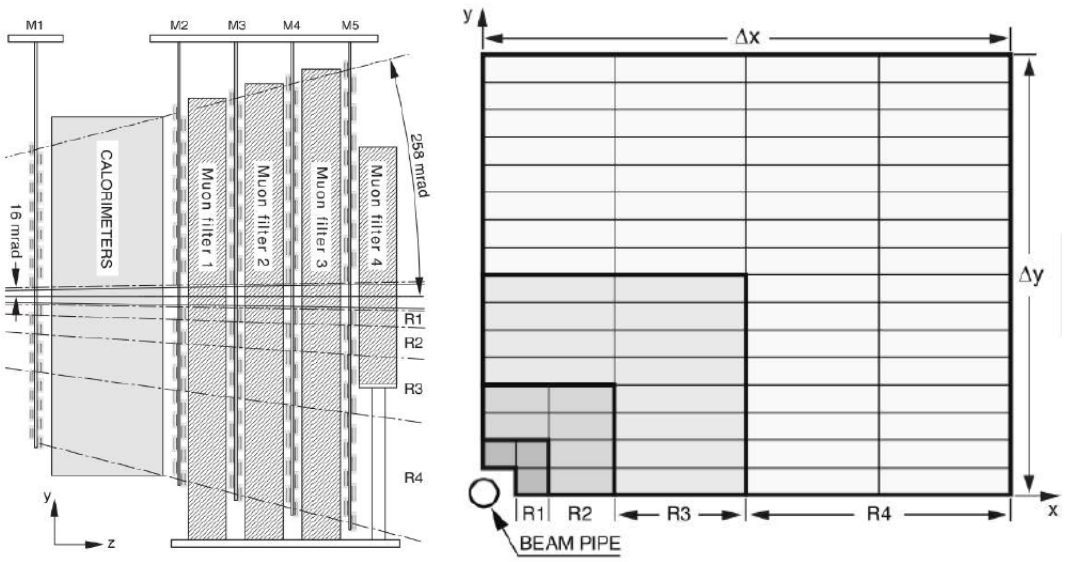

Figure 2.11: Muon system and its five stations (left) and front view of a quadrant of a muon station (right) where each rectangle represents a muon chamber.

The five stations are mainly Multi-Wire Proportional Chambers (MWPC) with the only exception of the inner part of the M1 station where triple-GEM detectors are used. Since the main purpose is to provide space measurements of tracks, the five stations are divided in logical pads: the M2-M5 stations are made by 276 chambers for a total of 1368 MWPC (with a typical layout shown in Fig2.11 (right), filled with a gas mixture of $Ar/CO_2/CF_4$ in a $40 : 55 : 5$ proportion. At a voltage of $2600 - 2700$ V the ionization gain is around 10^5 by means of 3×10^6 wires of gold-plated tungsten placed inside the chambers. The achieved time resolution is around 5 ns. The three GEM detectors in the inner

part of M1 consist of three gas ($Ar/CO_2/CF_4$ in 45 : 15 : 40 proportion) electron multiplier foils placed between anode and cathode plates. The time resolution is better than the one for the MWPC, reaching almost 3 ns; in this way GEMs are used instead of MWPCs to sustain the high rate (almost 500 kHz/cm^2) of charged particles in this region.

2.3 Trigger

As explained in Sec.2.2 LHCb operates at an average luminosity which is much lower than the maximum designed luminosity of LHC, with a bunch crossing rate around 40 MHz. At the beginning of the Run II data taking (2015) the storage system was able to write and save events at a rate of 12.5 kHz, corresponding to almost 0.6 GB/s. It is easy to understand that a trigger is necessary to reduce the rate from 40 MHz to 12.5 kHz, by selecting only a small fraction of potentially interesting events. The LHCb trigger is designed to operate in two steps: the first one is a hardware trigger stage called *Level 0* (L0) operating synchronously with the bunch crossing rate, while the second one is a software trigger called *High Level Trigger* (HLT) which operates on a processor farm. In Fig.2.12 a scheme of the LHCb trigger is shown.

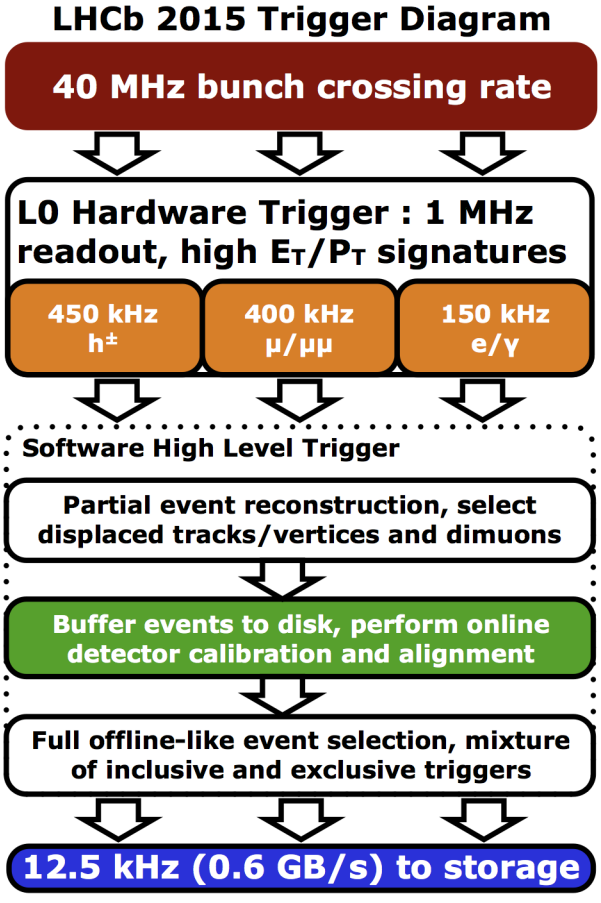

Figure 2.12: General scheme of the LHCb trigger used during Run II.

2.3.1 Level 0 (L0) trigger

The L0 trigger tries to spot particles with high transverse momentum p_T or high transverse energy E_T, since events concerning b and c hadrons are characterized by these signatures. Starting from the calorimeters and the muon system, which are connected to the L0 Decision Unit (L0-DU), the L0 trigger decides if an event is accepted or not, reducing the rate from 40 MHz to 1 MHz.
In order to reject events with multiple interactions, the L0 trigger applies a cut on the number of

hits inside the SPD, applying a so called *Global Event Cut* (GEC); having applied the GEC, an event (thus a track) is accepted and identified if one of the following conditions is satisfied:

- *L0-muon*: a particle is identified as a muon if its p_T is above a certain threshold;
- *L0-Dimuon*: two tracks are identified as muons if the product $p_T^1 \cdot p_T^2$ is above a given threshold;
- *L0-Photon*: a particle release a cluster in the ECAL and has a E_T above a given threshold;
- *L0-Hadron*: a particle release a cluster in the HCAL and has a E_T above a given threshold;
- *L0-Electron*: a particle has a E_T above a given threshold and before leaving a cluster in the ECAL it hits both the PS and (at least one time) the SPD.

The thresholds used in the L0 trigger are decided in order to maximize the trigger efficiencies of benchmark decays. The selected thresholds for this thesis are shown in Table2.2.

line	p_T [GeV/c]	$p_T^1 \cdot p_T^2$ [GeV^2/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

Table 2.2: Level 0 trigger thresholds considered in this thesis.

2.3.2 High Level Trigger (HLT)

Once an event is accepted by the L0 trigger, it is then processed by the HLT which runs on a large computing facility called *Event Filter Farm* and it reduces the rate from 1 MHz to a final rate of 12.5 kHz. The HLT is divided in two stages:

- HLT1: it partially reconstruct the L0-accepted events using also the informations coming from the VELO and the T1-T3 stations. Depending on the L0 decision there are different requirements, but generally speaking tracks with high p_T and high displacement from the interaction point are selected;
- HLT2: it performs a full reconstruction and selection of an event by means of inclusive and exclusive algorithms, respectively selecting all particles detected or considering a specific decay mode, after having performed a detector calibration and alignment. In this thesis a specific HLT2 trigger line has been considered: two reconstructed jets are selected with transverse momentum $p_T > 17$ GeV, and both jets have to include a Secondary Vertex SV within the jet cone. Details on the jet reconstruction and tagging algorithms, which are the same in the online system and in the offline analysis, will be given in the next chapter.

2.4 Tracking performances

To conclude the detector overview it is interesting to evaluate the tracking performances of the LHCb detector. In the tracking system particle momentum is determined by measuring the curvature of particle tracks; in order to do so pattern recognition algorithms use the hits coordinates in all the tracking sub-detectors and they reconstruct the particle trajectory performing a Kalman fit [38]. In this way particle tracks are classified in four categories:

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

- *VELO tracks*, which have hits only in the VELO.

In order to determine the momentum resolution ($\delta p/p$) of long tracks, benchmark processes are considered such as $J/\psi \to \mu\,\bar{\mu}$ decays, collected from a data sample with a particular trigger configuration to select high energy di-muons [39]; for momentum particles below 20 GeV/c resolution is about 0.5% while it is 0.8% for momentum particles around 100 GeV/c. For the invariant mass resolution (σ_m/m where σ_m is the peak width of the resonance at a mass m) six different resonances have been used, again observed in di-muons samples: J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and Z boson [39]; for particle mass below 10^4 MeV/c^2 resolution is 0.45% while it is around 1.9% for masses around 10^5 MeV/c^2. Results both for momentum (left) and mass (right) resolution are shown in Fig.2.13:

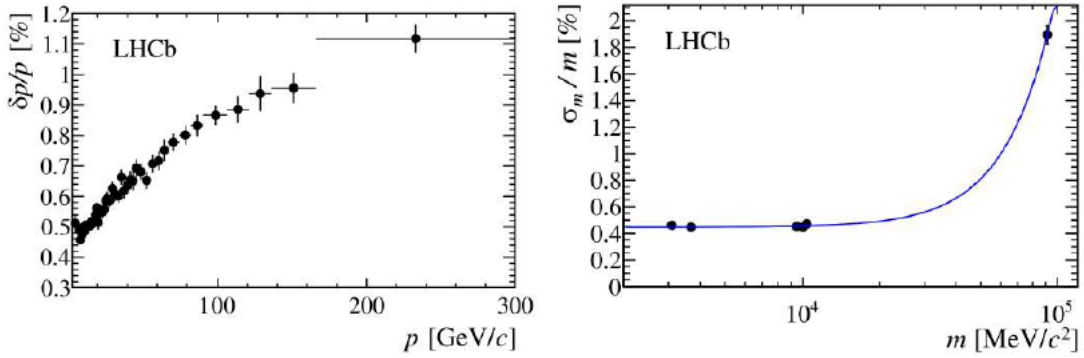

Figure 2.13: Momentum resolution (left) and mass resolution (right) for the tracking system [39].

From the detected and reconstructed tracks it is also possible to define the Primary Vertex (PV), which is the proton-proton interaction point where particles are generated. First of all tracks are clustered in seeds [40]; if the cluster has less than six tracks, it is discarded. After this, for each seed the PV position is calculated minimizing:

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \frac{d^2_{0i}}{\sigma^2_{d_{0i}}} \tag{2.3}$$

where d_{0i} is the impact parameter of the $i-$track and $\sigma_{d_{0i}}$ its error. If one or more tracks (inside a seed) have $\frac{d_0}{\sigma_{d_0}} > 4$ then the one with the highest $\frac{d_0}{\sigma_{d_0}}$ is removed from the cluster, and a new PV position is computed by minimizing the new χ^2_{PV}. This procedure goes on till there are no more tracks to discard. The PV reconstruction efficiency and resolution are obtained from simulations [40]: the average efficiency goes from 90.0% to 97.5% while the probability of reconstructing a false PV is about 1%; regarding resolution typical values are about 8 μm, 10 μm and 50 μm respectively for the x, y and z coordinates.

Chapter 3

Jets reconstruction, flavour tagging and event selection

3.1 Introduction

Jets are the main physical objects reconstructed in this analysis and here it will be described how they are reconstructed and tagged by LHCb algorithms. In the first part of this chapter an overview of the jets reconstruction algorithm will be given, while in the second part the flavour identification will be described, discussing the performance. From now on the speed of light is considered to be $c = 1$.

3.2 Simulations

Events are simulated by means of MC simulations. Different aspects of interactions inside a detector have to be simulated, and this is achieved by using different tools. The LHCb experiment uses several software packages to reconstruct events, apply the software trigger, analyse data and generate simulated events. All the software are based on ROOT [41] and on the Gaudi [42] framework, and they are described here in the following:

- first the pp interactions are generated. In this thesis PYTHIA6 [43] and PYTHIA8 [44] are used;
- PYTHIA is also used to reproduce the hadronisation process, while the decays of hadronic particles are described by EVTGEN [45];
- finally also the interaction between particles and the detector materials is simulated; this is achieved using GEANT4 [46, 47] as described in [48];
- as a last step BOOLE [49] manages the last part of the simulation providing sub-detectors responses using GEANT4. Digitization of the front end electronics is also reproduced.

3.3 Jets reconstruction algorithms

For LHCb it is not possible to rely only on calorimeters for jet reconstruction, since their resolution (particularly for the HCAL) is not good enough to measure jets energy; instead the tracking system is excellent, with an efficiency of 97% for charged tracks and a resolution of about 0.5%. It is thus clear that the jets reconstruction algorithm has to involve both calorimeters and the tracking system, resulting in a hybrid algorithm substantially different with respect to typical algorithms based on fully calorimetric reconstruction used by other experiments. The tracking system is used to select charged particles while calorimeters select neutral particles coming from jets. The LHCb jets reconstruction algorithm has several steps, which will be described in the following:

1. Particle flow: specific criteria are applied to tracks and calorimeter clusters as to select input particles for the algorithm;
2. Anti-k_T algorithm [50]: particles selected by the Particle Flow are clustered in jets;
3. $E-$recombination scheme: jets 4$-$momenta are computed as the sum of the jets' particles 4$-$momenta;
4. Jet Energy Correction: finally the jet 4$-$momentum is corrected by a factor depending on jet kinematics.

3.3.1 Particle flow

The first step of the jets reconstruction algorithm is the Particle Flow, where particles are selected via certain criteria and used as inputs for the anti-k_T algorithm. Depending on the type of particle there are different criteria:

- charged particles, particles related to tracks;
- isolated neutral particles, coming from requirements on clusters in the calorimeter not associated to tracks;
- non-isolated neutral particles, with requirements on calorimeters clusters associated to at least one track.

In the following a description of the different criteria is presented.

Charged particles

First of all, tracks are selected depending on the category they belong (*long*, *downstream*, *upstream* and VELO tracks) and afterwards requirements are applied considering the following quantities:

- the track transverse momentum p_T;
- the χ^2 of the Kalman fit [38];
- P_{ghost}, defined as the probability for a track to be reconstructed wrongly by the algorithm, thus not associated to a real particle;
- the momentum resolution $\frac{\sigma(q/p)}{(q/p)}$ where q is the particle charge and p the particle momentum.

In the following Table3.1 a list of all the requirements for all the tracks categories is presented.

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

Table 3.1: List of requirements for all tracks categories in the Particle Flow.

Moreover using RICHs informations it is possible to assign a particle type and mass to each track: the possible particle categories are $p/\bar{p}$, $\pi^{\pm}$, $\mu^{\pm}$, $e^{\pm}$ and $K^{\pm}$.

Isolated neutral particles

Neutral particles detected by the ECAL are photon and π^0 which decays into two photons. In order to distinguish between them, the shapes of the ECAL clusters are used to compute photon or π^0 hypotheses. Moreover the emission of two photons resulting in te π^0 decay can happen in two ways: *merged* π^0, when the two photon are emitted almost collinearly and thus produce a single cluster,

and *resolved* π^0, when the two photons are detected as two separate clusters. As for charged particles requirements are applied to specific quantities:

- the cluster transverse energy E_T;
- the likelihood for the photon hypothesis (called *PhotonID*);
- the $\chi^2_{track-cluster}$ which is the χ^2 associated to the probability that a cluster is associated to a track.

In Table 3.2 the requirements for all these quantities related to the three possible types of cluster are presented.

	merged π^0	resolved π^0	photons
E_T [MeV]	-	-	> 200
PhotonID	-	> -2	-
$\chi^2_{track-cluster}$	> 25	> 25	> 25

Table 3.2: List of requirements for all tracks categories in the Particle Flow.

For the HCAL there are no particle identification requirements, while for $\chi^2_{track-cluster}$ different requirements are applied for different energy thresholds: for $E < 10$ GeV a $\chi^2_{track-cluster}$ greater than 25 is required, while for $E > 10$ GeV the $\chi^2_{track-cluster}$ has to be greater than 15.

Non-isolated neutral particles

The last category analyzed by the Particle Flow considers clusters that are associated to tracks; in order to select non-isolated neutral particles there are several steps:

1. requirements on ECAL and HCAL $\chi^2_{track-cluster}$ are applied: $\chi^2_{track-cluster,ECAL} < 25$ and $\chi^2_{track-cluster,HCAL} < 25$ (< 15) for energies below (above) 10 GeV;
2. clusters are grouped such that different groups do not share tracks, both in the ECAL and HCAL;
3. an empirical parametrization of E/p as a function of p (where E is the cluster energy released by charged particles and p the track momentum) is used in order to compute the expected released energy E^{exp} in the calorimeters;
4. if $E^{exp} > 1.8E^{meas}$ where E^{meas} is the measured energy of the cluster then the cluster is discarded;
5. if otherwise $E^m > 1.8E^{exp}$ then E^{exp} is subtracted from E^m obtaining $E_{subtracted}$;
6. finally if $E_T > 2$ GeV then $E_{subtracted}$ is selected as a non-isolated neutral particle and used as input in the following anti-k_T jet clustering algorithm.

3.3.2 Jet clustering with anti-k_T algorithm

The output 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 Sec.2.4 and afterwards the clustering algorithm is applied to tracks that share the same PV; if more PVs are found the clustering algorithm is performed for all of them. LHCb uses the so called *anti-k_T* algorithm [50] that works as follow:

- a list of all the input particles is created;
- for each combination of two particles i and j the following quantity ("distance") is computed

$$d_{ij} = \min\left(k_{t,i}^{-2}, k_{t,j}^{-2}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

where $k_{T,i}$ is the particle i transverse momentum, $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y_i and ϕ_i respectively the rapidity and the azimuthal angle of the particle i, and R is the radius parameter to be chosen;

- the same quantity is computed for each particle i and the beam B, by doing

$$d_{iB} = k_{t,i}^{-2} \tag{3.2}$$

- the smallest quantity between d_{ij} and d_{iB} is found;
- if d_{ij} is the smallest then the two particle i and j are combined in a single particle summing their 4−momenta. The new particle substitutes the two particles in the input list and the algorithm restarts from step 2;
- if the smallest quantity is d_{iB} then the particle i is removed from the list and it is considered a jet;
- if there are no more particle the algorithm stops, otherwise it starts again from step 2.

The main consequence is that soft particles will tend to cluster with hard particles instead of clustering among themselves, since for hard particles k_T^{-2} is smaller; in this way if there are no hard particles within a $2R$ distance, a hard particle will accumulate all soft particles aroud within a R distance, resulting in a conical jet. This is not always the case, since another hard particle could be within a $[R,2R]$ distance and thus two non-perfectly-conical jets are created.
As a last point, the choice for the radius parameter R has to be optimized. R depends on the experimental environment, the detector properties and the processes under study: large R may bring to merge jets that are originated by different particles, while small R can fragment a single jet. At LHCb the optimal radius parameter has been found to be $0.5 < R < 0.7$, and for this thesis a value of $R = 0.5$ has been chosen.

3.3.3 E−recombination scheme

Once all the particles have been clustered the jet 4−momentum is computed using the "E−recombination scheme": considering particles 4−momenta $p_i = (E_i, \vec{p}_i)$ the jet 4−momentum p_{jet} is defined as

$$p_{jet} = (E_{jet}, \vec{p}_{jet}) \tag{3.3}$$

where

$$E_{jet} = \sum_i E_i \qquad \vec{p}_{jet} = \sum_i \vec{p}_i \tag{3.4}$$

At this point it is possible to validate jets reconstruction by means of Monte Carlo (MC) samples. In these samples both jets and MC jets are reconstructed. The main difference between MC jets (jet_{MC}) and reconstructed jets (jet_{reco}) is that jet_{MC} are clustered by the anti-k_T algorithm using all the stable truth-level particles (with lifetime $\tau > 10^{-8}$s) which thus have the true kinematical values, while for jet_{reco} reconstructed particles are used (invisible particles as neutrinos are removed from the list, not to bias the evaluation of the reconstructed energy). In order to associate a jet_{MC} with a jet_{reco} the distance[1] ΔR between them has to be lower than 0.4, and if more than one jet_{MC} is found in this range then the closest in distance is selected. It is thus possible to compute the ratio between the energy of jet_{reco} and jet_{MC} for charged particles, for neutral particles. and for isolated neutral particles as functions of p_T, η and ϕ of jet_{reco}. Results are shown in Figs.3.1, 3.2 and 3.3.

The figures above show that for charged particles the agreement between MC and data is excellent, while this does not happen for neutral particles (both isolated and non-isolated): this is due to the worse resolution of the calorimeters with respect to the tracking system; that is also why results are better for neutral particles (where informations from the trackers are used, since clusters are associated to tracks) than for isolated neutral particles.

[1]Remember that in the (η,ϕ) plane ΔR is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

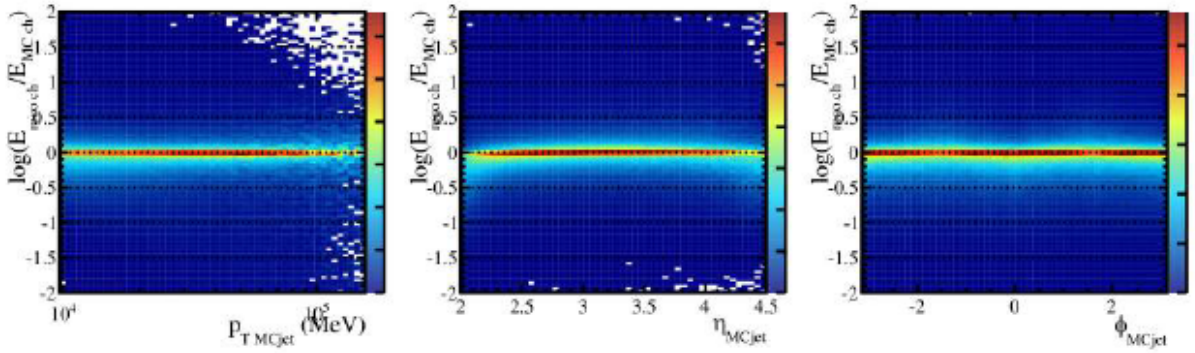

Figure 3.1: jet_{reco}/jet_{MC} energy ratio for charged particles as a function of p_T, η and ϕ of the MC jet.

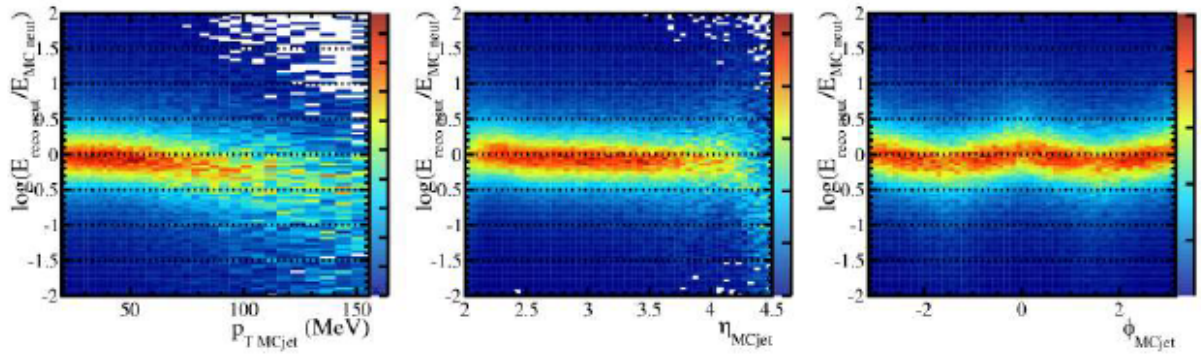

Figure 3.2: jet_{reco}/jet_{MC} energy ratio for neutral particles as a function of p_T, η and ϕ of the MC jet.

