

STUDY OF THE $t\bar{t}$ AND $Wb\bar{b}$ BACKGROUND IN HIGGS SEARCHES WITH THE LHCb DETECTOR

Summer Student Report

Work done during Summer 2015

Summer Student: Fabrizio Napolitano
Heidelberg Universität

Supervisor: Clara Matteuzzi
Università & INFN, Milano-Bicocca

Abstract

The increase of the center of mass energy to 13 TeV for Run 2 in the Large Hadron Collider