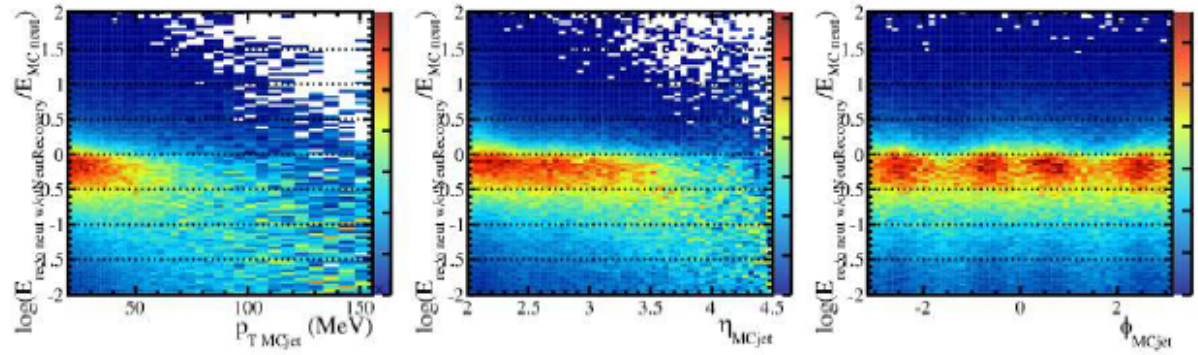

Figure 3.3: jet_{reco}/jet_{MC} energy ratio for isolated neutral particles as a function of p_T, η and ϕ of the MC jet.

3.3.4 Jet Energy Correction

As a final step it is necessary to correct the energy of the reconstructed jet jet_{reco} since it will be different with respect to the energy of the associated MC jet jet_{MC}. In order to perform this correction, a multiplicative factor is computed in simulation:

$$E_{jet_{MC}} = k_{MC} E_{jet_{reco}} \tag{3.5}$$

The factor k_{MC} takes into account all the effects such as the pile-up, the noise and the non-uniformity of the detector, and since the jet_{MC} direction is found to be compatible with jet_{reco} then the same correction is applied to each component of the 4−momentum. It is possible to evaluate k_{MC} starting from simulations of b, c. light quarks and gluons jets at a CoM of 13 TeV; it is found to be non-uniform with respect to the jet η, ϕ and the fraction of charged particles inside the jet *cpf*, and a parametrization of k_{MC} as a function of p_T has been obtained with a cubic model. The various distributions are shown in Fig.3.4, both for different R in the jet clustering algorithm ($R = 0.5$ in blue, $R = 0.7$ in red). k_{MC} may also depend on the jet flavour; the relative systematic uncertainty on the jet of flavour i is defined as $max\left(\frac{k^i_{MC} - k^j_{MC}}{k^i_{MC} + k^j_{MC}}\right)$ where $k^{i,j}_{MC}$ is the correction calculated for a jet of flavour i,j: the uncertainty obtained is around 2.7% (2.6%) for jets with $R = 0.5$ ($R = 0.7$). Other systematic uncertaintes are due to fake tracks (1.2% per track), to track p_T resolution (about 1% per track) and to calorimeters energy resolution.

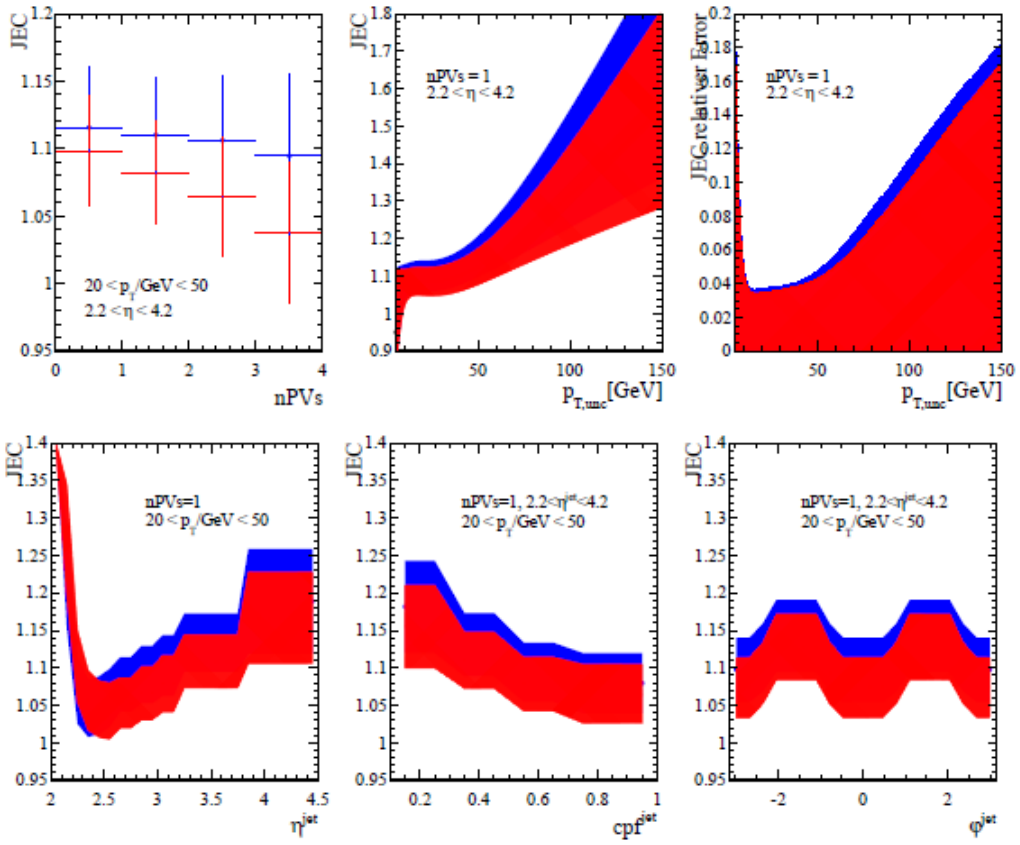

Figure 3.4: Correction factor k_{MC} as a function of different jet variables: nPV, p_T, η, ϕ and cpf for both $R = 0.5$ (blue) and $R = 0.7$ (red).

An additional multiplicative factor is computed in order to take in account possible differences between jet energies in real data and MC samples: this factor is called *Jet Energy Scale*.

3.3.5 Jet identification efficiencies

The jet identification efficiency is defined as the number of reconstructed jets over the number of true jets:

$$\varepsilon_{jets} = \frac{n_{reco-jets}}{n_{true-jets}} \tag{3.6}$$

The following requirements are applied to reject jets coming from background noise and high energetic isolated leptons:

- $nPVtrks \geq 2$, where $nPVtrks$ is the number of tracks pointing to the PV;
- $mpf < 0.8$, where mpf is the maximum fraction of transverse momentum carried by a single Particle Flow particle;
- $mpt > 1.2$ GeV, where mpt is the maximum transverse momentum carried by a track;
- $cpf > 0.1$, with cpf the fraction of charged particles inside the jet.

A MC sample of $Z \rightarrow \mu\mu$+jet at $\sqrt{s} = 7$ TeV has been used to measure ε_{jet} [51]. In Fig.3.5 jet identification efficiency ε_{jet} is plotted as a function of jet p_T.

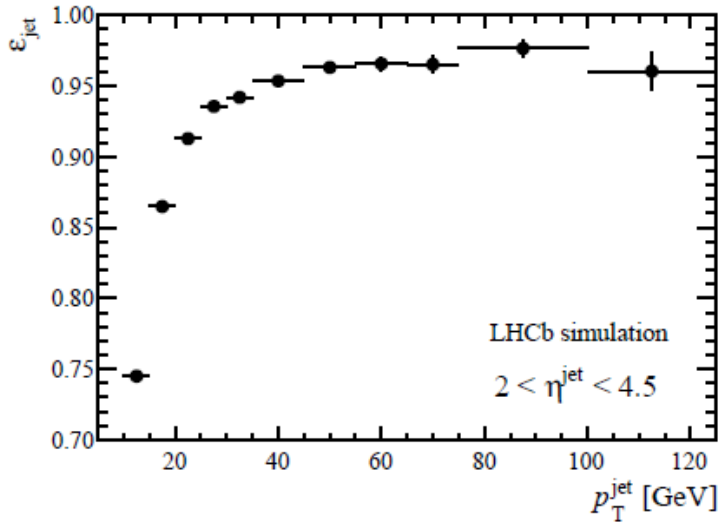

Figure 3.5: Jet identification efficiency as a function of jet p_T.

It is evident that the identification efficiency is increasing with respect to increasing jet p_T, particularly good results are obtained for $p_T^{jet} > 20$ GeV.

3.4 Flavour tagging algorithm

The *flavour tagging algorithm* is the method used to identify the flavour of the heavy quark (b,c) that originates a jet and it is based on the analysis of jet properties. In this thesis *Secondary Vertex* (SV) tagging algorithm is used, and in the following sections it is described and compared to other algorithms.

3.4.1 SV tagging algorithm

The SV tagging algorithm [52] can be summed up in this way: since b and c hadrons have a certain lifetime τ they travel for a distance d inside the detector before decaying. The average distance is computed as $< d >= \gamma c\tau\beta$ where c is the speed of light, $\beta = \frac{v}{c}$, v is the particle velocity and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ is the Lorentz factor. In this way a SV is present, and it is generated by the decay of hadrons detached from the PV where the main interaction occurs. If a SV is found inside a jet then it can be tagged as generated from a heavy flavour quark, separating light quarks from heavy quarks and b-jets from

c-jets.
The SV tagging algorithm procedes through several steps, as described next:

1. the following requirements on selected tracks are considered:
   - *long* tracks;
   - $p > 5$ GeV/c and $p_T > 0.5$ GeV/c;
   - the χ^2/ndof < 3, where χ^2ndof is associated to the track fit;
   - $\chi^2_{IP} > 16$, where χ^2_{IP} is the variation of the χ^2 for the PV fit after removing a track from the fit result;
   - $P_{ghost} < 0.3$.
2. Using selected tracks all possible 2−body SVs in the 3−dimensional space are built, in order to determine the SV position. If two tracks are associated to the same SV then they are combined to form one particle whose flight direction is parallel to the vector pointing to the SV from the PV, with 4−momentum defined as the sum of tracks 4−momenta, assuming the π mass for the particle.
3. The 2−body particles need to fulfill some requirements;
   - $d_{DOCA} < 0.2$ mm, where d_{DOCA} is the distance of closest approach (DOCA) between two tracks;
   - the $\chi^2_{SV} < 10$, with χ^2_{SV} associated to the SV fit;
   - the invariant mass must be between 400 MeV/c^2 and 5279.4 MeV/c^2 which is the mass of the B^0 meson, such that strange-hadrons decays are removed but B vertices are still considered (thanks to the assigning the π mass).
4. In order to consider a 2−body particle inside a jet $\Delta R < 0.5$ is required, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, with $\Delta\eta = \eta_{jet} - \eta_{SV}$ and $\Delta\phi = \phi_{jet} - \phi_{SV}$. All 2−body particles in the same jet that share at least one track are associated to each other, thus creating n−body particles that do not have tracks in common. The resulting n−body particle is called tag_{SV}, and its position is averaged from the 2−body vertex fit. Its flight direction is parallel to the vector that points from the PV to its position, while its 4−momentum is the sum of all the tracks 4−momenta, again assuming the π mass.
5. To remove light jets contamination, the following criteria are applied to tag_{SV}:
   - $p_T > 20$ GeV;
   - z−position of the jet less than 200 mm;
   - $\frac{d}{p} < 1.5$ mm/(GeV/c) where d is the flight distance;
   - $\chi^2_d > 5\sigma$ where χ^2_d is the flight distance χ^2 obtained from the PV fit by adding the tag_{SV} tracks;
   - a tag_{SV} is rejected if it has only two tracks;
   - the tag_{SV} must have at most one track with $\Delta R > 0.5$.
6. The corrected mass of the tag_{SV} is defined as:

$$m_{cor} = \sqrt{m^2 + p^2 \sin^2\theta} + p \sin\theta \tag{3.7}$$

   where m and p are respectively the invariant mass and the momentum of tag_{SV}, while θ is the angle between the tag_{SV} momentum and the flight direction.

7. If more than one tag_{SV} inside a jet satisfies the requirements listed above, then the one with greater p_T is chosen. In order to consider a jet as generated by a heavy flavour quark at least one tag_{SV} has to be found with this procedure.

8. As a last point of this algorithm, it is necessary to separate heavy quarks from light ones and b quarks from c quarks. A Multivariate Algorithm (MVA) is used by employing two Boosted Decision Trees (BDTs) [53–55]: one for the heavy/light quarks separation (called $\mathrm{BDT}_{bc|udsg}$ or BDT_0) and the other one for b/c quarks separation (called $\mathrm{BDT}_{b|c}$ or BDT_1). The variables used as inputs for the BDTs are those related to tag_{SV}:

   - the tag_{SV} mass m;
   - the tag_{SV} corrected mass m_{cor};
   - the transverse flight distance d_T of the tag_{SV} 2−body particle closest to the PV;
   - the fraction of the jet p_T carried by tag_{SV}, $\frac{p_{T,SV}}{p_{T,jet}}$;
   - the number of tracks n_{tracks} that form tag_{SV};
   - the number of tracks $n_{tracks,\Delta R}$ inside tag_{SV} with $\Delta R < 0.5$;
   - the total charge Q of the tracks in tag_{SV};
   - the χ^2_d of tag_{SV};
   - the sum of χ^2_{IP} for all the tracks in tag_{SV}.

MC samples considering b, c and light quarks coming from proton-proton collisions at 8 TeV are generated with PYTHIA 8 [44] and used to train the BDTs, simulating both heavy/light quarks and b/c quarks as signal/background samples. In order to select jets samples with a particular flavour it is possible to apply cuts on the two BDTs. In Fig.3.6 a 2-dimensional graph for $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ is shown, presenting different distributions for b (red), c (green) and q (blue) jets.

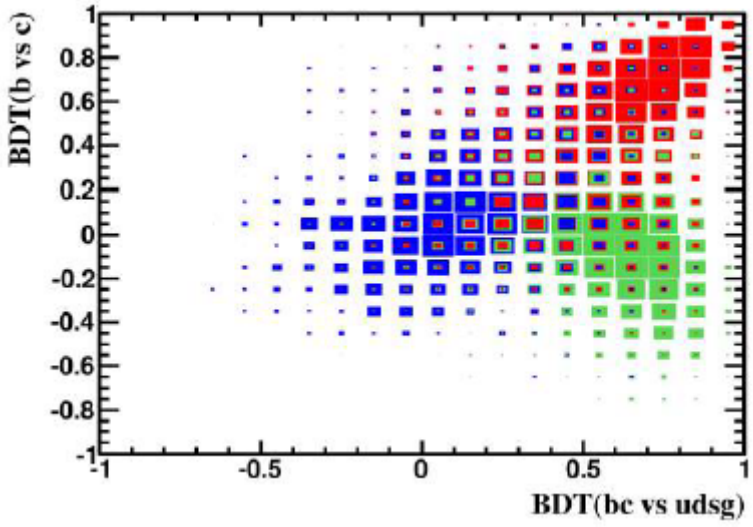

Figure 3.6: 2-dimensional graph for both BDTs showing the different distribution of b (red), c (green) and light quarks (blue) [52].

3.4.2 Tagging perfomance

The tagging algorithm performances are evaluated through two parameters:

- the $b(c)$ tagging efficiency $\varepsilon_{b(c)}$ defined as the number of tagged $b(c)$ jets $N^{tag}_{b(c)}$ over the total

number of reconstructed $b(c)$ jets $N^{tot}_{b(c)}$:

$$\varepsilon_{b(c)} = \frac{N^{tag}_{b(c)}}{N^{tot}_{b(c)}} \tag{3.8}$$

- the light jets misidentification ε_q defined as the ratio between the number of tagged light jets N^{tag}_q (generated by $q = u$, d, s and g) to the total number of reconstructed light jets N^{tot}_q:

$$\varepsilon_q = \frac{N^{tag}_q}{N^{tot}_q} \tag{3.9}$$

The SV tagging performances have been evaluated on MC samples for $pp \to$ di-jets processes at $\sqrt{s} = 13$ TeV generating b, c and q jets simulated with PYTHIA 8 [44]. Moreover both $\varepsilon_{b(c)}$ and ε_q have been computed as functions of p_T and pseudorapidity (η) of the reconstructed jet. The following results are obtained:

- for $p_T > 20$ GeV/c the average SV-tagging efficiency $\varepsilon_{b(c)}$ is around 60% (25%) for $b(c)$ jets;
- for $p_T < 20$ GeV/c the tagging efficiencies are significantly lower, about 10% (15%) for $b(c)$ jets;
- the SV-tagging mis-identification efficiency ε_q is about 0.3%, increasing with the increasing p_T;
- for $2.2 < \eta < 4.2$ $\varepsilon_{b(c)}$ have the same behaviour (and almost same values) as for varying p_T while ε_q is almost constant.

In Fig.3.7 SV-tagging efficiencies and mis-identification efficiency are shown as functions of reconstructed jet p_T (left column) and η (right column), compared to other jet tagging algorithms used at LHCb: the *topological* algorithm (TOPO) which is based on online selection of events at HLT2 trigger level, and its *loose* version, where a muon track is found and thus looser BDT requirements can be applied.

The jet tagging algorithm performance should also be computed for data in order to find discrepancies with MC samples; in order to do so particular data events (called *event-tag*) that have a fully reconstructed b or c jet or a high p_T muon are selected. Requirements are applied to a *test jet* associated to the same PV: $|\Delta\phi| > 2.5$ so that there is no contamination from the event-tag and $p_T > 100$ GeV, since it is not possible to measure efficiency for greater p_T. The general procedure is the following:

- templates of $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ distributions are obtained from MC samples, for b, c and q jets;
- a combined fit of both BDTs distributions of data is performed. In the fit the yields of tagged b, c and q jets (respectively n^{tag}_b, n^{tag}_c and n^{tag}_q) are measured;
- given $\chi^2_{IP(max-p_t)}$ as the χ^2_{IP} of the track with the highest p_T in the jet, templates of $\chi^2_{IP(max-p_t)}$ for b, c and q jets are built using MC samples;
- a combined fit of the $\chi^2_{IP(max-p_t)}$ of data events is performed. In the fit the yields of b, c and q jets (respectively n^{tot}_b, n^{tot}_c and n^{tot}_q) with no SV-tagging are measured;
- finally the $b(c)$ tagging efficiencies and the light jets mis-identification are computed:

$$\epsilon_{b(c)} = \frac{n^{tag}_{b(c)}}{n^{tot}_{b(c)}} \qquad \epsilon_q = \frac{n^{tag}_q}{n^{tot}_q} \tag{3.10}$$

Different event-tag process have been used to compute $b(c)$ tagging efficiencies and the q jets mis-identification: in the former the processes considered are B+jet, D+jet and μ+jet while in the latter the process considered is W+jet. In Fig.3.8 ratios between efficiency computed in data and in MC samples are presented, for $b(c)$ jets (left) and for q jets (right); the error bars take into account both

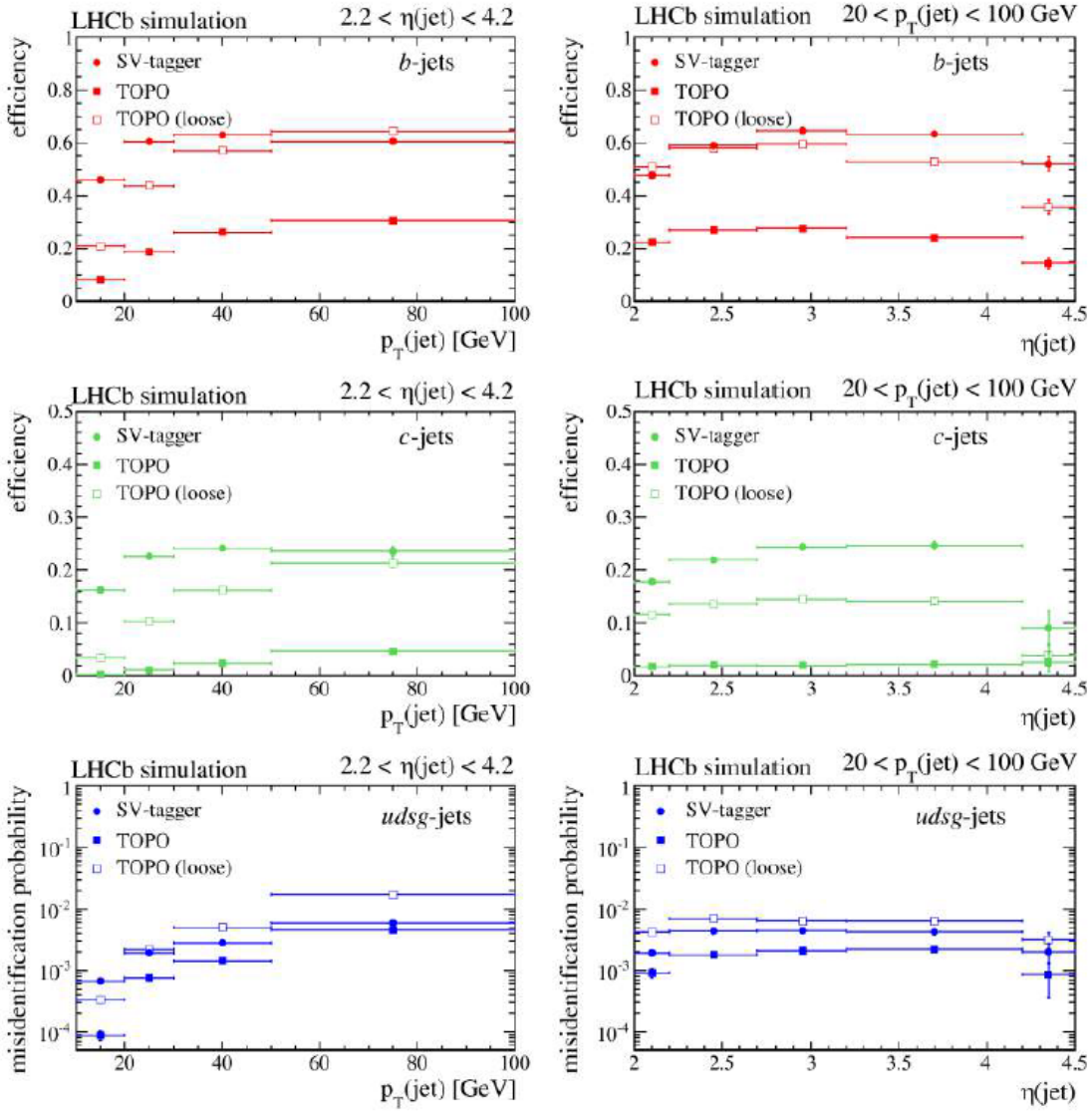

Figure 3.7: Comparison between different jet tagging efficiencies (different markers) for b, c and q jets (respectively in red, green and blue) and as functions of p_T (left column) and η (right column) [52].

statistical and systematic uncertainties, the latter coming from the mis-modelling of $\chi^2_{IP(max-p_t)}$ and to the gluon splitting creating $b\bar{b}$ and $c\bar{c}$ couples that can be tagged. On an average results coming from data are consistent with MC simulations.

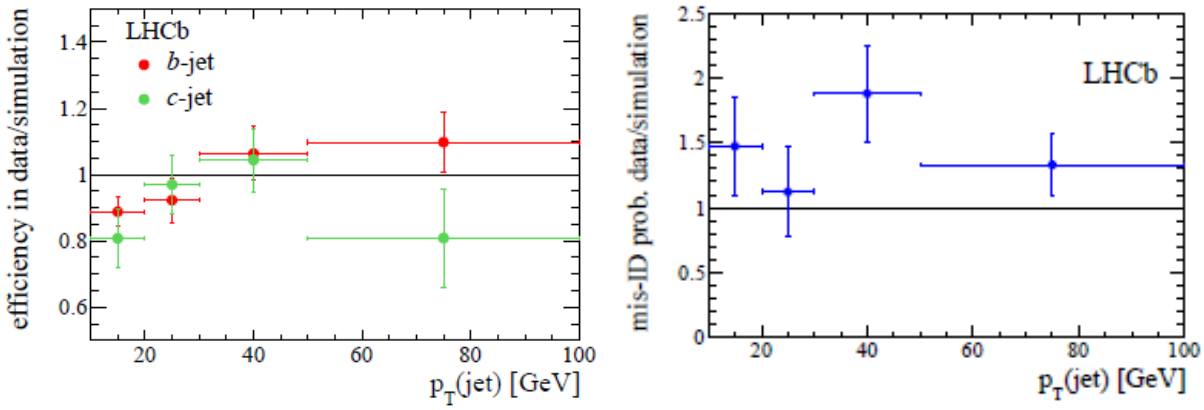

Figure 3.8: Ratio between tagging efficiencies in data and MC for b and c quarks (left), and between mis-identification efficiencies in data and MC for light quarks (right) as functions of jet p_T [52].

Previous results on tagging performances have been obtained during Run I campaign, but compatible results have been found also during Run II campaign at 13 TeV. In Fig.3.9 plots for $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ for QCD MC samples simulated for Run II are shown for $b\bar{b}$ (red line), $c\bar{c}$ (green) and $q\bar{q}$ (blue) di-jets candidates. It is evident that by choosing cuts on the BDTs it is possible to select a specific flavour for di-jets candidates.

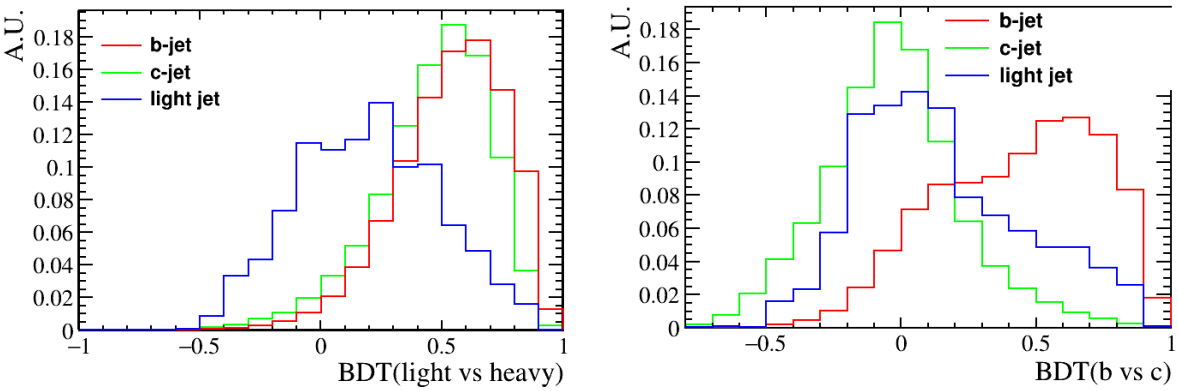

Figure 3.9: $\mathrm{BDT}_{bc|udsg}$ (left) and $\mathrm{BDT}_{b|c}$ (right) distributions for di-jets candidates simulated in QCD MC samples.

3.5 Data event selection

Data taken at LHCb during Run II in 2016 are used. In 2016 a total integrated luminosity of (1.67 ± 0.06) fb^{-1} has been collected, corresponding to a CoM energy $\sqrt{s} = 13$ TeV. Only a small fraction of this data has been used to perform the analysis, around (200 ± 8) pb^{-1}. In order to properly select events necessary for analysis some requirements are considered (the speed of light is considered $c = 1$):

- requirements on the trigger, in particular:
  - for the *L0* hardware trigger and for HLT1 the requirement is that at least one line is considered;
  - for the HLT2 two jets with $p_T > 17$ GeV are required to have a SV each;
- kinematical cuts, such as:

  - the jet p_T is greater than 20 GeV since there is the best p_T resolution and to avoid QCD background at low masses but still preserving most of the background;
  - $2.2 < \eta_{jet} < 4.2$ since there is an efficiency plateau and to ensure that jets are fully inside the instrumented LHCb acceptance;
  - $\Delta\phi > 1.5$, where $\Delta\phi$ is the angular distance between the two jets, in order to avoid that a jet is reconstructed and selected as two jets together;
- presence of a Secondary Vertex (SV) so that both jets have to be tagged by the SV-tagging algorithm, thus forming a *di-jets* candidate;
- if more than one di-jets candidate is found in the same event, thus having multiple candidates, the one with the highest jets p_T is selected, so that there is only one candidate per event; this procedure is called *multiple candidates removal*.

In Fig.3.10 invariant mass distribution for di-jets candidates and jet p_T distribution are shown.

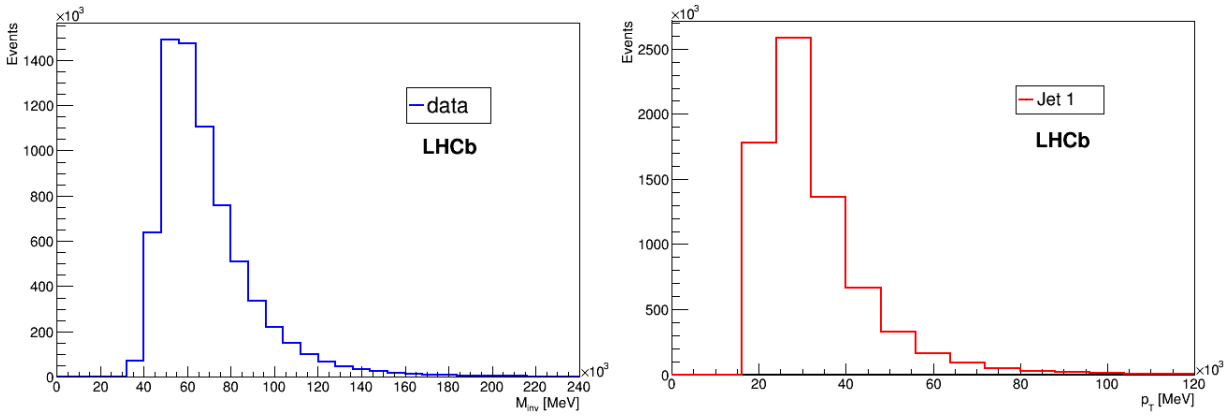

Figure 3.10: Invariant mass distribution for di-jets candidates (left) and jet p_T distribution.

Chapter 4

Background determination

4.1 Introduction

In this thesis $b-$ and $c-$di-jets resonances with low production cross section are searched and the background coming from QCD jets have to be precisely determined. MC simulation fails to reproduce background processes, therefore a data-driven approach has to be used. In data a so called *Control Region* (CR) is identified, where no signal is expected but the characteristics are similar to the *Signal Region* (SR), where a signal is expected. The following procedure is considered:

- SR and CR are defined by cuts on the SV-tagging BDTs introduced in section 3.4.1, requiring for the CR that one of the two jets is tagged as a non-b (non-c) quarks for $b\bar{b}$ ($c\bar{c}$) di-jets candidates;
- having defined SR and CR, a *Transfer Function* (TF) is obtained from the MC samples, so that the di-jets invariant mass distribution in the CR is used to describe the background in the SR without creating any bias;
- a fit to the di-jets invariant mass distribution in SR is performed. In this fit the parameters of the TF obtained in MC are left free to vary, in order to account for data/MC differences;
- the fitting procedure is validated by extracting the number of $Z \to b\bar{b}$ ($Z \to c\bar{c}$) events in the $b\bar{b}$ ($c\bar{c}$) sample and comparing it with the number of expected Z events.

4.2 QCD and Z MC samples

QCD MC samples are obtained from simulations: the production of $b\bar{b}$, $c\bar{c}$ and $q\bar{q}$ jets is simulated using Pythia8 [44] as generator and Geant4 [46, 47] to describe interactions with the detector. The two jets are required to be in the LHCb acceptance (which is $2 < \eta < 5$). As a second source of background the processes $Z \to b\bar{b}$ and $Z \to c\bar{c}$ are considered; again they are simulated with Pythia8 and Geant4, requiring the two jets coming from the Z decay in the LHCb acceptance. Other sources of background have not been considered, since at this value of luminosity they are negligible.

4.2.1 Event selection

Kinematical requirements applied to data are also applied to MC. They are recalled here for completeness:

- the jet p_T is greater than 20 GeV for both jets;
- $2.2 < \eta_{jet} < 4.2$ for both jets;
- $\Delta\phi > 1.5$, where $\Delta\phi$ is the angular distance between the two jets.
- presence of a Secondary Vertex (SV) in both jet cones;

- multiple candidates removal;

Once these requirements are applied, it is possible to compare MC samples with data. In Fig.4.1 normalized events for data and MC are shown: it is evident that MC sample can not reproduce data, thus requiring a new strategy to analyse data.

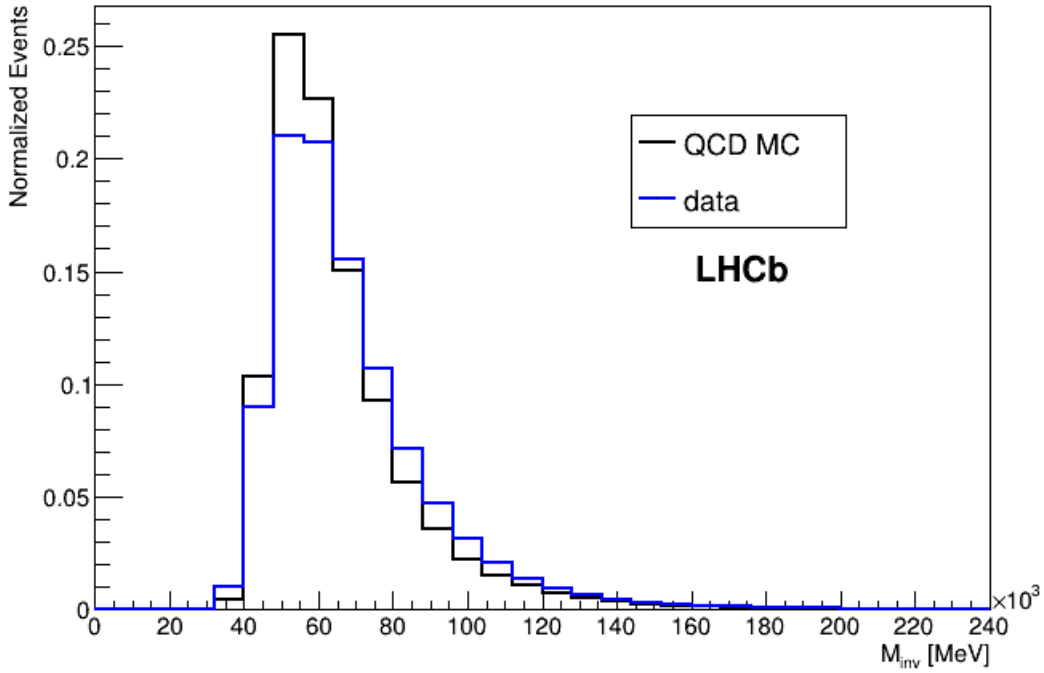

Figure 4.1: Invariant mass distribution for di-jets candidates of data (blue) and QCD MC samples (black). The two distributions are normalized to the same area (unity).

4.2.2 Yield prediction

Before dealing with the construction of the background model, it is important to know how many $Z \to b\bar{b}$ and $Z \to c\bar{c}$ events are expected in the selected data sample: this will be really important since it is the second source of background (after QCD) and in order to validate the fit procedure afterwards. The number of expected events N^{exp} is computed with the following formula:

$$N^{exp} = \mathcal{L} \cdot \sigma_{th} \cdot A \cdot \epsilon_{cut} \tag{4.1}$$

where:

- $\mathcal{L}$ is the integrated luminosity which is around 200 pb^{-1};
- σ_{th} is the theoretical cross section of the considered process, obtained by multiplying the Z production cross section for the branching ratio $\mathcal{BR}$ of the considered decay. For $Z \to b\bar{b}$ and $Z \to c\bar{c}$ the values for cross sections and relative uncertainties are presented in the following table, obtained with PYTHIA6 [43] in the MADGRAPH [63] framework:

Process	σ_{th} [pb]	Uncertainty
$Z \to b\bar{b}$	5307	3.5%
$Z \to c\bar{c}$	4154	3.9%

Table 4.1: Theoretical cross sections for Z boson decays.

- A is the LHCb acceptance factor computed with PYTHIA 6 [43] by requiring the decay products inside the LHCb acceptance ($2 < \eta < 5$): it depends on the considered process;

- ϵ_{cut} is the selection efficiency and it is defined as

$$\epsilon_{cut} = \frac{n_{MC,sel}}{n_{MC,gen}} \tag{4.2}$$

  where $n_{MC,sel}$ is the number of MC events selected after applying kinematical and SV-tagging cuts, while $n_{MC,gen}$ is the number of generated MC events, which is 200000.

In the following table the number of expected events in the data sample for each process and their relative selection efficiencies are presented:

Process	Acceptance	Selection efficiency	Expected events
$Z \to b\bar{b}$	0.161	16.13%	27564 ± 2067
$Z \to c\bar{c}$	0.213	0.94%	1663 ± 131

Table 4.2: Expected number of Z boson events in the selected data sample. Acceptance factor and selection efficiencies are also reported.

It is important to compute these yields since they will be compared with the number of observed Z decay events resulting from the fitting procedure.

4.3 Definition of Signal and Control Region

As seen in the previous section MC simulation can not describe well data, therefore we rely on data to model the QCD multi-jet background. Data are divided in two sub-samples: the *Signal Region* (SR) where signal is expected and the *Control Region* (CR) where there should be no signal and the events have the same kinematical characteristics of the SR. In order to select SR events cuts on SV-tagging BDTs variables are required: first of all a requirement on the $\text{BDT}_{bc|udsg}$ for both jets is applied, in order to select only heavy quarks; then a cut on $\text{BDT}_{b|c}$ for both jets is applied depending on the final state studied ($b\bar{b}$ or $c\bar{c}$). In this way two SRs are defined, one to search for $b\bar{b}$ resonances and one to search for $c\bar{c}$ resonances.

The CRs are defined as the regions where just one of the two jets passes the SR BDTs cuts, while the other jet does not fulfill the requirements: in this way one jet is tagged as an heavy quark with flavour i while the other one is tagged as a non-i quark, thus the possible final states in the CR are $b\bar{c}$, $c\bar{b}$, $q\bar{b}$, $b\bar{q}$, $q\bar{c}$ and $c\bar{q}$. The principle of working of this strategy is sketched in Fig.4.2 for the simpler case if only two flavours b (heavy) and q (light) are present.

The values of where to cut the SV-tagging BDTs in order to define the SR are chosen using a $H \to b\bar{b}$ simulated sample. The sample has been simulated using Pythia8 [44] and Geant4 [46,47] and all the kinematical and SV-tagging requirements are applied. The number of selected H events is afterwards compared with the number of data events that pass the same cuts, and the significance of the signal is computed in a defined mass range. The mass range chosen is $[100, 150]$ GeV, symmetric to the expected Higgs mass. The significance of the signal is defined as:

$$\mathcal{S} = \frac{N_{sig}}{\sqrt{N_{bckg} + N_{sig}}} \simeq \frac{N_{sig}}{\sqrt{N_{bckg}}} \tag{4.3}$$

where N_{sig} is the number of expected signal events (obtained applying the formula 4.1) while N_{bckg} is the number of selected data events. The assumption of no signal in data is good enough since very few events are expected. In Fig.4.3 the di-jets invariant mass distributions for simulated Higgs and data events normalized to 1 are shown.

Cuts on BDTs are applied to both jets and they are chosen such that $\mathcal{S}$ is maximized. This maximation procedure produces the cuts for di-jets candidates reported in Table 4.3. In the table cuts for selecting

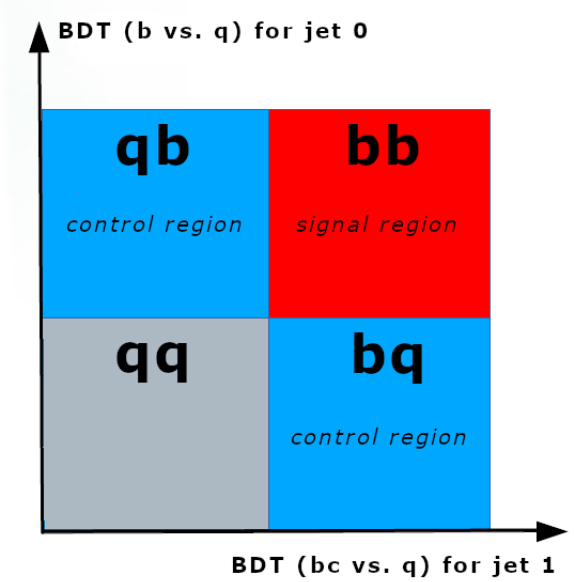

Figure 4.2: General working scheme for application of BDTs cuts in the simpler case where only two flavours (an heavy one b and a light one q) are present.

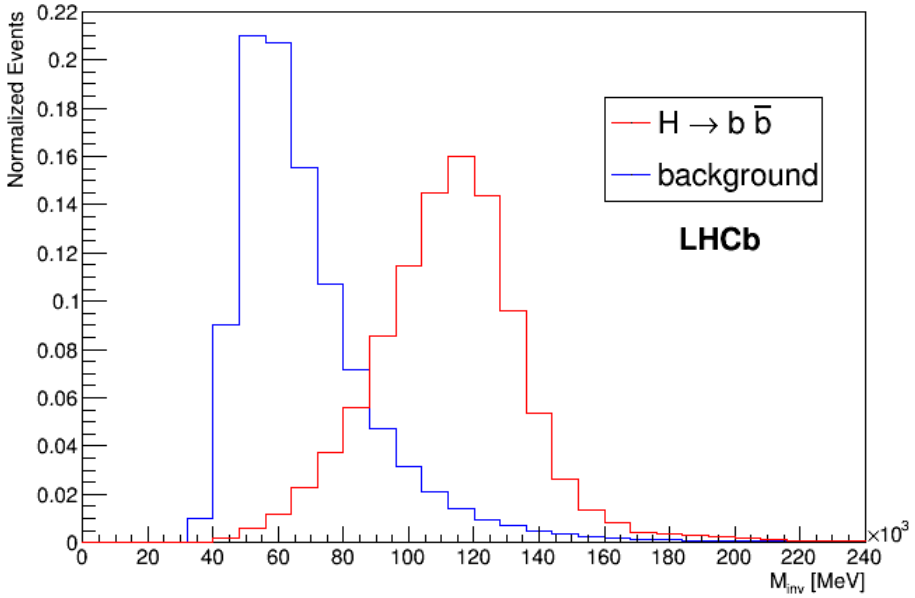

Figure 4.3: Dijet invariant mass distributions for data (blue) and simulated $H \to b\bar{b}$ (red) both normalized to 1, prior to applying BDTs cuts.

the CR are also reported, the symbol $not\{$ means that the complementary region to the applied cuts is considered.

events set	BDTs cuts applied on jet 0	BDTs cuts applied on jet 1
$b\bar{b}$ signal region	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$
$b\bar{b}$ control region	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$	$not\{ \mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$
	$not\{ \mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} > -0.03$
$c\bar{c}$ signal region	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$
$c\bar{c}$ control region	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$	$not\{ \mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$
	$not\{ \mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$	$\mathrm{BDT}_{bc\|udsg} > 0.24$ $\mathrm{BDT}_{b\|c} < -0.03$

Table 4.3: Cuts on BDTs

Once the cuts on BDTs are computed, they are applied to di-jets candidates on data so to select SR and CR both for $b\bar{b}$ and $c\bar{c}$ di-jets candidates. For $b\bar{b}$ di-jets candidates results obtained by applying BDTs cuts to data are shown in Table4.4.

set	data events
$b\bar{b}$ signal region	6275092
$b\bar{b}$ control region	769231
total	7103323

Table 4.4: Number of data events divided in two regions SR and CR for $b\bar{b}$ di-jets candidates (see text for comment on the total number of events).

From Table 4.4 it is evident that most of data (around 88%) are tagged as $b\bar{b}$ di-jets candidates. The sum of the yields of the two regions does not match the total number of events: this is due to the fact that in the CR no resonances are allowed, thus all $q\bar{q}$ di-jets candidates (with $q \neq b$) are not selected. In Fig.4.4 the di-jets invariant mass distribution for data events in SR, CR and prior to BDTs cuts for $b\bar{b}$ di-jets candidates are shown.

For $c\bar{c}$ di-jets candidates the situation is complementary: a small SR is now present overcomed by a CR almost 20 times bigger, on the same order as the CR for $b\bar{b}$ di-jets candidates. The result is consistent since the c-jets SV-tagging efficiency is smaller with respect to the b-jets, therefore most of the SV-tagged data sample is composed by $b\bar{b}$ events, that are removed by the BDT cuts for the $c\bar{c}$ SR. In Table 4.5 yields for SR and CR for $c\bar{c}$ di-jets candidates are given by applying BDTs cuts to data:

In Fig.4.5 the di-jets invariant mass distribution for data events in SR, CR and prior to BDTs cuts for $c\bar{c}$ di-jets candidates are shown: the SR is really small, as seen in the previous table.

Since the Z resonance is considered as a background source the same BDTs cuts are also applied to the $Z \to b\bar{b}$ and $Z \to c\bar{c}$ resonances. In this way it is possible to predict the proper number of Z events expected in the SR, to be compared to results from the fitting procedure. The number of selected Z MC events in SR and CR are reported in table 4.6.

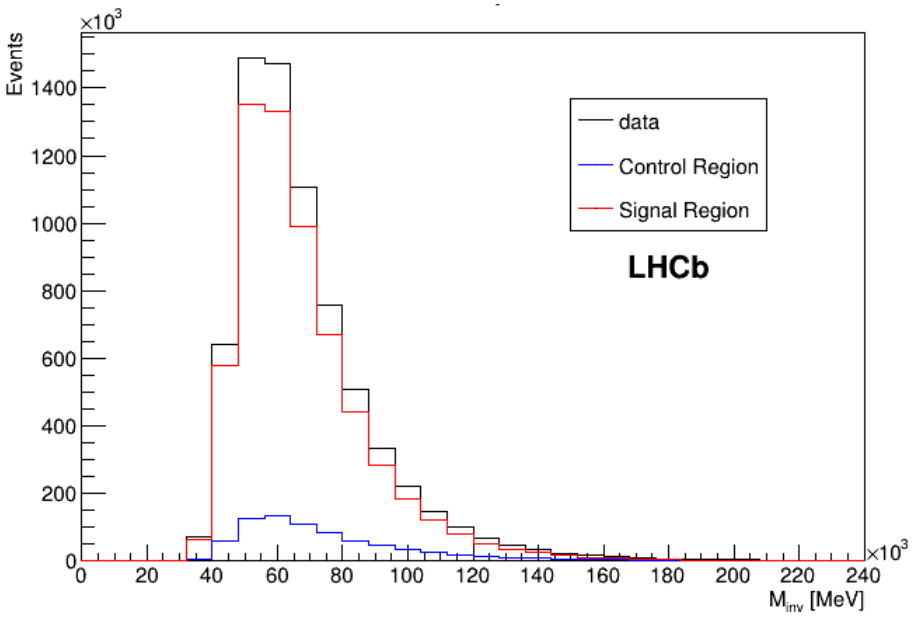

Figure 4.4: Dijet invariant mass distribution of data events divided in SR and CR applying $b\bar{b}$ cuts on BDTs. The label data indicates the selected events prior to BDT cuts.

set	data events
$c\bar{c}$ signal region	24823
$c\bar{c}$ control region	409399
total	7103323

Table 4.5: Number of data events in SR and CR with respect to all data for $c\bar{c}$ di-jets candidates.

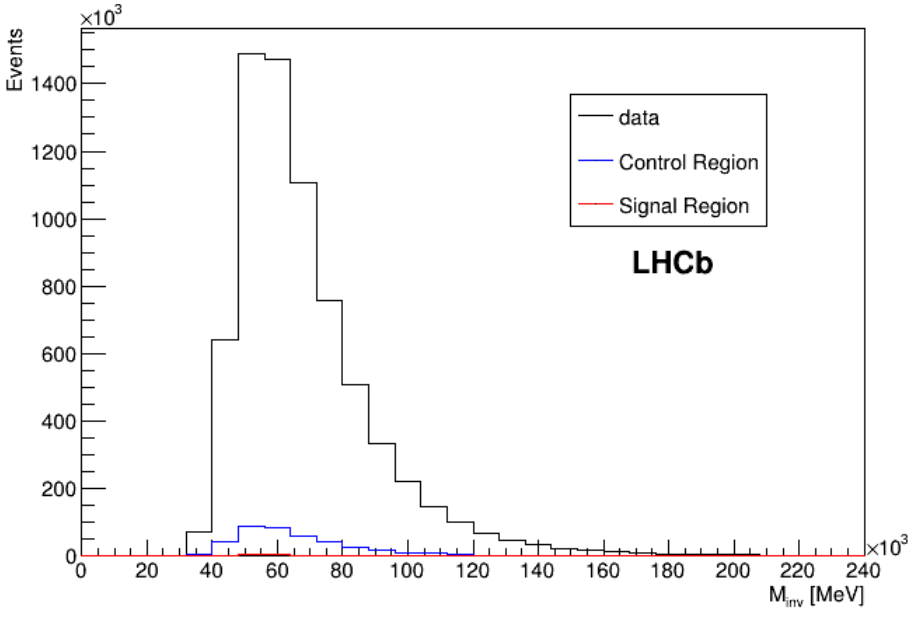

Figure 4.5: Dijet invariant mass distribution of data events divided in SR and CR applying $c\bar{c}$ cuts on BDTs. The label data indicates the selected events prior to BDT cuts.

process	set	MC events
$Z \to b\bar{b}$	$b\bar{b}$ signal region	30250
	$b\bar{b}$ control region	1967
	total MC	32253
$Z \to c\bar{c}$	$c\bar{c}$ signal region	69
	$c\bar{c}$ control region	325
	MC	821

Table 4.6: Number of SR and CR events for Z boson decay in MC samples.

Table 4.6 shows that for $Z \to b\bar{b}$ almost all the events (around 93.8%) are considered in the SR, while for $Z \to c\bar{c}$ around 8.4% are events in the SR. In this way it is possible to compute how many Z events are expected in SR and CR for data, as reported in Table 4.7. Notice that even if the expected number of $Z \to c\bar{c}$ events in the CR is higher with respect to the SR, the ratio of the expected $Z \to c\bar{c}$ events to the number of selected data events is greater in the SR with respect to CR.

process	set	expected events
$Z \to b\bar{b}$	$b\bar{b}$ signal region	25855
	$b\bar{b}$ control region	1681
	total	27564
$Z \to c\bar{c}$	$c\bar{c}$ signal region	140
	$c\bar{c}$ control region	658
	total	1663

Table 4.7: Expected SR and CR events for Z boson decay in data.

In Fig.4.6 invariant mass for MC events divided in SR and CR for $Z \to b\bar{b}$ and $Z \to c\bar{c}$ are shown.

4.3.1 Definition of the Transfer Function

The Transfer Function is used to evaluate the di-jets invariant mass distribution of the QCD multi-jet background in the SR starting from the CR. It is not possible to directly use the CR to describe the SR, since the insertion of cuts on the BDTs defines different shapes to the sub-samples. In searching for signal resonances (see Chapter 5) it is important for the TF to not have any bias by artificially inserting a resonance. In order to do so the TF is computed starting from the QCD MC samples. The TF is defined as the ratio between the number of events in the SR over the number of events in the CR per bin of invariant mass:

$$\mathrm{TF} = \frac{n_{events,SR}}{n_{events,CR}} \tag{4.4}$$

The TF is fitted in MC with a polynomial curve. In the case of $b-$di-jets the $\mathrm{TF}_{MC,b}$ is parametrized as a polynomial curve of seven-grade

$$\mathrm{TF}_{MC,b}(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 + a_4 \cdot x^4 + a_5 \cdot x^5 + a_6 \cdot x^6 + a_7 \cdot x^7 \tag{4.5}$$

where x is the invariant mass of the di-jets candidate. In the Table 4.8 a list of all the parameters is shown while in Fig.4.7 the fitted function $\mathrm{TF}_{MC,b}$ is presented. The fit has been performed in the invariant mass range $[35; 240]$ GeV.

It is possible to check if the TF_{MC} obtained from the QCD MC samples is good enough to describe data; in order to do so it is possible to compare $\mathrm{TF}_{MC,b}$ with the TF obtained from data $\mathrm{TF}_{data,b}$, as

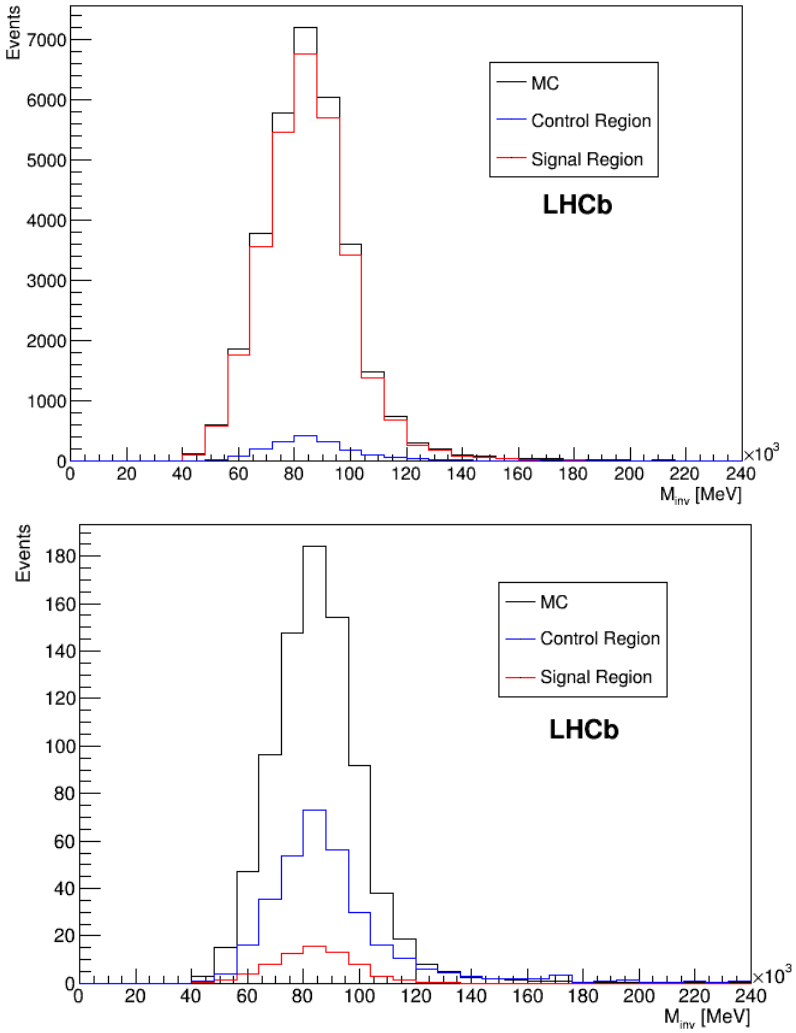

Figure 4.6: MC samples for $Z \to b\bar{b}$ (up) and $Z \to c\bar{c}$ (down) resonances with SR and CR.

name	value
a_0	(-75 ± 47)
a_1	$(5 \pm 3) \times 10^{-3}$ MeV^{-1}
a_2	$(-1.2 \pm 0.8) \times 10^{-7}$ MeV^{-2}
a_3	$(1.5 \pm 1.1) \times 10^{-12}$ MeV^{-3}
a_4	$(-1.1 \pm 0.9) \times 10^{-17}$ MeV^{-4}
a_5	$(4.6 \pm 4.3) \times 10^{-23}$ MeV^{-5}
a_6	$(-1.1 \pm 1) \times 10^{-28}$ MeV^{-6}
a_7	$(1.1 \pm 1) \times 10^{-34}$ MeV^{-7}

Table 4.8: Parameters for $\mathrm{TF}_{MC,b}$ after fitting the TF for QCD MC samples.

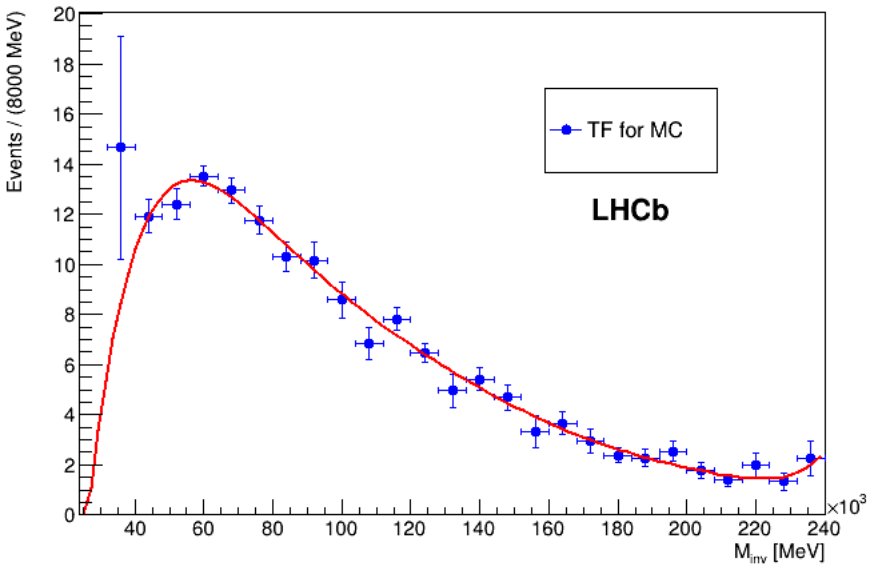

Figure 4.7: Fitting function for $\mathrm{TF}_{MC,b}$.

shown in Fig.4.8, where it is clear that the two TFs for data and MC for $b\bar{b}$ di-jets candidates are not the same, but they have a similar shape. In this check it is assumed that resonances are not present in data and that $\mathrm{TF}_{data,b}$ is not modified by them.

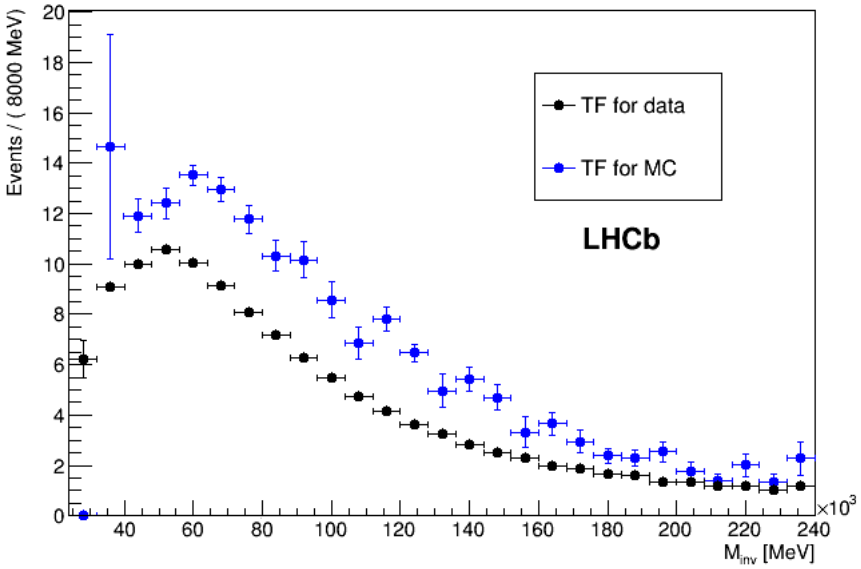

Figure 4.8: TFs for data (black dots) and MC (blue dots) for $b\bar{b}$ di-jets candidates.

In order to study the differences between the two shapes it is possible to compute the ratio ρ_b between the two TFs, by dividing their values per single bin of invariant mass:

$$\rho_b = \frac{\mathrm{TF}_{data,b}}{\mathrm{TF}_{MC,b}} \tag{4.6}$$

Then it is possible to fit this function with a function $h(x)$ defined as

$$h(x) = p_b + q_b e^{-r_b x} \tag{4.7}$$

where x is the invariant mass and p_b, q_b and r_b are real positive constants.As shown in Fig.4.9 ρ_b can be fitted by $h(x)$ in the mass range $[35; 240]$ GeV, obtaining the following values for the parameters:

$$\begin{aligned} p_b &= 0.56 \pm 0.02 \\ q_b &= 1.3 \pm 0.05 \\ r_b &= (3.2 \pm 0.8) \times 10^{-5}\ \text{MeV}^{-1} \end{aligned} \tag{4.8}$$

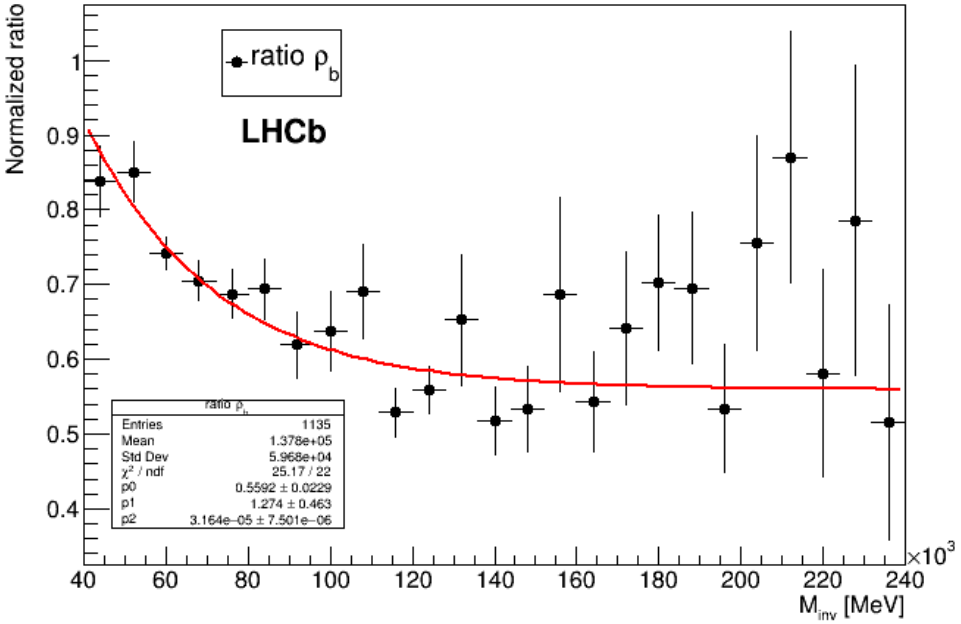

Figure 4.9: Ratio ρ_b between MC and data TFs and fitting function $h(x)$.

In this way it is possible to check that the two TFs are the same for high invariant masses (where $e^{-r_b x} \to 0$), while at low masses there is a correction coming from the exponential term. *A posteriori* it is also evident that no resonant structures are present.

The same procedure is used for the TF in the case of search for $c\bar{c}$ di-jets candidates: a nine-grade polynomial curve $\text{TF}_{MC,c}$ is chosen to fit TF from QCD MC samples

$$\text{TF}_{MC,c}(x) = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + b_4 \cdot x^4 + b_5 \cdot x^5 + b_6 \cdot x^6 + b_7 \cdot x^7 + b_8 \cdot x^8 + b_9 \cdot x^9 \tag{4.9}$$

where x is the invariant mass of the di-jets candidate. In Table 4.9 parameters for $\text{TF}_{MC,c}$ are presented, while in Fig.4.10 the fitting function $\text{TF}_{MC,c}$ is presented. The fit has been performed in the invariant mass range $[40; 220]$ GeV.

It is also possible to check if $\text{TF}_{MC,c}$ describes well the TF obtained from data $\text{TF}_{data,c}$, as shown in Fig.4.11: it is clear that the two TFs are different but have similar shapes.

Finally as done for $b\bar{b}$ di-jets candidates the ratio ρ_c is computed in order to study the differences between the two shapes, by fitting ρ_c with a function $g(x)$ defined as

$$g(x) = p_c + q_c e^{-r_c x} \tag{4.10}$$

where x is the invariant mass and p_c, q_c and r_c are real positive constants. Via a fit procedure in the mass range $[40; 220]$ GeV the following values for $g(x)$ parameters are obtained

parameter	value
b_0	(4.029 ± 0.008)
b_1	$(-3.977 \pm 0.001) \times 10^{-4}$ MeV^{-1}
b_2	$(1.6597 \pm 0.0001) \times 10^{-8}$ MeV^{-2}
b_3	$(-3.80804 \pm 0.00006) \times 10^{-13}$ MeV^{-3}
b_4	$(5.32051 \pm 0.00003) \times 10^{-18}$ MeV^{-4}
b_5	$(-4.71524 \pm 0.00002) \times 10^{-23}$ MeV^{-5}
b_6	$(2.66187 \pm 0.00001) \times 10^{-28}$ MeV^{-6}
b_7	$(-9.26845 \pm 0.00004) \times 10^{-34}$ MeV^{-7}
b_8	$(1.81325 \pm 0.00002) \times 10^{-39}$ MeV^{-8}
b_9	$(-1.52392 \pm 0.00008) \times 10^{-45}$ MeV^{-9}

Table 4.9: Parameters of the TF$_{MC,c}$ for MC sample.

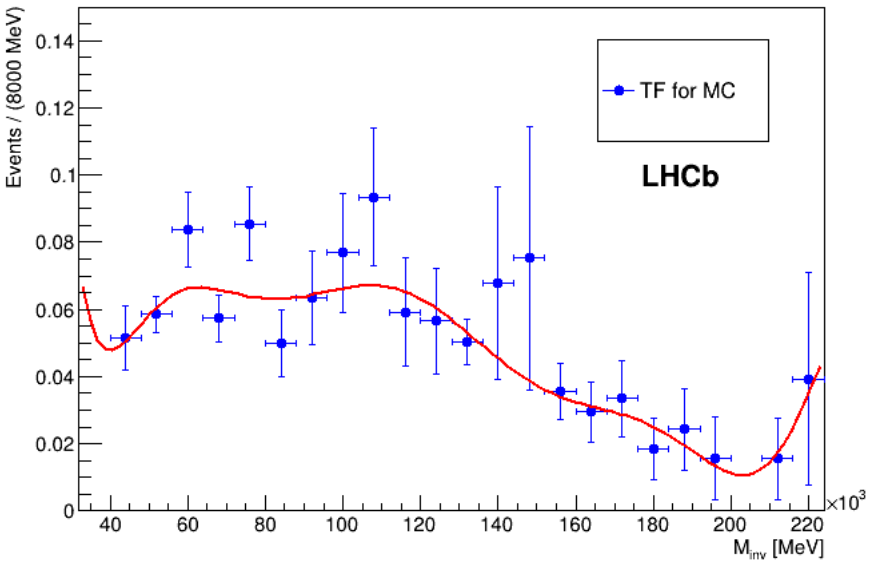

Figure 4.10: Fitting function for TF$_{MC,c}$.

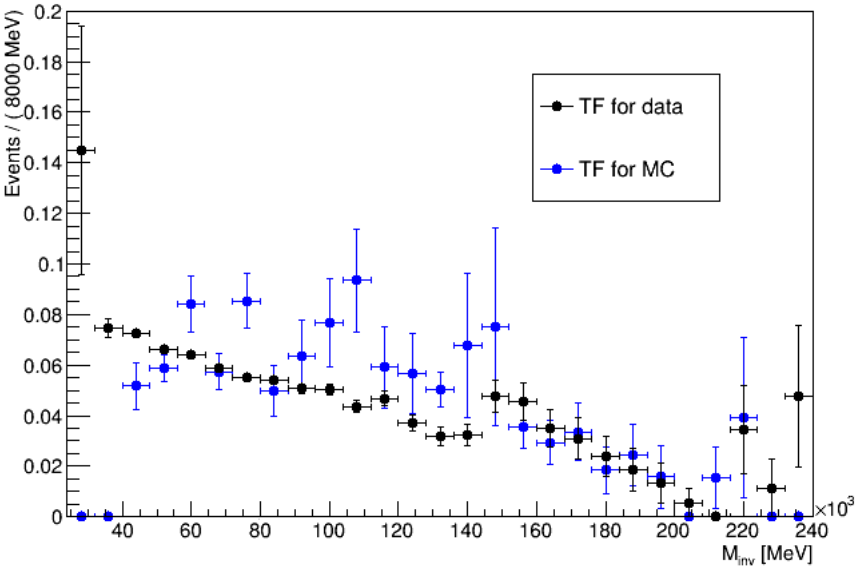

Figure 4.11: TFs for data (black dots) and MC (blue dots) for $c\bar{c}$ di-jets candidates.

$$\begin{aligned} p_c &= 0.61 \pm 0.06 \\ q_c &= 5 \pm 6 \\ r_c &= (4.7 \pm 0.4) \times 10^{-5}\ \text{MeV}^{-1} \end{aligned} \tag{4.11}$$

and the fitting plot is reported in Fig.4.12.

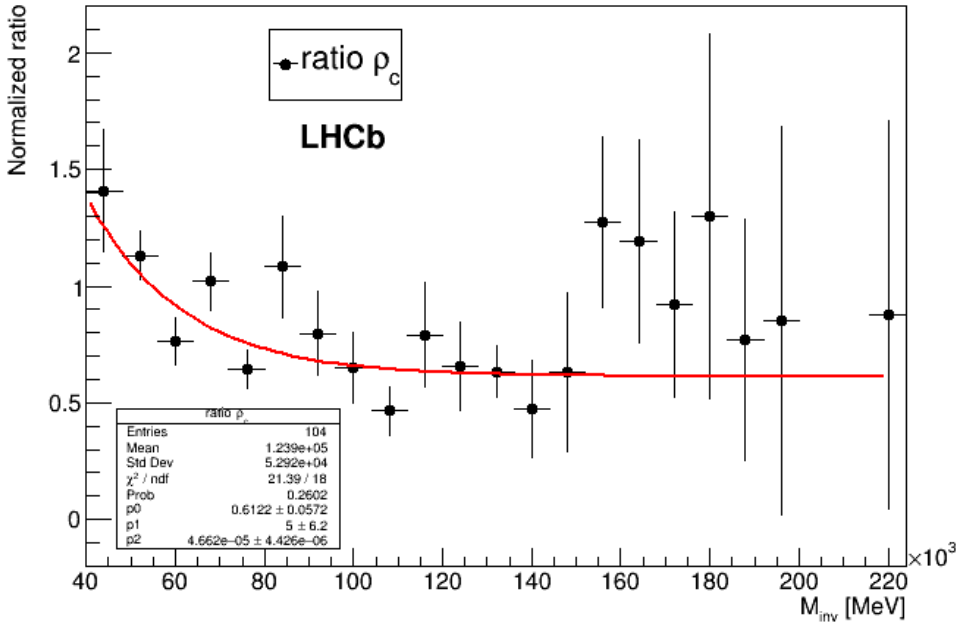

Figure 4.12: Ratio ρ_c between MC and data TFs and fitting function$g(x)$.

It is clear that as for $b\bar{b}$ di-jets candidates for $c\bar{c}$ di-jets candidates at high values for invariant mass the two shapes are the same, while at low masses there is a correction coming from the exponential term. Again it is evident that no resonant structures are present.

4.4 Fit to the Signal Region

A fit to the di-jets invariant mass distribution in the SR is performed to validate the background model. A binned maximum likelihood approach is followed, therefore distributions in the fit model are obtained as histograms.

Since for both $b\bar{b}$ and $c\bar{c}$ di-jets candidates the TF_{MC} is not the same as TF_{data}, it is necessary to take account of data/MC differences. For this reason in the fit to the SR, the same parametrization of TF_{MC} is used for TF_{data}, but the parameters for TF_{data} are left free to vary. The TF_{MC} parameters obtained in MC are used as starting points in the fit and they can vary with a certain range centered around them.

Once the fitting functions for $\text{TF}_{data,b}$ and $\text{TF}_{data,c}$ are obtained as explained in the previous Section, it is possible to build the background model, which basically consists in two contributions: QCD processes and the decay of the Z boson into b or c quarks. Other sources of background are considered negligible. To build the QCD background B_{QCD} it is sufficient to multiply the CR invariant mass distribution for the TF_{data}: in this way a model for QCD processes in the SR is created, without considering any resonance. Then the Z resonance can be added, the di-jets invariant mass shape is taken from simulation. In order to check the validity of the TF procedure it is possible to fit the model to data leaving free the number of QCD and Z events, and compare the number of fitted Z events with the theoretical expectations. Therefore the model used in the fit is the following:

$$f_i(x) = N^{obs}_{QCD,i} \cdot B_{QCD,i}(x) + N^{obs}_{Z,i} \cdot B_{Z,i}(x) \tag{4.12}$$

where x is the invariant mass, $B_{QCD,i}(x)$ is the QCD background model obtained as the product of the function fitting the TF_{data} times the CR for $i = b, c$ di-jets candidates, and $B_{Z,i}(x)$ is the Z invariant mass distribution for $i = b, c$ quarks in the final state, obtained from MC. $N^{obs}_{QCD,i}$ and $N^{obs}_{Z,i}$ are respectively the observed number of QCD and Z events for $i = b, c$ di-jets candidates. The fit is performed with the binned maximum likelihood technique using the *Roofit* [56] package of the software ROOT [41]. The fitting algorithm used is *Minuit2* [57].

In Tables 4.10 and 4.11 the fitted parameters for $\mathrm{TF}_{data,b}$ and $\mathrm{TF}_{data,c}$ are shown.

name	value
$\hat{a}_0$	(-38.6 ± 0.1)
$\hat{a}_1$	$(3.1 \pm 0.2) \times 10^{-3}$ MeV^{-1}
$\hat{a}_2$	$(-7.6 \pm 0.5) \times 10^{-8}$ MeV^{-2}
$\hat{a}_3$	$(9.3 \pm 0.7) \times 10^{-13}$ MeV^{-3}
$\hat{a}_4$	$(-6.6 \pm 0.6) \times 10^{-18}$ MeV^{-4}
$\hat{a}_5$	$(2.7 \pm 0.3) \times 10^{-23}$ MeV^{-5}
$\hat{a}_6$	$(-5.9 \pm 0.7) \times 10^{-29}$ MeV^{-6}
$\hat{a}_7$	$(5.5 \pm 0.8) \times 10^{-35}$ MeV^{-7}

Table 4.10: Parameters for $\mathrm{TF}_{data,b}$ after fitting the TF for data.

parameter	value
$\hat{b}_0$	(1.3 ± 0.1)
$\hat{b}_1$	$(-1.3 \pm 0.7) \times 10^{-4}$ MeV^{-1}
$\hat{b}_2$	$(5.9 \pm 0.4) \times 10^{-9}$ MeV^{-2}
$\hat{b}_3$	$(-1.5 \pm 0.5) \times 10^{-13}$ MeV^{-3}
$\hat{b}_4$	$(2.3 \pm 0.8) \times 10^{-18}$ MeV^{-4}
$\hat{b}_5$	$(-2.2 \pm 0.2) \times 10^{-23}$ MeV^{-5}
$\hat{b}_6$	$(1.4 \pm 0.3) \times 10^{-28}$ MeV^{-6}
$\hat{b}_7$	$(-5.3 \pm 0.5) \times 10^{-34}$ MeV^{-7}
$\hat{b}_8$	$(1.1 \pm 0.7) \times 10^{-39}$ MeV^{-8}
$\hat{b}_9$	$(-1.0 \pm 0.1) \times 10^{-45}$ MeV^{-9}

Table 4.11: Parameters of the $\mathrm{TF}_{data,c}$ after fitting the TF for data.

For both $b\bar{b}$ and $c\bar{c}$ di-jets candidates the number of fitted QCD and Z events are the following:

$$\begin{aligned} N^{obs}_{QCD,b} &= 6253915 \pm 3194 \quad N^{obs}(Z \to b\bar{b}) = 22194 \pm 1575 \\ N^{obs}_{QCD,c} &= 24740 \pm 174 \qquad N^{obs}(Z \to c\bar{c}) = 152 \pm 64 \end{aligned} \tag{4.13}$$

with a χ^2/ndof of 55.2 (4.1) for $b\bar{b}$ ($c\bar{c}$) di-jets candidates. The sum of the two contributions gives the total number of background events, which is in agreement with the total number of events selected in the SR. In Fig.4.13 and 4.14 the fitted $Z \to b\bar{b}$ and $Z \to c\bar{c}$ resonances, the QCD background and data points in SR are shown.

4.5 Systematic uncertainties

In the next chapter limits on the number of signal events ($b\bar{b}$ or $c\bar{c}$ resonances) will be computed, but in order to get valid results it is mandatory to compute systematic uncertainties on the background model, at least the major contributions. In this thesis a very preliminary evaluation of the systematic uncertainties is performed. In the following a list of the major systematic uncertainties is presented:

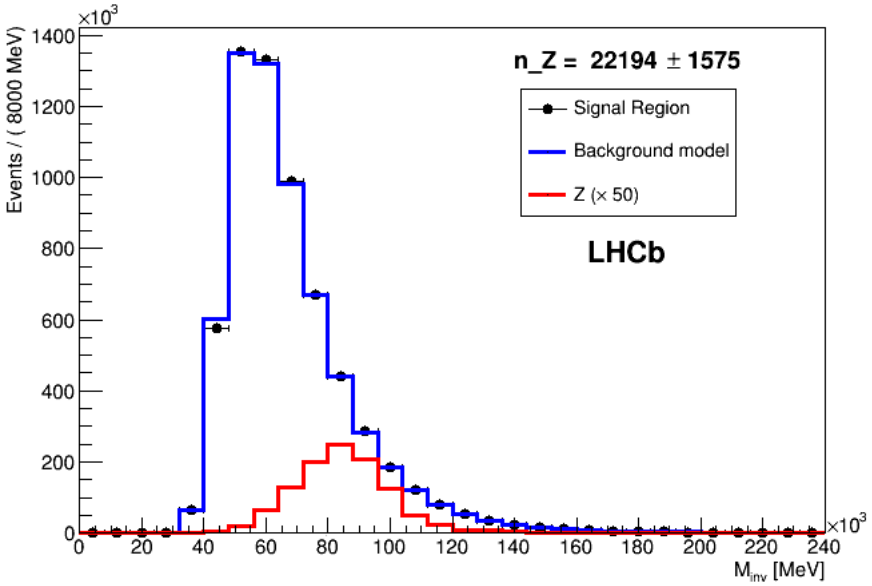

Figure 4.13: Dijet invariant mass distribution for the $Z \to b\bar{b}$ resonance fitted to data (red), background model (blue) and data (black dots).

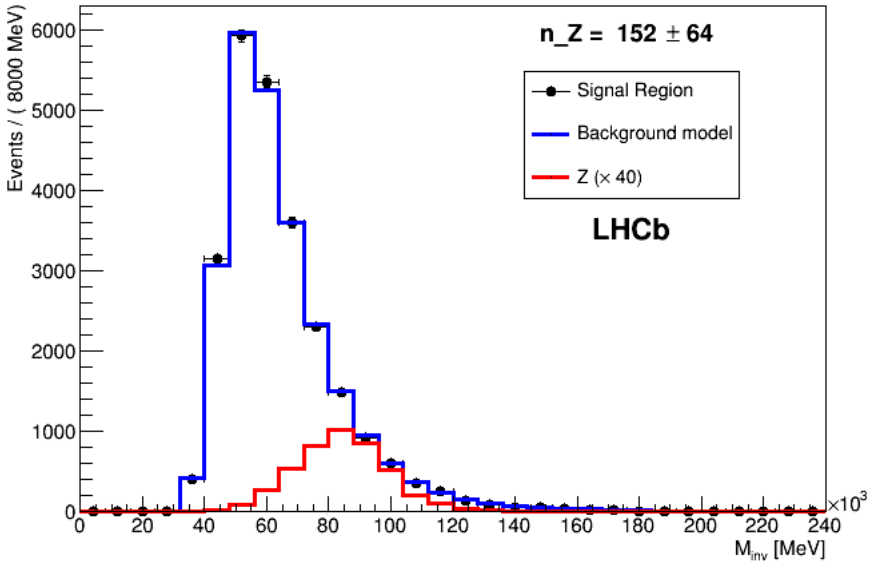

Figure 4.14: Dijet invariant mass distribution for the $Z \to c\bar{c}$ resonance fitted to data (red), background model (blue) and data (black dots).

- the greatest contribution comes from the uncertainty of the TF, which describes the background in the SR starting from the CR. For both $b\bar{b}$ and $c\bar{c}$ di-jets candidates polynomial curves of seven and nine degree are used respectively. In order to compute a systematic error in the choice of those functions, for thr moment it is considered an *upper* (*lower*) transfer function TF^{+} (TF^{-}) where all parameters are changed by adding (subtracting) to the original values their error σ_{TF} obtained from the SR fit: this results in two new TFs that define two new models for QCD background. In Fig.4.15 TF^{+} and TF^{-} are shown together with the original TF, both for $b\bar{b}$ and $c\bar{c}$ di-jets candidates.

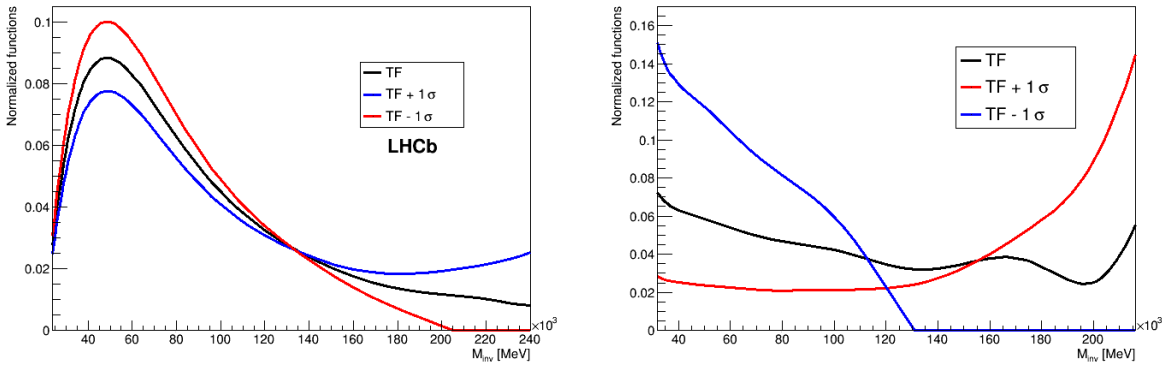

Figure 4.15: Systematics uncertainties for TFs for $b\bar{b}$ (left) and $c\bar{c}$ di-jets candidates. Variations of $\pm 1\sigma$ for all TFs parameters are also shown.

More detailed studies of the uncertainties due to the TF are in progress.

- in the limits computation the expected number of Z events, obtained with efficiencies from simulation and theoretical cross sections, will be used. Therefore an other important source of systematic uncertainty is due to the jet $b-$ and $c-$tagging efficiency uncertainty, affecting the number of predicted Z events. This error depends on the jet p_T and on the jet flavour as shown in section 3.4.1, and in order to compute it MC Z events are weighted considering an upper weight w^{up}_{tag} and a lower one w^{down}_{tag}. In this way new predicted yields on the Z events in the SR are obtained, one for the upper variation and one for the lower one, respectively $N^{up}_{SR,tag}$ and $N^{down}_{SR,tag}$. Once they are computed the systematic uncertainty is defined as

$$\sigma_{tag} = \frac{N^{up}_{SR,tag} - N^{down}_{SR,tag}}{2} \tag{4.14}$$

In this way a systematic uncertainty of $\pm 20\%$ ($^{+40}_{-20}\%$) in the number of Z events in the SR is obtained for $Z \to b\bar{b}$ ($Z \to c\bar{c}$) di-jets candidates.

- another important source of systematic uncertainty comes from the arbitrariness in defining the CR. When applying cuts on the BDTs a bias can be inserted in the fit result: it can arise from artifact structures created in the invariant mass distribution. In order to account for this source of error, cuts on $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ are varied, computing new CRs that afterwards can create new QCD background models. A variation of $\pm 20\%$ is applied to the BDTs cuts, computing an upper (lower) cut $\mathrm{BDT}^{up}_{b,c}$ ($\mathrm{BDT}^{down}_{b,c}$) both for $\mathrm{BDT}_{bc|udsg}$ and $\mathrm{BDT}_{b|c}$ and obtaining two new CRs, an upper one CR^{up} and a lower one CR^{down}. In Fig.4.16 CR^{up}, CR^{down} and original CR are shown both for $b\bar{b}$ and $c\bar{c}$ di-jets candidates.

- as studied in the MC samples, some Z events are considered in the CR, while all the events should be recognized as events in the SR. This Z events in the CR are therefore transferred in the SR as QCD background model via the TF. In order to account for this error, the following systematic uncertainty is applied to the expected number of Z events in SR:

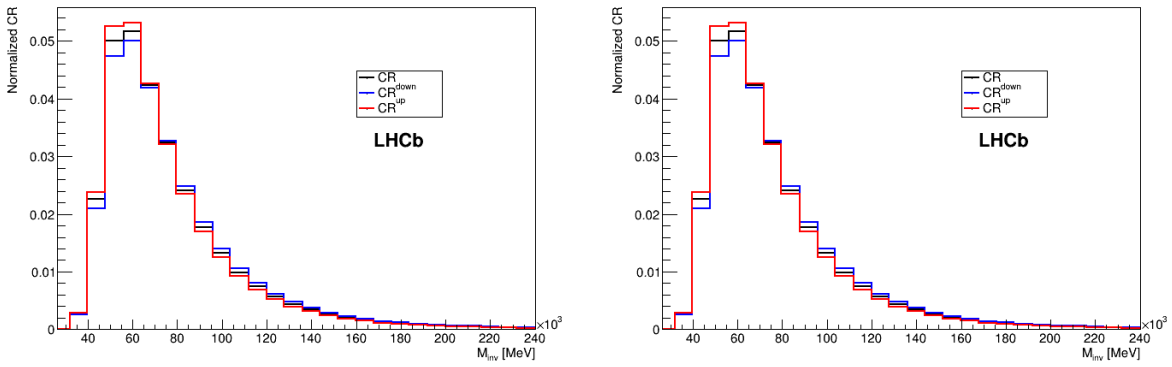

Figure 4.16: Di-jet invariant mass of $b\bar{b}$ candidates (left) and $c\bar{c}$ candidates (right) obtained by varying the BTD cuts.

$$\sigma_{Z,CR} = \frac{N_{CR}}{N_{total}^{CR}} N_{total}^{SR} \tag{4.15}$$

where N_{CR} is the number of Z events in the CR while N_{total}^{CR} is the number of data events in the CR and N_{total}^{SR} is the number of data events in the SR. For $Z \to b\bar{b}$ ($Z \to c\bar{c}$) this error amount to 6.1% (39.6%).

- other contributions to systematic uncertainties on to the Z yield prediction, related to the knowledge of the Z cross section in the forward region of pp collisions are included. These contributions have been computed in previous searches [51], and the overall systematic uncertainty affects the number of Z events with an error of $\pm 15\%$, both for $b\bar{b}$ and $c\bar{c}$ di-jets candidates.

4.6 Results

To conclude this chapter related to the construction of the background model the number of fitted Z bosons decaying into $b\bar{b}$ and $c\bar{c}$ final states is reported, where the first error is the statistic uncertainty while the second error is the systematic uncertainty due to the QCD multi-jet background modeling described in the previous paragraph:

$$\begin{aligned} N(Z \to b\bar{b}) &= 22194 \pm 1575(\text{stat}) \pm 5711(\text{sys}) \\ N(Z \to c\bar{c}) &= 152 \pm 64(\text{stat}) \ ^{+88}_{-71}(\text{sys}) \end{aligned} \tag{4.16}$$

Results are compatible with our expectations (see Table 4.7): this means that the fitting procedure performed to obtain the background model is valid, since no Z resonances are included in the QCD background.

Chapter 5

Search for $H \to b\bar{b}$ and $H \to c\bar{c}$ and high-mass resonances

5.1 Introduction

In this chapter it is described the search for particles decaying to $b\bar{b}$ and $c\bar{c}$ jets. The search is performed without following any theoretical model. The invariant mass region goes from 45 to 175 GeV in steps: 45, 60, 80, 100, 125 and 175 GeV. The resonance at 125 GeV describes the SM Higgs boson decaying in a pair of b or c quarks. For the other masses no description in the SM is present, thus they are considered as new particles that decay into b and c quarks final states. To pursue a model independent approach the following procedure is considered:

- apply event selection and BDTs cuts in order to properly select SR and CR;
- fit to SR di-jets invariant mass distribution are performed in order to search for signal events;
- if no signal events are found upper limits on events and on cross sections for the considered processes are set by using CL_s method;
- for Higgs resonances limits are compared with SM predictions.

5.2 Signal MC samples

MC samples for the signal processes with a number of events corresponding to $\mathcal{L} = 200$ pb^{-1} are generated using Pythia8 [44] and Geant4 [46–48], requiring a SM Higgs boson H and particles h_m with mass $m = 45, 60, 80, 100$ and 175 GeV to decay into $b\bar{b}$ and $c\bar{c}$ final states inside LHCb acceptance ($2 < \eta < 5$). h_m resonances are simulated following a MSSM prescription (see Sec.1.5.1). The considered production mechanism is gluon-gluon fusion, which is the most present (see Table1.1).

5.2.1 Yield prediction

Concerning kinematical and SV-tagging cuts, the same criteria discussed in Sec.4.2.1 are used; no other selection criteria are used, in order to consider inclusive processes in the most general possible way. For the Higgs boson the numbers of expected events are calculated. The theoretical cross sections in table 5.1 are used, computed by multiplying the production cross section with the $\mathcal{BR}$ of the considered process.

The acceptance and the selection efficiencies are evaluated by applying the requirements described in Chapter 4 to the MC samples. The values obtained and the final signal events expected are summarised in Table 5.2.

By using the same procedure the acceptances and the selection efficiencies are calculated for the other resonances and results are shown in Table 5.3. These numbers will be used to extract the upper limits

Process	σ_{th} [pb]	Uncertainty
$H \to b\bar{b}$	29.6	12%
$H \to c\bar{c}$	1.47	5%

Table 5.1: Theoretical production cross sections times branching ratios for Higgs boson decays.

Process	Acceptance	Selection efficiency	Expected events
$H \to b\bar{b}$	0.051	12.23%	37 ± 6
$H \to c\bar{c}$	0.064	2.15%	0.4 ± 0.1

Table 5.2: Expected number of SM Higgs boson events in the selected data sample.

on the production cross sections. In these cases there are no expected number of events since no theoretical model is assumed. It is evident that selection efficiencies for $c\bar{c}$ resonances are definitely lower than for $b\bar{b}$ resonances.

Process	Selection efficiency
$h_{45} \to b\bar{b}$	3.11%
$h_{60} \to b\bar{b}$	3.72%
$h_{80} \to b\bar{b}$	7.56%
$h_{100} \to b\bar{b}$	10.67%
$h_{175} \to b\bar{b}$	9.31%
$h_{45} \to c\bar{c}$	1.06%
$h_{60} \to c\bar{c}$	1.14%
$h_{80} \to c\bar{c}$	2.12%
$h_{100} \to c\bar{c}$	2.01%
$h_{175} \to c\bar{c}$	1.27%

Table 5.3: Selection efficiencies for the considered processes.

5.2.2 Application of SR and CR to signal resonances

Same cuts decided in Sec.4.3 are used for $b\bar{b}$ and $c\bar{c}$ di-jets SR selection. It is important to check how many MC events are selected in the SR. In Table 5.4 events in SR and CR are compared to the total number of events in the samples: for $b\bar{b}$ resonances around 90% of events are in the SR while for $c\bar{c}$ resonances the number of SR events is ten times lower.

In Figs.5.1 and 5.2 di-jets invariant mass distribution for different MC signal resonances is presented, showing also the sub-samples of SR and CR. No modifications in shape are inserted by selecting the SR.

5.3 CL_s method

If there is no evidence of signal in data, limits on the production cross sections are evaluated. The CL_s method [58] is used where two hypotheses are considered: the *null* hypothesis which assumes that the signal does not exist, and the *alternate* hypothesis which assumes that the signal is present. In order to decide which hypothesis is right a test-statistics Q is computed. Q is defined as:

$$Q = \frac{\mathcal{L}(H_{s+b})}{\mathcal{L}(H_b)} \tag{5.1}$$

where $\mathcal{L}(H_{s+b})$ ($\mathcal{L}(H_b)$) is the likelihood for the alternative (null) hypothesis. Q can be considered as

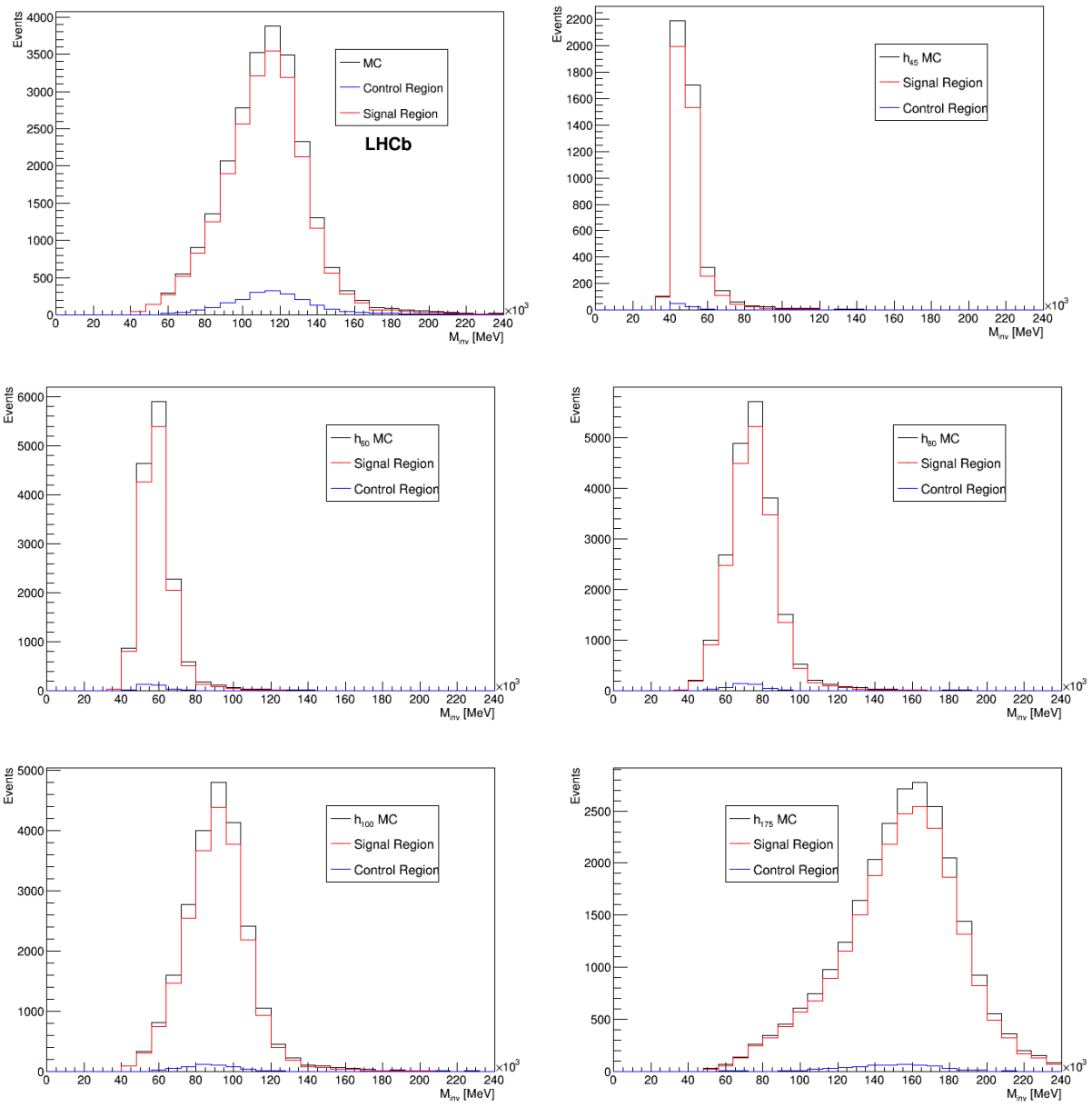

Figure 5.1: Di-jets invariant mass distribution for MC events prior to SR/CR cuts (black), SR MC events (red) and CR MC events (blue) for $b\bar{b}$ resonances. The distributions are normalized to the number of selected events in MC. From left to right and up to down: Higgs, h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

Process	total MC events	SR events	CR events	% events in SR
$H \to b\bar{b}$	24451	22174	2210	90.7%
$H \to c\bar{c}$	2418	191	1002	7.9%
$h_{45} \to b\bar{b}$	3507	3241	258	92.4%
$h_{60} \to b\bar{b}$	4633	4347	280	93.8%
$h_{80} \to b\bar{b}$	11454	10653	785	93.1%
$h_{100} \to b\bar{b}$	17598	16140	1416	91.7%
$h_{175} \to b\bar{b}$	15227	13022	2118	85.5%
$h_{45} \to c\bar{c}$	1593	133	611	8.4%
$h_{60} \to c\bar{c}$	1694	134	674	7.9%
$h_{80} \to c\bar{c}$	3086	260	1233	8.4%
$h_{100} \to c\bar{c}$	3167	261	1265	8.2%
$h_{175} \to c\bar{c}$	1429	105	533	7.3%

Table 5.4: SR, CR and total number of MC events for the considered processes. SR events for $b\bar{b}$ ($c\bar{c}$) resonances are approximately 90% (8%) of total events.

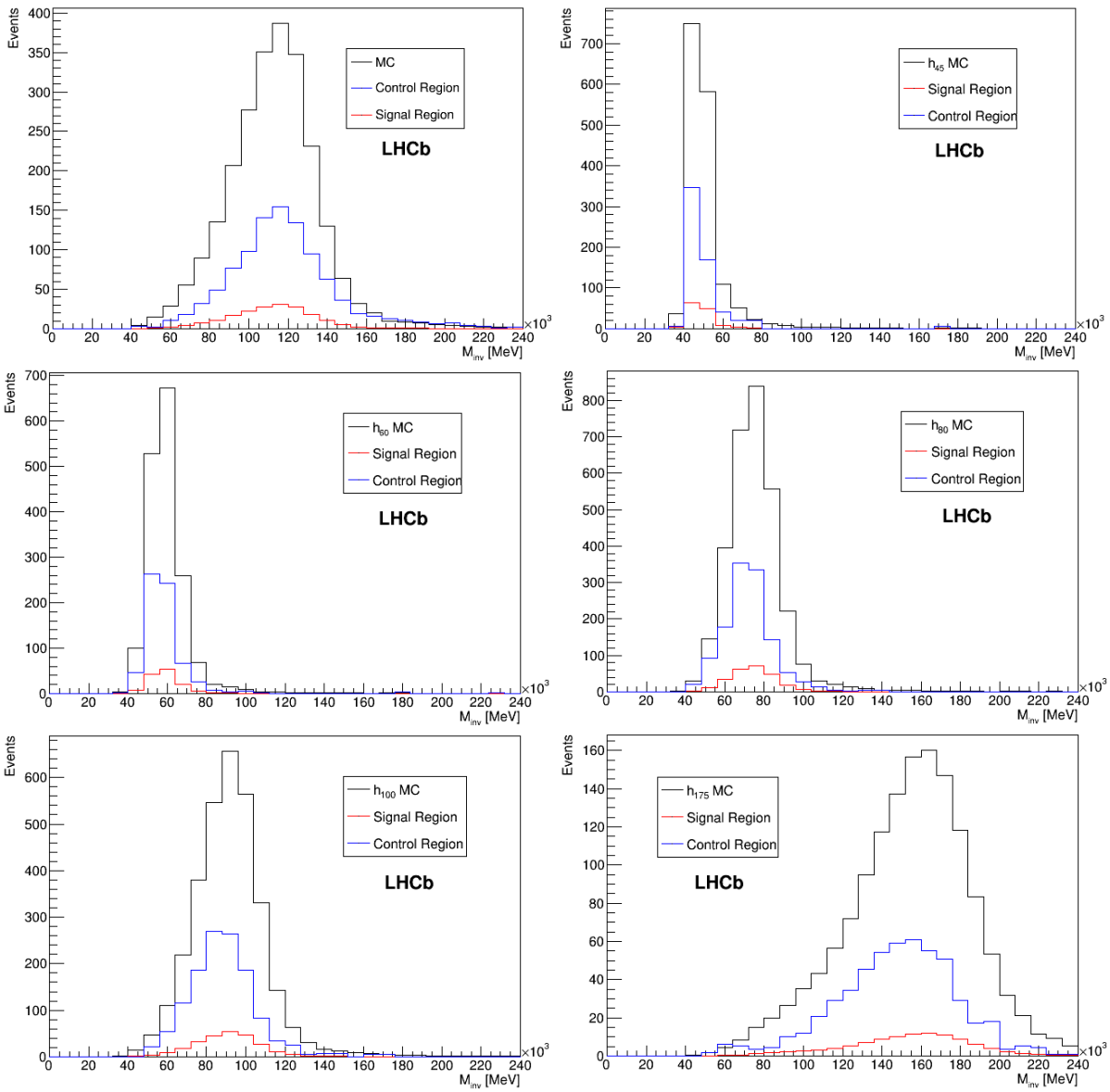

Figure 5.2: Di-jets invariant mass distribution for MC events prior to SR/CR cuts (black), SR MC events (red) and CR MC events (blue) for $c\bar{c}$ resonances. The distributions are normalized to the number of selected events in MC. From left to right and up to down: Higgs, h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

a random observable that depends on the parameters of the signal model (e.g. the mass or the cross section of the Higgs boson. Q_{obs} is calculated on data and the one-sided probability is defined as:

$$P_{s+b}(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_{s+b}}{dQ} dQ \tag{5.2}$$

where $\frac{dP_{s+b}}{dQ}$ is the probability distribution function for Q in the alternative hypothesis. The confidence level is then defined as

$$CL_{s+b} = P_{s+b}(Q \leq Q_{obs}) \tag{5.3}$$

The analogous procedure is followed for the null hypothesis:

$$CL_b = P_b(Q \leq Q_{obs}) = \int_{-\infty}^{Q_{obs}} \frac{dP_b}{dQ} dQ \tag{5.4}$$

where $\frac{dP_b}{dQ}$ is the probability distribution function for the background-only hypothesis. Both CL_{s+b} and CL_b can be computed with different methods: in this thesis the so called *asymptotic calculator* [59] is used, which uses the asymptotic formula for the test-statistic distribution and results can be obtained without generating toys (differently from the so called *frequentist calculator* [60]).
Finally the CL_s is defined by the following ratio:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{5.5}$$

In order to compute limits at a certain confidence level CL the CL_s is compared to the chosen CL, so that a signal is excluded if

$$1 - CL_s \leq \text{CL} \tag{5.6}$$

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:

$$1 - CL_s = \text{CL} \tag{5.7}$$

5.4 Systematic uncertainties

Systematic uncertainties on the background model have been determined in Sec.4.5, and they are applied as nuisance parameters to the limits computation. For each cross section hypothesis the number of expected signal events have to be calculated, applying the efficiencies obtained in MC. Therefore nuisance parameters have to be considered for the signal yield prediction in the limit computation. In the following the major systematic contributions on the tested signal yield are described:

- a first source of systematic uncertainty is due to the jet $b-$ and $c-$ tagging efficiency, affecting the number of signal events. The efficiencies are obtained from simulation, but an uncertainty due to data/MC differences have to be applied. As shown in section 3.4.1 this error depends on the jet p_T and on the jet flavour and in order to compute it signal events are weighted considering an upper weight w_{tag}^{up} and a lower one w_{tag}^{down}. In this way new yields on the signal events in the SR are obtained, one for the upper variation and one for the lower one, respectively $N_{signal\ SR,tag}^{up}$ and $N_{signal\ SR,tag}^{down}$. Once they are computed the systematic uncertainty is defined as:

$$\sigma_{signal,tag} = \frac{N^{up}_{signal\ SR,tag} - N^{down}_{signal\ SR,tag}}{2} \tag{5.8}$$

For $b\bar{b}$ ($c\bar{c}$) di-jets candidates this systematic uncertainty on the number of signal events is about $\pm 20\%$ ($^{+20\%}_{-40\%}$).

- another contribution to systematic errors comes from the fact that some signal events contaminate the CR, and they are transferred to the SR background model via the TF. Thus the number of signal events in the CR can be considered as a systematic uncertainty due to the application of cuts on BDTs; with N_{CR} the number of signal events in the CR, N^{CR}_{total} the total number of events in the CR, and N^{SR}_{total} the total number of events in the SR, $\sigma_{signal,CR}$ is defined as

$$\sigma_{signal,CR} = \frac{N_{CR}}{N^{CR}_{total}} N^{SR}_{total} \tag{5.9}$$

As shown in Table5.2 $\sigma_{signal,CR}$ depends on the considered process; values of $\sigma_{signal,CR}$ for all processes are reported in Table 5.5. It is evident that for $c\bar{c}$ resonances this systematic uncertainty is greater since the selection efficiencies for the SR for $c\bar{c}$ resonances are consistently lower than for $b\bar{b}$ resonances.

Process	total events	CR events	$\sigma_{signal,CR}$
$H \to b\bar{b}$	24451	2210	9%
$H \to c\bar{c}$	2418	1002	41.5%
$h_{45} \to b\bar{b}$	4657	90	1.9%
$h_{60} \to b\bar{b}$	14880	331	2.2%
$h_{80} \to b\bar{b}$	21005	487	2.3%
$h_{100} \to b\bar{b}$	23185	546	2.4%
$h_{175} \to b\bar{b}$	24768	588	2.4%
$h_{45} \to c\bar{c}$	1593	611	36.1%
$h_{60} \to c\bar{c}$	1694	674	39.8%
$h_{80} \to c\bar{c}$	3086	1233	39.9%
$h_{100} \to c\bar{c}$	3167	1265	39.9%
$h_{175} \to c\bar{c}$	1429	533	37.3%

Table 5.5: $\sigma_{signal,CR}$ for the considered processes. For $b\bar{b}$ ($c\bar{c}$) resonances this systematic error is around 3% (40%).

Systematic uncertainties coming from both background model and signal resonances have to be included in the CL_s method, since they take part in defining the error in the limit procedure. Other systematic uncertainties, like these related to the jet energy scale and jet energy resolution, are neglected in this thesis since they are considered sub-dominant.

5.5 Results

In order to search for heavy flavour di-jets resonances first a fit to data is performed, and if result is compatible with zero, i.e. no signal is found, the upper limit on the production cross section is calculated.

5.5.1 Fit results

The number of events for each process is computed fitting data with a function composed by *pdf* for background and *pdf* for signal. These *pdfs* are obtained as histograms since a binned maximum likelihood approach is used. The model used in the fit is the following:

$$f_{i,j}(x) = N^{obs}_{QCD,i} \cdot B_{QCD,i}(x) + N^{exp}_{Z,i} \cdot B_{Z,i}(x) + N^{obs}_{S_{i,j}} \cdot S_{i,j}(x) \tag{5.10}$$

where x is the invariant mass, $B_{QCD,i}(x)$ is the QCD background model for $i = b, c$ di-jets candidates obtained in Section 4.4, $B_{Z,i}$ is the Z boson model obtained from MC for $i = b, c$, $S_{i,j}$ is the signal model obtained from MC and shown in Figs.5.1 and 5.2, for $i = b, c$ resonances and for each resonance with mass j. $N^{obs}_{QCD,i}$ is the number of observed QCD background events, $N^{exp}_{Z,i}$ is the number of Z events (fixed to expectation) while $N^{obs}_{S_{i,j}}$ is the number of observed signal events. The parameters of the QCD background model are fixed to those obtained in Section 4.4. The number of QCD and signal events are left free to vary. The fit is performed with the binned maximum likelihood technique using the *Roofit* package [56] of the software ROOT [41]. The fitting algorithm used is *Minuit2* [57]. Di-jets invariant mass distribution for data with the results of the fit superimposed is shown in Figs.5.3 and 5.4 for $b\bar{b}$ and $c\bar{c}$ resonances searches.

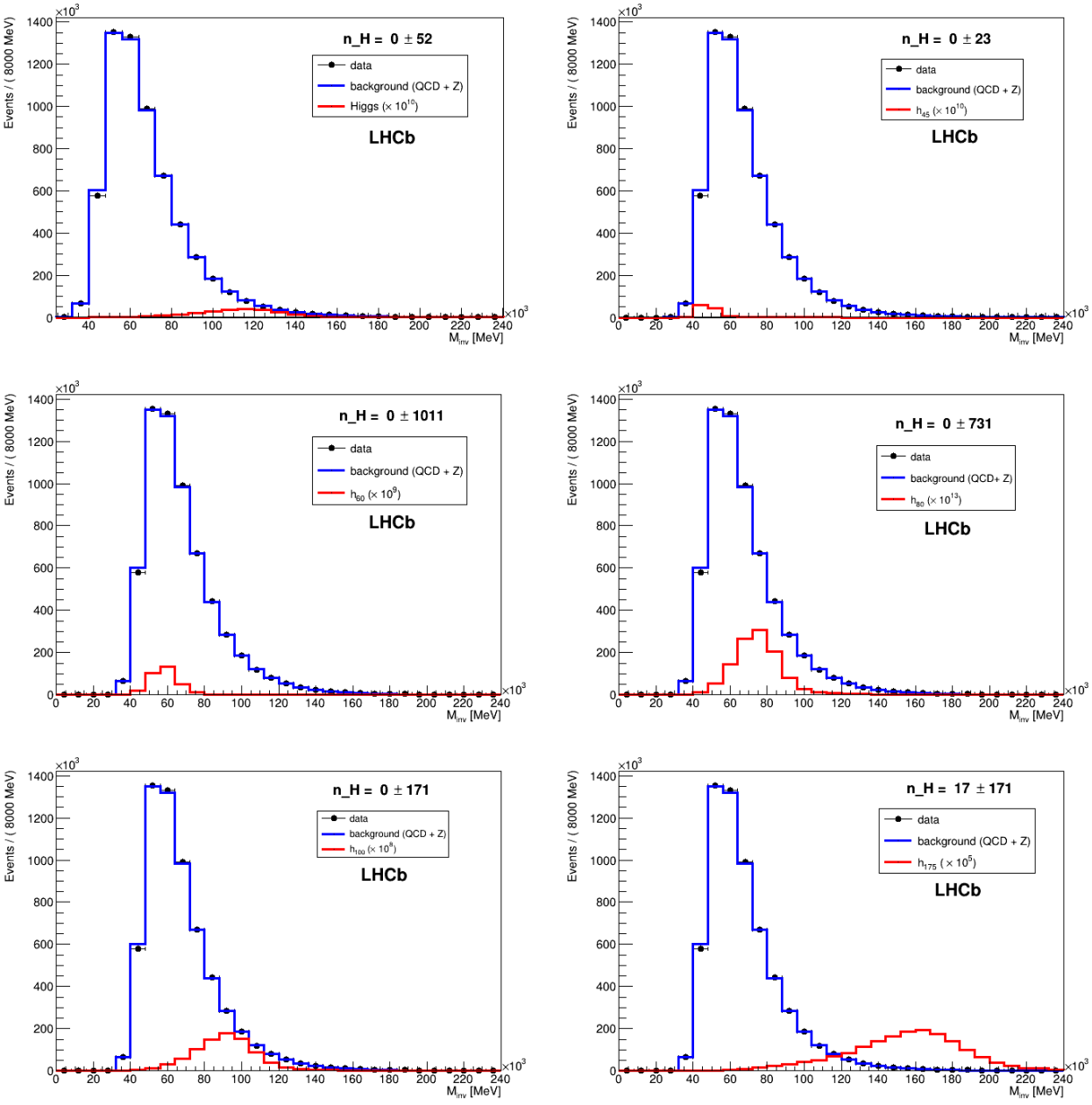

Figure 5.3: $b\bar{b}$ di-jets invariant mass distribution with the result of the fit superimposed. The number of events of searched signal is multiplied by power of 10. From left to right and up to down: Higgs, h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

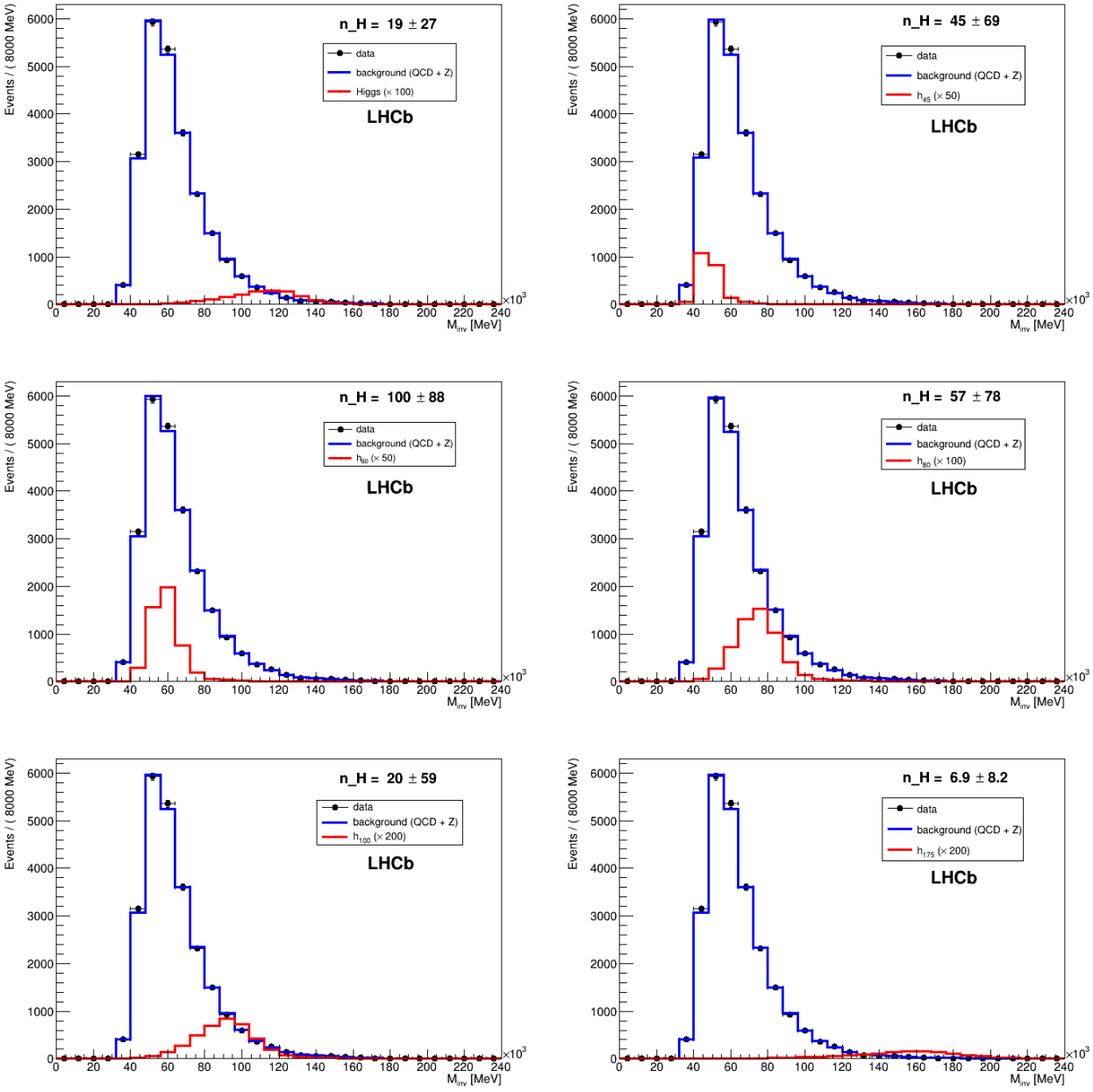

Figure 5.4: $c\bar{c}$ di-jets invariant mass distribution with the result of the fit superimposed. The number of events of searched signal is multiplied by power of 10. From left to right and up to down: Higgs, h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

The number of fitted events for each search is summarized here:

$$\begin{aligned} N(H \to b\bar{b}) &= 0 \pm 52 \\ N(h_{45} \to b\bar{b}) &= 0 \pm 23 \\ N(h_{60} \to b\bar{b}) &= 0 \pm 1011 \\ N(h_{80} \to b\bar{b}) &= 0 \pm 731 \\ N(h_{100} \to b\bar{b}) &= 0 \pm 171 \\ N(h_{175} \to b\bar{b}) &= 17 \pm 171 \\ N(H \to c\bar{c}) &= 19 \pm 27 \\ N(h_{45} \to c\bar{c}) &= 45 \pm 69 \\ N(h_{60} \to c\bar{c}) &= 100 \pm 88 \\ N(h_{80} \to c\bar{c}) &= 57 \pm 78 \\ N(h_{100} \to c\bar{c}) &= 20 \pm 59 \\ N(h_{175} \to c\bar{c}) &= 7 \pm 8 \end{aligned} \tag{5.11}$$

5.5.2 Limits with CL_s method

In the previous section the fit results show that no signal is present for any of the considered resonances, since the fitted value is compatible with zero. Limits are computed by means of CL_s method. Starting with the processes involving Higgs boson decay, in the following pictures the CL_s results are shown.

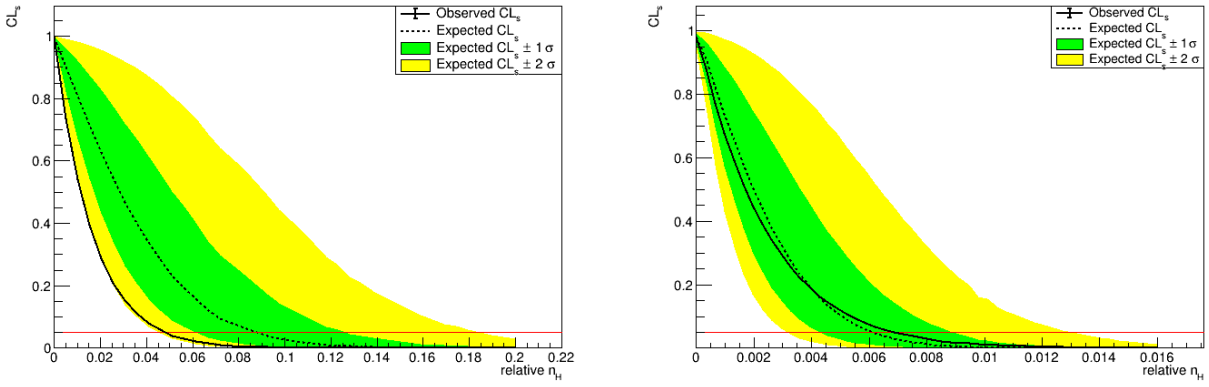

Figure 5.5: CL_s limits for Higgs $b\bar{b}$ ($c\bar{c}$) resonance on the left (right).

The dotted (filled) black line represents the expected (observed) number of events in the only-background (signal+background) hypothesis, while the green (yellow) bands define the $\pm 1\sigma$ ($\pm 2\sigma$) uncertainties which include the systematic uncertainties computed above. The red line defines the 95% CL and on the $x-$axis the relative number is defined as the ratio of events in the SR with respect to total number of events, as reported in Table5.2. In order to get the expected and the observed limits on the number of events it is sufficient to intersect the dotted and filled black lines with the red one, getting respectively the expected and the observed limit. It is thus possible to get the following results for expected and observed limits:

$$\begin{aligned} N_{obs}(H \to b\bar{b}) &= 1086 \quad & N_{exp}(H \to b\bar{b}) &= 1951^{+842}_{-576} \\ N_{obs}(H \to c\bar{c}) &= 29 \quad & N_{exp}(H \to c\bar{c}) &= 86^{+44}_{-25} \end{aligned} \tag{5.12}$$

The same approach is followed for all the other mass resonances. In Figs.5.6 and 5.7 CL_s limits are shown respectively for $b\bar{b}$ and $c\bar{c}$ resonances.

Expected and observed limits can be computed for $b\bar{b}$ resonances

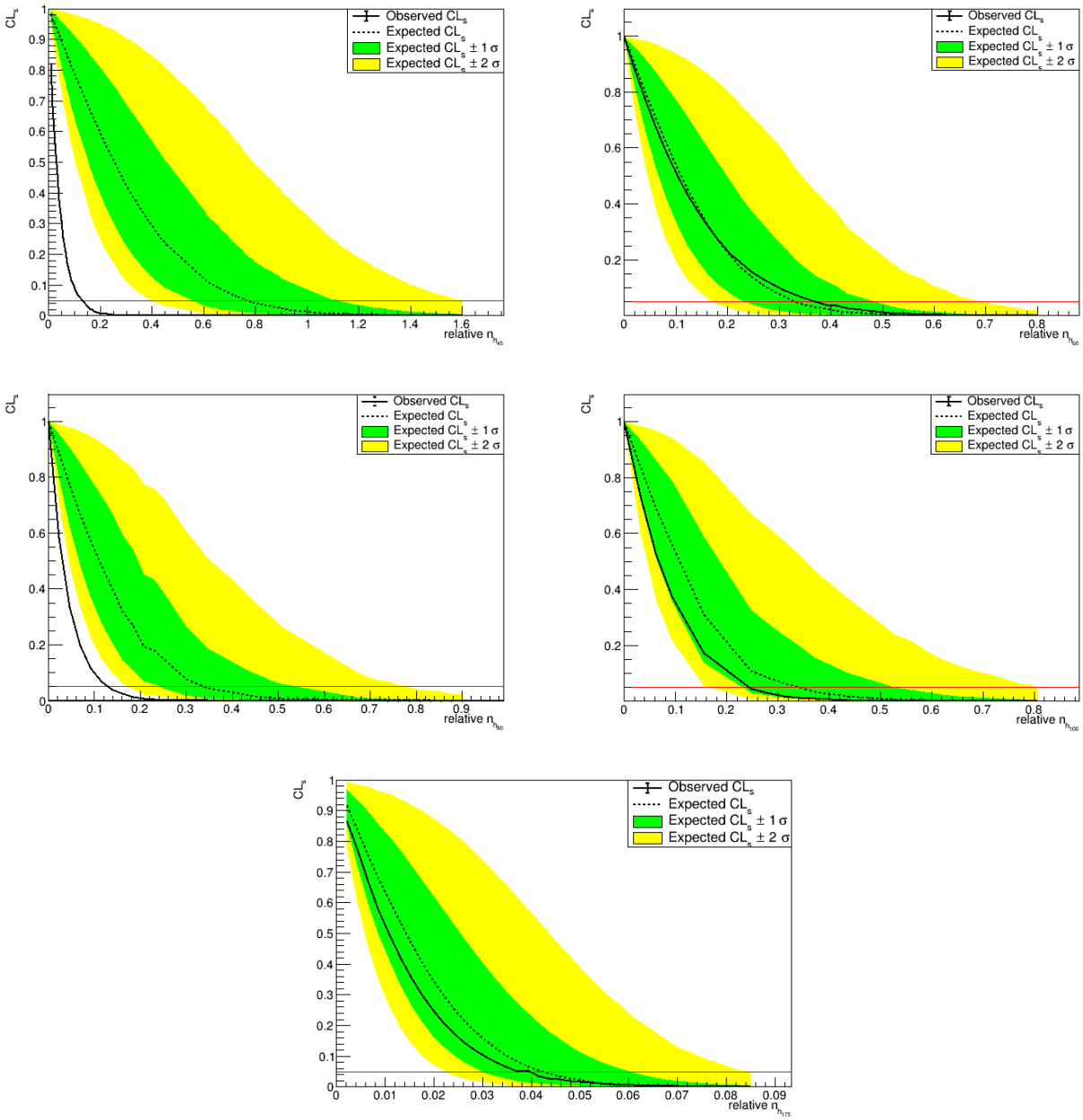

Figure 5.6: CL_s limits for $b\bar{b}$ resonances. From left to right and up to down: h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

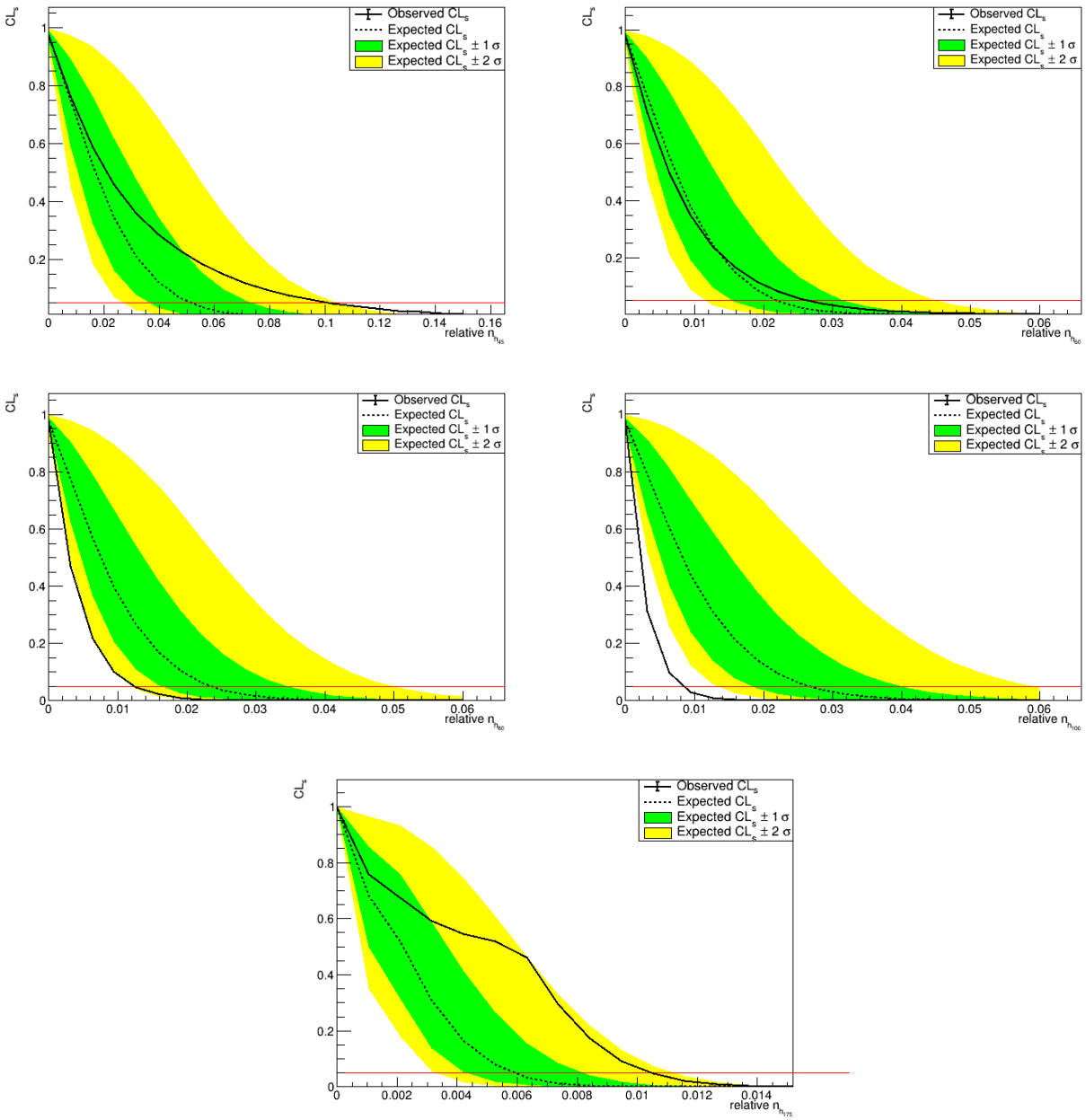

Figure 5.7: CL_s limits for $c\bar{c}$ resonances. From left to right and up to down: h_{45}, h_{60}, h_{80}, h_{100} and h_{175}.

$$
\begin{aligned}
N_{obs}(h_{45} \to b\bar{b}) &= 421 & N_{exp}(h_{45} \to b\bar{b}) &= 2463^{+1166}_{-680} \\
N_{obs}(h_{60} \to b\bar{b}) &= 1608 & N_{exp}(h_{60} \to b\bar{b}) &= 1434^{+695}_{-391} \\
N_{obs}(h_{80} \to b\bar{b}) &= 1384 & N_{exp}(h_{80} \to b\bar{b}) &= 3622^{+1811}_{-1065} \\
N_{obs}(h_{100} \to b\bar{b}) &= 3873 & N_{exp}(h_{100} \to b\bar{b}) &= 5487^{+2582}_{-2098} \\
N_{obs}(h_{175} \to b\bar{b}) &= 508 & N_{exp}(h_{175} \to b\bar{b}) &= 547^{+260}_{-156}
\end{aligned}
\tag{5.13}
$$

and for $c\bar{c}$ resonances

$$
\begin{aligned}
N_{obs}(h_{45} \to c\bar{c}) &= 159 & N_{exp}(h_{45} \to c\bar{c}) &= 79^{+36}_{-16} \\
N_{obs}(h_{60} \to c\bar{c}) &= 22 & N_{exp}(h_{60} \to c\bar{c}) &= 39^{+18}_{-11} \\
N_{obs}(h_{80} \to c\bar{c}) &= 40 & N_{exp}(h_{80} \to c\bar{c}) &= 71^{+33}_{-21} \\
N_{obs}(h_{100} \to c\bar{c}) &= 28 & N_{exp}(h_{100} \to c\bar{c}) &= 85^{+44}_{-25} \\
N_{obs}(h_{175} \to c\bar{c}) &= 14 & N_{exp}(h_{175} \to c\bar{c}) &= 8^{+7}_{-4}
\end{aligned}
\tag{5.14}
$$

As a last result, the selection efficiencies computed in Sec.5.2.1 can be used in order to get the production cross sections for the considered process using the following equation:

$$\sigma = \frac{N}{\mathcal{L} \cdot A \cdot \epsilon_{cut}} \tag{5.15}$$

with N the number of events. In this way an expected (observed) production cross section times branching ratio σ_{exp} (σ_{obs}) can be computed considering the expected (observed) number of events N_{exp} (N_{obs}) computed previously, and in the case of the Higgs processes it can be compared with SM predictions. In Fig.5.9 limits computed above with respect to the resonance mass $m_{H,h}$, both for $b\bar{b}$ and $c\bar{c}$ resonances are shown.

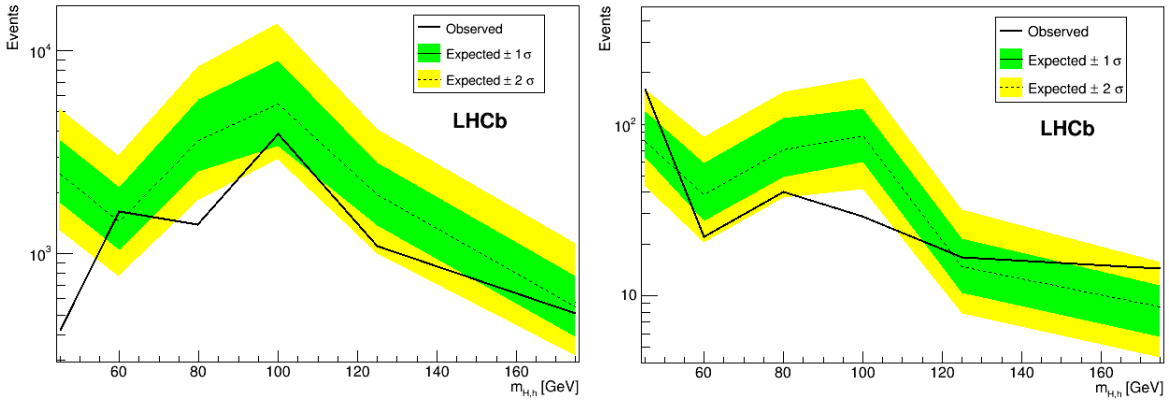

Figure 5.8: CL_s limits for $b\bar{b}$ and $c\bar{c}$ resonances (respectively on the left and right) with respect to the mass resonance $m_{H,h}$.

A few comments regarding these last results:

- for almost all resonance searches, values for observed events are inside the $\pm 2\sigma$ error. Therefore no significant deviations from the background-only model are found;
- better limits are obtained for values of the invariant mass $m > 100$ GeV, where it is easier to model the QCD background. Indeed a greater error band is present in the mass invariant region of the QCD background peak;

- considered that the systematic uncertainties estimation is only preliminary, values for expected and observed number of events may change.

As a last step using Eq. 5.15 and selection efficiencies reported in Table5.3 the same two graphs can be shown, this time representing observed and expected production cross sections:

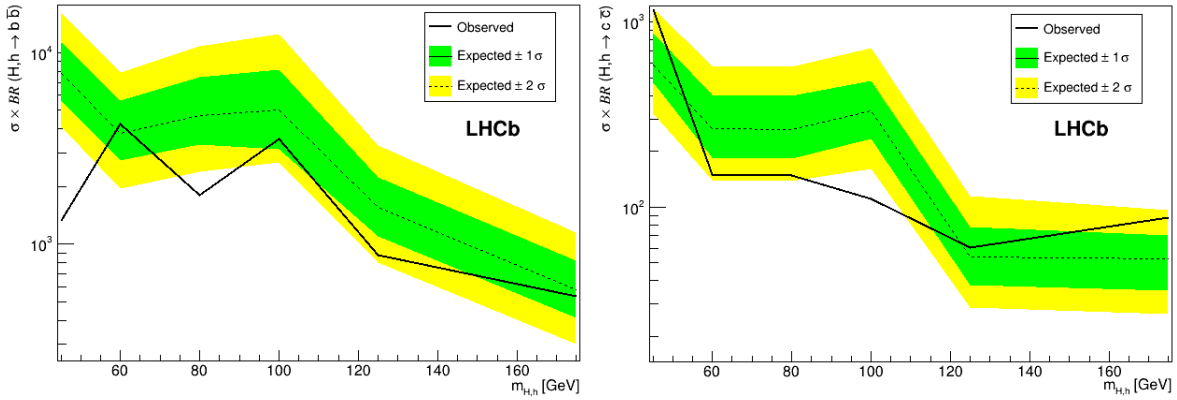

Figure 5.9: CL_s limits on production cross sections for $b\bar{b}$ and $c\bar{c}$ resonances (respectively on the left and right) with respect to the mass resonance $m_{H,h}$.

The following limits at 95% CL on observed production cross sections are evaluated:

$$
\begin{aligned}
&\sigma(H \to b\bar{b}) = \sigma(pp \to H) \times \mathcal{BR}(H \to b\bar{b}) < 871 \text{ pb} \\
&\sigma(H \to c\bar{c}) = \sigma(pp \to H) \times \mathcal{BR}(H \to c\bar{c}) < 61 \text{ pb} \\
&\sigma(h_{45} \to b\bar{b}) = \sigma(pp \to h_{45}) \times \mathcal{BR}(h_{45} \to b\bar{b}) < 1328 \text{ pb} \\
&\sigma(h_{60} \to b\bar{b}) = \sigma(pp \to h_{60}) \times \mathcal{BR}(h_{60} \to b\bar{b}) < 4238 \text{ pb} \\
&\sigma(h_{80} \to b\bar{b}) = \sigma(pp \to h_{80}) \times \mathcal{BR}(h_{80} \to b\bar{b}) < 1795 \text{ pb} \\
&\sigma(h_{100} \to b\bar{b}) = \sigma(pp \to h_{100}) \times \mathcal{BR}(h_{100} \to b\bar{b}) < 3559 \text{ pb} \\
&\sigma(h_{175} \to b\bar{b}) = \sigma(pp \to h_{175}) \times \mathcal{BR}(h_{175} \to b\bar{b}) < 534 \text{ pb} \\
&\sigma(h_{45} \to c\bar{c}) = \sigma(pp \to h_{45}) \times \mathcal{BR}(h_{45} \to c\bar{c}) < 1174 \text{ pb} \\
&\sigma(h_{60} \to c\bar{c}) = \sigma(pp \to h_{60}) \times \mathcal{BR}(h_{60} \to c\bar{c}) < 150 \text{ pb} \\
&\sigma(h_{80} \to c\bar{c}) = \sigma(pp \to h_{80}) \times \mathcal{BR}(h_{80} \to c\bar{c}) < 147 \text{ pb} \\
&\sigma(h_{100} \to c\bar{c}) = \sigma(pp \to h_{100}) \times \mathcal{BR}(h_{100} \to c\bar{c}) < 110 \text{ pb} \\
&\sigma(h_{175} \to c\bar{c}) = \sigma(pp \to h_{175}) \times \mathcal{BR}(h_{175} \to c\bar{c}) < 87 \text{ pb}
\end{aligned}
\tag{5.16}
$$

Considering the Higgs resonance, it is interesting to compare these limits at 95% CL with SM predictions: for the decay $H \to b\bar{b}$ the observed production cross section times the branching ratio is 29 times the SM expectation, while for the decay $H \to c\bar{c}$ the observed production cross section times the branching ratio is 42 times the SM expectation. These results show excellent improvements with respect to latest LHCb [26] analysis considering Higgs production with a vector boson, where for $H \to b\bar{b}$ ($H \to c\bar{c}$) the observed production cross section times branching ratio was 50 (6400) times the SM expectation.

Chapter 6

Future developments

6.1 Introduction

Run II campaign ended in 2018 and LHCb has stored a data set larger than the one used in this thesis is ready to be analyzed. For the next data taking campaign the experiment is undergoing an intense upgrade of the detector and data acquisition. On a parallel path, new analysis techniques are under development, considering new computational tools that, hopefully, will be more efficient than the ones already present. In this chapter the main aspects that will allow to improve the searches presented in the thesis are described:

- the improved limits on cross sections times the branching ratio of heavy flavor di-jets high-mass resonances due to the increased number of collected data thanks to the upgraded data acquisition;
- upgrade of LHCb's Vertex Locator (VELO) for the Run III campaign;
- new computational approach based on a *Deep-Neural-Network* (DNN) architecture to identify $b-$ and $c-$ jets.

6.2 Increased data set

This analysis uses only a small fraction of data corresponding to an integrated luminosity ($\mathcal{L} = 200$ pb^{-1}). The LHCb experiment is performing an important upgrade to be able to collect as many as possible data in the next years. Moreover, the LHC machine will be upgraded to the so called High luminosity (HL-LHC) after 2025. The results obtained in this thesis can be scaled for the larger number of events. The luminosity factor is defined as

$$f = \sqrt{\frac{\mathcal{L} = 200 \text{ pb}^{-1}}{\mathcal{L}_{new}}} \tag{6.1}$$

where $\mathcal{L}_{new}$ is the new integrated luminosity. Assuming the same centre-of-mass energy, $\sqrt{s} = 13$ TeV, two different scenarios are considered:

- data from Run II, Run III and Run IV are used for the measurement. The corresponding integrated luminosity is $\mathcal{L}_{new-1} = 50$ fb^{-1};
- HL-LHC is included, data from Run II up to 2035 are used. The integrated luminosity is $\mathcal{L}_{HL-LHC} = 300$ fb^{-1}.

Using Eq.6.1 the luminosity factors for both scenarios are

$$\begin{aligned} f_{new-1} &= 0.063 \\ f_{HL-LHC} &= 0.026 \end{aligned} \tag{6.2}$$

In this way production cross sections times branching ratio results from the CL_s method for the Higgs resonances are multiplied by these factors in order to get new limits in the two possible scenarios, as shown in Table6.1.

process	new-1 scenario ($\mathcal{L} = 50$ fb^{-1})	HL-LHC scenario ($\mathcal{L} = 300$ fb^{-1})
$H \to b\bar{b}$	$\sigma < 54.9$ pb	$\sigma < 21.2$ pb
$H \to c\bar{c}$	$\sigma < 3.85$ pb	$\sigma < 1.59$ pb

Table 6.1: New limits on production cross section times branching ratio $\sigma \times \mathcal{BR}$ for the two scenarios described in the text.

In the HL-LHC scenario it will be possible to measure the branching ratio of $H \to b\bar{b}$ and $H \to c\bar{c}$, as the limits are very close to the SM expectations. In the new-1 scenario with no HL-LHC LHCb will be close to the reconstruct the two decay channels but additional improvements are needed.

6.3 LHCb VELO upgrade

As already said, before the start of Run III campaign, the second *Long Shutdown* (LS2) gives the opportunity to upgrade LHCb experiment. This is needed because the LHC peak luminosity will be five times higher than Run II, with a nominal value of $\mathcal{L}_{peak} = 4 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ and the trigger and data acquisition systems will be upgraded to sustain a data rate of about 40 MHz. All the LHCb sub-detectors are improved but the more relevant for this measurements is the tracking system and in particular the vertex locator. The VELO is going through an intense upgrade [62], in order to satisfy the following requirements:

- collecting data rates of 2.8 Tb/s;
- sensors kept at a temperature $T < -20°$ C;
- improved performances, resolution in track reconstruction and tagging performances.

The fully upgraded VELO is expected to be mounted by September 2019, and its major improvements are:

- a new pixel detector with more and thinner modules replacing the r and ϕ strips, in order to reduce the number of ghost tracks and get a faster reconstruction algorithm;
- new front-end electronics in order to cope with higher readout data rates and higher HV tolerance (around 1000 V);
- thinner RF-aluminium foils with a thickness of ~ 250 μm, in order to improve the resolution of the Interaction Point (IP);
- a more efficient cooling interface made by 500 μm thick silicon substrate with integrated micro channels;

In Fig.6.1 [62] a schematic layout of the upgraded VELO is shown, particularly the new pixel detector is evident.

Upgrades on VELO will in general guarantee better reconstruction efficiency for particles which are reconstructible as VELO tracks. For this analysis improvements on the resolution of the Impact Parameter (IP) are fundamental, since in this way the SVs selection efficiency will increase and it will be possible to better separate b from c jets, e.g. by a better measurement of their time of flight.

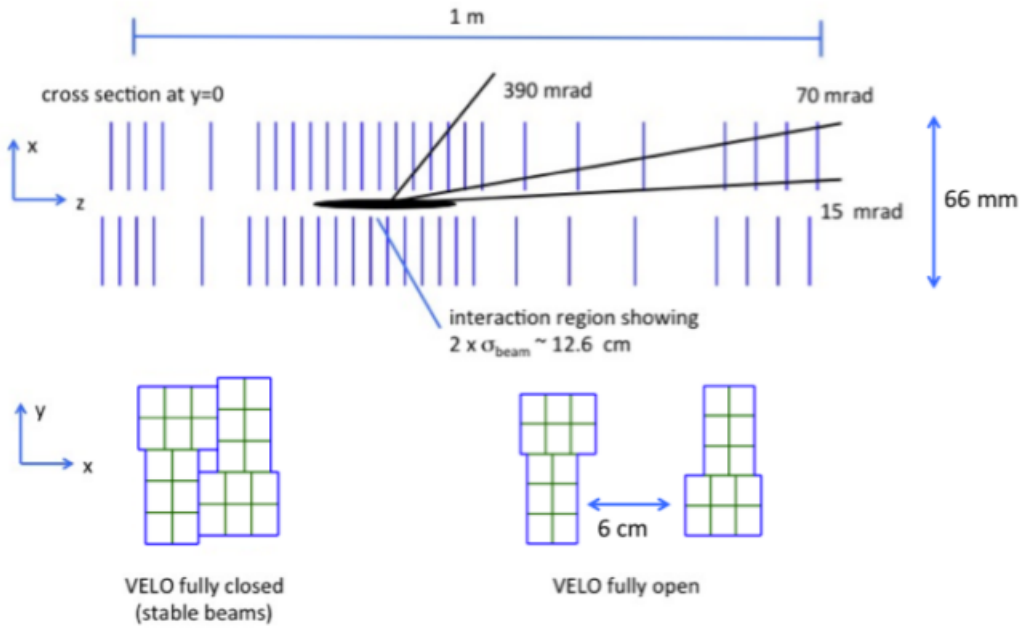

Figure 6.1: General structure of the new upgraded VELO [62].

In Fig.6.2 $x-$axis and $3D$ resolution for the IP are shown with respect to the inverse of the track transverse momentum p_T, comparing actual VELO (black dots) with future upgraded VELO (red squares).

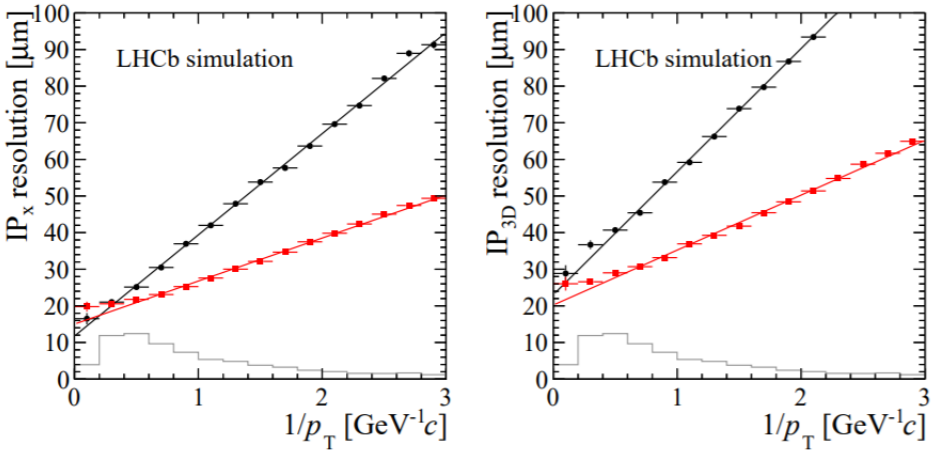

Figure 6.2: $x-$axis (left) and $3D$ (right) resolution for the IP are shown with respect to the inverse of the track transverse momentum p_T, comparing actual VELO (black dots) with future upgraded VELO (red squares) [62]. $x-$ and $y-$ axis resolutions are similar.

6.4 New analysis techniques

The tagging performances of the flavour tagging algorithm have been discussed in Chapter 3, analyzing efficiencies for b and c quarks and mis-identification probability for light quarks and gluons. New computational instruments can be considered instead of relying on the BDTs, such as a DNN architecture, described in the following.

6.4.1 DNN scheme

An Artificial Neural Network (ANN) is a computational instrument inspired by biological neural networks: such systems use examples to learn how to do a particular task, generally without task-

specific programming. An ANN is made by artificial neurons (called *nodes*) which are connected to each other by *synapses* in order to transmit signals; to nodes and synapses particular weights are associated, so that starting from an *input* signal given to the ANN an *output* signal is produced. Changing these weights changes the output, so that if the output does not resemble the expected result weights are changed to "learn" how to get the proper output. Nodes are typically organized in layers, and if multiple layers are present than the ANN is called DNN. DNNs can model complex non-linear relationships and usually they are feedforward, which means that signals are transmitted from the input layer to the output layer without going back. In Fig.6.3 a pictorial representation of a DNN is shown.

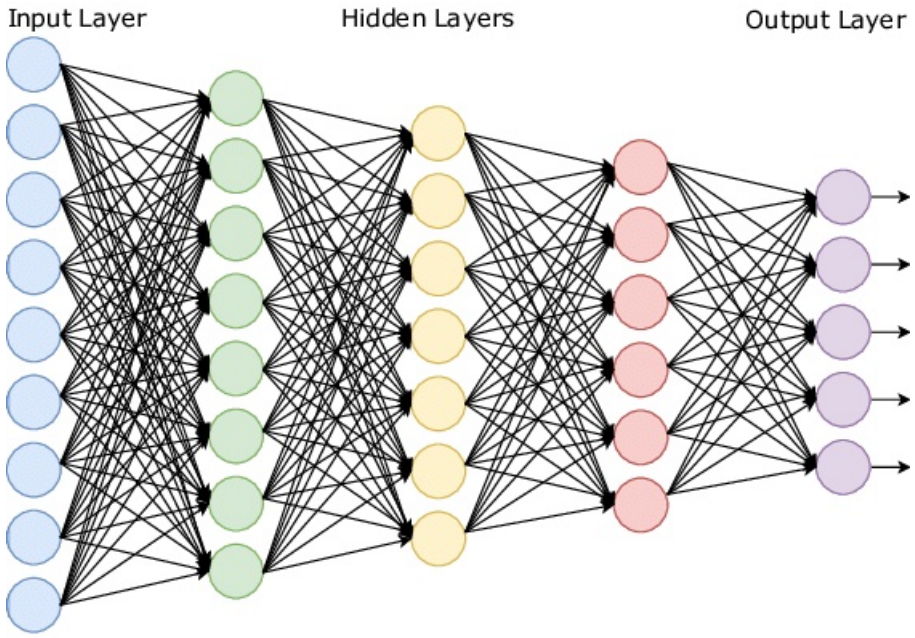

Figure 6.3: DNN general scheme.

6.4.2 DNN for jet tagging

The DNN architecture can be used to tag jets flavour: generally speaking by giving to the DNN jets with known flavour (so that the flavour is b, c or q) the algorithm learns how to tag a jet autonomously. In order to do so the QCD MC samples discussed in Chapter 4 are used, divided in $b\bar{b}$, $c\bar{c}$ $q\bar{q}$ di-jets candidates. The general idea is that the DNN algorithm should replace te SV tagging algorithm: variables related to the jet sub-structure are used as input for the DNN without requiring the presence of a secondary vertex. Thus the DNN selects jets with and without SVs and by means of its multiple layers it tries to tag di-jets candidates, defining also their flavour. A jet constituent matrix is created and used as input for the DNN. There are 369 elements, and each element describes an observable for a particle. A fixed number of particles is considered, if there are less particles empty elements are put at zero. In the jet constituent matrix particles are also ordered using a feature that depends on the particle type. The elements are grouped in the following way:

- *charged particles* ordered by decreasing impact parameter: 16 variables used to describe 10 charged particles, in total there are 160 variables;
- *neutral particles* ordered by decreasing transverse energy: 12 variables describing 15 neutral particles, in total there are 180 variables;
- *global features*: 14 variables that globally describe the jet;
- *SV features*: 15 variables describing the secondary vertex, already used in for the usual jet tagging. If no SV is found then these elements are set to zero.

As output the DNN gives the probability that a jet is tagged with a specific flavour, thus the possible

results are:

- the distribution probability that a jet is tagged as a $b-$jet P_b;
- the distribution probability that a jet is tagged as a $c-$jet P_c;
- the distribution probability that a jet is tagged as a light jet P_q.

It is obvious that $P_b + P_c + P_q = 1$. In Fig.6.4 a general scheme for the DNN used for jet tagging is shown, where all variables are listed. In the following section the application of an already-trained DNN to QCD MC samples containing di-jets candidates is described.

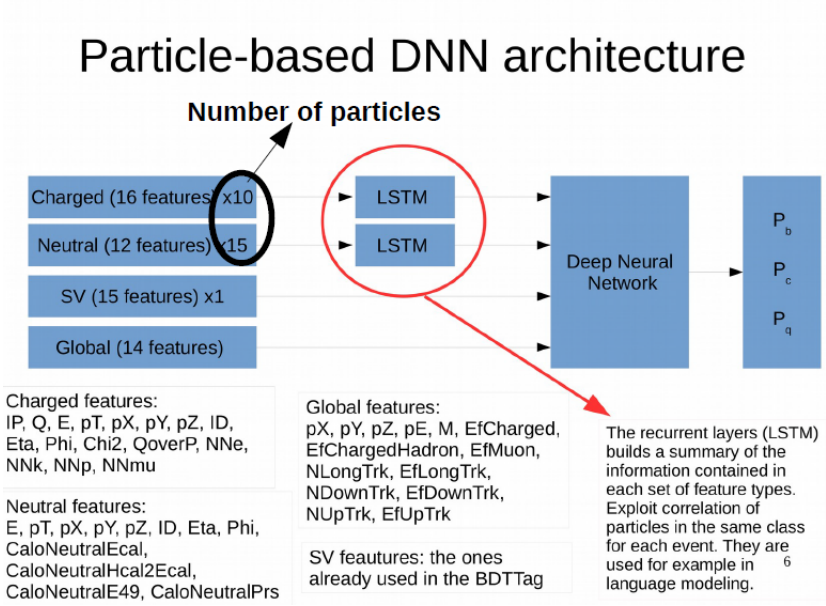

Figure 6.4: General scheme for the DNN used for jet tagging.

6.4.3 Application of DNN to di-jets candidates

Dijet candidates are selected using the same kinematical cuts applied in previous chapters, but requiring no criteria for the selection of a secondary vertex. Kinematical cuts are recalled here for completeness:

- the jet transverse momentum p_T is greater than 20 GeV;
- $2.2 < \eta_{jet} < 4.2$;
- $\Delta\phi > 1.5$, where $\Delta\phi$ is the angular distance between the two jets.

The output of the DNN is shown in Fig.6.5 for b (left), c (center) and q (right) di-jets candidates: the distribution probalities that a jet is tagged as a b, c or q are respectively pictured in red, green and blue curves.

It is evident that only for $b\bar{b}$ di-jets candidates jets are quite correctly tagged, with peaks at $P = 1$ for P_b and at $P = 0$ for P_c and P_q. For all di-jets candidates the P_c distribution has a typical pattern for $0.2 < P < 0.6$ that is still unclear. Generally speaking some changes should be applied to the DNN in order to get better probability distributions.

6.4.4 DNN efficiency

Besides previous results it is still interesting to compare the DNN architecture with the actual SV-tagging algorithm by computing efficiencies for both computational tools. Here no kinematical cuts are applied, and efficiencies are computed for different p_T intervals (expressed in GeV): $p_T < 20$,

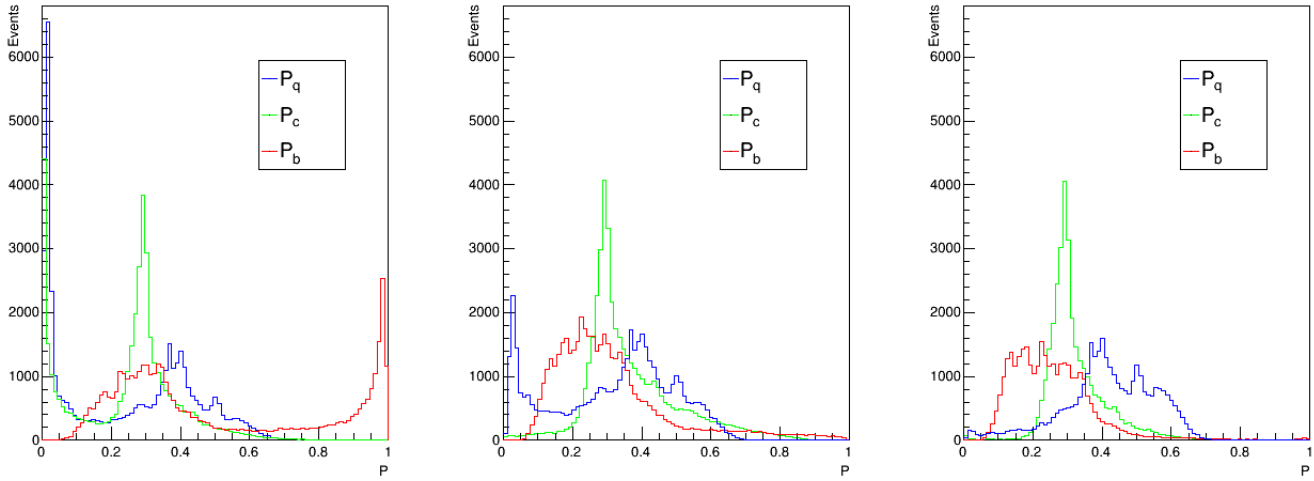

Figure 6.5: Different probability distributions P_b (red), P_c (green) and P_q (blue) for $b-$ (left), $c-$ (center) and $q-$ di-jets candidates (right).

$20 < p_T < 30$, $30 < p_T < 50$ and $50 < p_T < 100$. Since QCD MC samples are considered (thus all events are di-jets candidates) the tagging efficiency (for each p_T interval) is defined as

$$\epsilon_{tag} = \frac{N_{tag}^{jet}}{N_{total}^{jet}} \tag{6.3}$$

where N_{tag}^{jet} is the number of tagged jets and N_{total}^{jet} is the total number of jets. Considering light q jets (light quarks and gluons) the DNN is asked to have a mis-identification probability equal to the one for the SV-tagging algorithm, so that the DNN algorithm is going to wrongly tag a jet as much as the SV-tagging already does. In order to achieve this condition, cuts on the value of P_q should be considered; depending on the p_T interval considered the following cuts are applied:

interval [GeV]	P_q
$p_T < 20$	< 0.0445
$20 < p_T < 30$	< 0.0525
$30 < p_T < 50$	< 0.0699
$50 < p_T < 100$	< 0.1017

Table 6.2: Cuts on P_q.

Results are shown in Fig.6.6, where it is evident that the mis-identification probabilities of tagging a q jet is the same both for the DNN and the SV-tagging algorithm.

With the same cuts applied to P_q the b and c tagging efficiencies $\epsilon_{tag,b}$ and $\epsilon_{tag,c}$ are computed for each p_T interval and compared to SV-tagging efficiencies. In Fig.6.7 results for $\epsilon_{tag,b}$ and $\epsilon_{tag,c}$ are shown both for DNN and SV-tagging.

From the previous plots two aspects are clear:

- for $p_T < 20$ GeV efficiencies are lower also for the DNN algorithm. This is not a problem because in analyzing a signal the kinematical cut $p_T > 20$ GeV is considered;
- for both situations there is an increase in efficiency around $2\% - 5\%$, increasing with increasing p_T.

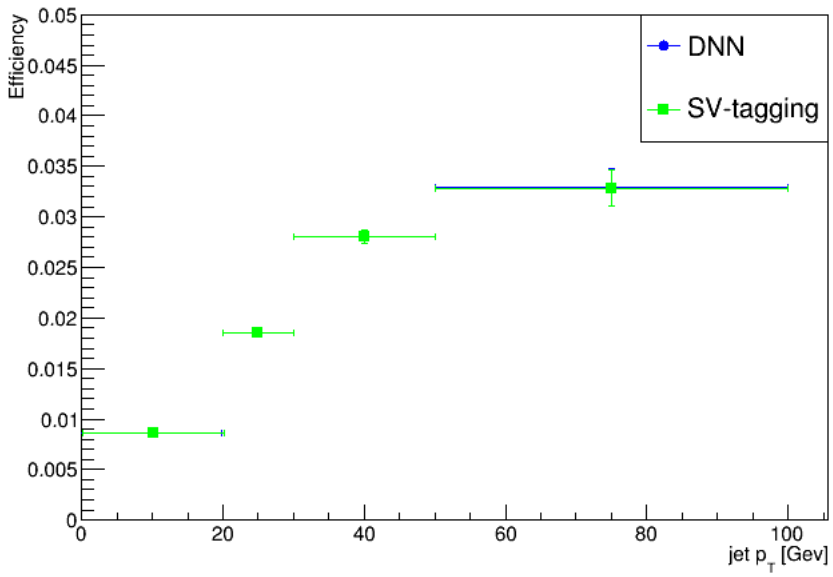

Figure 6.6: Light jets selection efficiencies applying DNN cuts and SV-tagging as a function of the jet p_T.

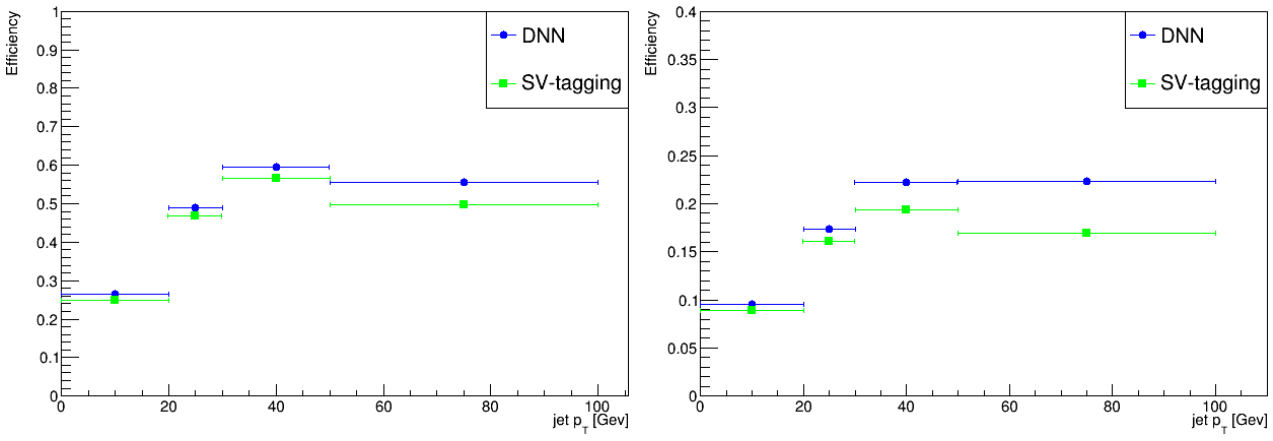

Figure 6.7: b-jet efficiencies (left) and c-jet efficiencies (right) applying DNN cuts and SV-tagging as a function of the jet p_T.

6.4.5 DNN application to signal

In Sec.6.4.3 the DNN performed pretty good in tagging $b-$jets. Since in this thesis the purpose is to study signals coming from high-mass resonances decaying to heavy flavour jets, it is possible to check if the DNN algorithm performs better than the SV-tagging algorithm when dealing with a signal over QCD background. Signals MC samples with different invariant masses m_h are considered (with $m_h = 75, 125, 175, 225$ GeV) decaying to $b\bar{b}$ final states, and the significance of a signal is computed as

$$\mathcal{S} = \frac{N_{sig}}{\sqrt{N_{bckg} + N_{sig}}} \sim \frac{N_{sig}}{\sqrt{N_{bckg}}} \tag{6.4}$$

where N_{sig} (N_{bckg}) is the number of signal (background) events in a defined mass range; symmetric mass ranges with respect to the invariant masses are considered. Particularly significance for both algorithm is maximized by putting cuts: the significance $\mathcal{S}_{SV}$ is maximized by putting cuts on the BDTs (as done in the previous chapters) while for $\mathcal{S}_{DNN}$ cuts on P_q and P_b are considered (since only two probability distributions are independent).

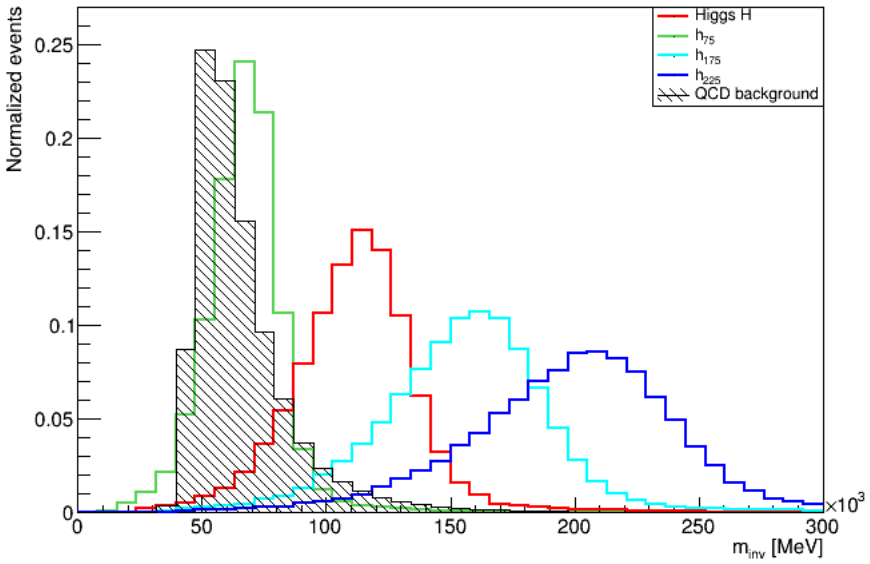

Figure 6.8: QCD MC background and signals for different resonances masses m_h.

To check the performance of the DNN with respect to the SV-tagging algorithm, the following factor can be computed for each m_h

$$\varphi = \frac{\mathcal{S}_{DNN}}{\mathcal{S}_{SV}} \tag{6.5}$$

and results are shown in Fig.6.9 with respect to the threshold value $\varphi_{threshold} = 1$.

It is clear that considering the DNN architecture instead of the usual SV-tagging algorithm for all resonances there is an improvement of about $20\% - 40\%$ in the signal significance. The next step will be the application of the DNN algorithm to data, in order to perform the same study for high-mass resonances done with the SV-tagging algorithm.

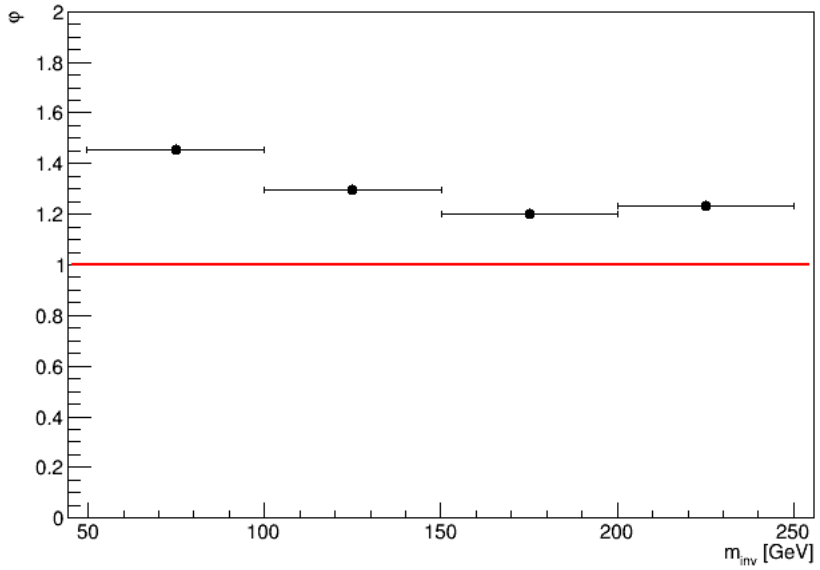

Figure 6.9: Ratio φ as a function of the resonances masses.

Chapter 7

Conclusions

In this thesis data taken by LHCb experiment during Run II campaign in 2016 have been analyzed to search for heavy flavour di-jets high-mass resonances. The total integrated luminosity is $\mathcal{L}_{int} = (200 \pm 8)$ pb^{-1}. The contribution of the QCD background is evaluated with a technique data-driven based on the identification of signal and control regions. The extrapolation of the background shape from the control to the signal region is performed with a transfer function, determined starting from the MC samples. The number of background events is a free parameter of the fit to the data together with the characteristics of the searched signal. The fitting procedure is validated by extracting the number of $Z \to b\bar{b}$ and $Z \to c\bar{c}$ decays and by comparing them to the number of expected events. The results are:

$$N_{obs}(Z \to b\bar{b}) = 22194 \pm 1575(\text{stat}) \pm 5711(\text{sys})$$
$$N_{obs}(Z \to c\bar{c}) = 152 \pm 64(\text{stat}) \, ^{+88}_{-71}(\text{sys})$$

In agreement with the SM expectations:

$$N_{exp}(Z \to b\bar{b}) = 25855 \pm 1939$$
$$N_{exp}(Z \to c\bar{c}) = 140 \pm 11$$

Heavy flavour di-jets high-mass resonances are searched using the procedure validated on the Z boson. For each search, the number of events found is compatible with zero and limit on the production cross section times the branching ratio in $b\bar{b}$ or $c\bar{c}$ is set at 95% confidence level. Here, the results for the Higgs searches:

$$\sigma(H \to b\bar{b}) = \sigma(pp \to H) \times \mathcal{BR}(H \to b\bar{b}) < 871 \text{ pb}$$
$$\sigma(H \to c\bar{c}) = \sigma(pp \to H) \times \mathcal{BR}(H \to c\bar{c}) < 61 \text{ pb}$$

The upper limit on $H \to b\bar{b}$ production cross section times the branching ratio is 29 times the SM value while the upper limit for the $H \to c\bar{c}$ process is 42 times the SM value. The improvements discussed in chapter 6 tell us that in the near future LHCb may have enough sensitivity to perform these measurements.

Bibliography

[1] J. Strandberg on behalf of ATLAS and CMS Collaborations, *Higgs and Electro-weak Physics at the LHC*, LP2019.

[2] CMS Collaboration, *Search for the standard model Higgs boson decaying to charm quarks*, CMS-PAS-HIG-18-031.

[3] CMS Collaboration, *Search for low-mass resonances decaying into bottom quark-antiquark pairs in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Phys. Rev. D 99 (2019) 012005.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. 22 (1961) 579.

[5] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, Nuovo Cim. 11 (1959) 568.

[6] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. 19 (1967) 1264.

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. 13 (1964) 321.

[8] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. 12 (1964) 132.

[9] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. 13 (1964) 508.

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. 13 (1964) 585.

[11] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. 145 (1966) 1156.

[12] T. W. B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, Phys. Rev. 155 (1967) 1554.

[13] Particle Data Group, J. Beringer et al., *Review of Particle Physics (RPP)*, Phys. Rev. D86 (2012) 010001.

[14] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. 10 (1963) 531.

[15] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. 49 (1973) 652.

[16] 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`.

[17] HL/HE WG2 group, *Higgs Physics at the HL-LHC and HE-LHC*, CERN-LPCC-2018-04 (2019), `arXiv:1902.00134`.

[18] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN-2017-002-M (2016), `arXiv:1610.07922`.

[19] CMS Collaboration, *Measurement and interpretation of differential cross sections for Higgs boson production at* $\sqrt{s} = 13$ *TeV*, Phys. Lett. B 792 (2019) 369-396.

[20] ATLAS Collaboration, *Combined measurement of differential and total cross sections in the $H \to \gamma\gamma$ and the $H \to ZZ^* \to 4l$ decay channels at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B 786 (2018) 114-133.

[21] CMS Collaboration, *Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B 791 (2019) 96.

[22] ATLAS Collaboration, *Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B 789 (2019) 508-529.

[23] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D 99, 072001.

[24] CMS Collaboration, *Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP 03 (2019) 026.

[25] ATLAS and 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.

[26] The LHCb Collaboration, *Search for $H^0 \to b\bar{b}$ or $c\bar{c}$ in association with a W or Z boson in the forward region of pp collisions*, CERN-LHCb-CONF-2016-006 (2016).

[27] Wess J., Bagger J., *Supersymmetry and supergravity*, Princeton University Press (1992).

[28] D. J. Castano, E. J. Piard, P. Ramond, *Renormalization group study of the Standard Model and its extensions. 2. The Minimal Supersymmetric Standard Model*, Phys. Rev. D49 (1994) 4882, arXiv:hep-ph/9308335.

[29] D. S. Gorbunov, V. A. Rubakov, *Introduction to the theory of the Early Universe*, World Scientific Publishing (2011).

[30] A. de Simone, T. Jacques, *Simplified Models vs. effective field theory approaches in dark matter searches*, Eur. Phys. J. C (2016) 76:367.

[31] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.

[32] ATLAS, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003.

[33] CMS, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST 3 (2008) S08004.

[34] ALICE, K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST 3 (2008) S08002.

[35] LHCb, A. A. Alves, Jr. et al., *The LHCb Detector at the LHC*, JINST 3 (2008) S08005.

[36] F. Follin, D. Jacquet, *Implementation and experience with luminosity levelling with offset beam*, 10.5170/CERN-2014-004.183 (2014), arXiv:1410.3667.

[37] LHCb Collaboration, Aaij, R. et al., *Performance of the LHCb Vertex Locator*, JINST 9 (2014) P09007.

[38] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A262 (1987) 444.

[39] LHCb, R. Aaij et al., *LHCb Detector Performance*, Int. J. Mod. Phys. A30 (2015), no. 07 1530022, arXiv:1412.6352.

[40] M. Kucharczyk et al., *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044 (2014).

[41] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A389 (1997) 81.

[42] G. Barrand et al., *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. 140 (2001) 45.

[43] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP 05 (2006) 026, arXiv:hep-ph/0603175.

[44] T. Sjostrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. 178 (2008) 852, arXiv:0710.3820.

[45] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A462 (2001) 152

[46] Geant4 collaboration, S. Agostinelli et al., *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. A506 (2003) 250.

[47] Geant4 collaboration, J. Allison et al., *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. 53 (2006) 270.

[48] M. Clemencic et al., *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. 331 (2011) 032023.

[49] G. Corti et al., *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. 53 (2006) 1323.

[50] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_T jet clustering algorithm*, JHEP 04 (2008) 063, arXiv:0802.1189.

[51] LHCb Collaboration, 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.

[52] LHCb Collaboration, R. Aaij et al., *Identification of beauty and charm quark jets at LHCb*, JINST 10 (2015), no. 06 P06013, arXiv:1504.07670.

[53] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[54] 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.

[55] 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.

[56] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf C0303241 (2003) MOLT007, arXiv:physics/0306116, [,186(2003)].

[57] F. James and M. Winkler, *C++ MINUIT User's Guide.*

[58] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. A434 (1999) 435, arXiv:hep-ex/9902006.

[59] Cowan G., Cranmer K., Gross E. and Vitells O., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C71 (2011), arXiv:1007.1727.

[60] Read A. L., *Presentation of search results: The CL(s) technique*, Advanced Statistical Techniques in Particle Physics, J. Phys. G28 (2002).

[61] LHCb Collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, LHCb-TDR-12 (2012).

[62] Karol Hennessy, *LHCb VELO upgrade*, Nucl. Instrum. Meth. A845 (2017) 0168-9002.

[63] J. Alwall et al, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, arXiv:1405.0301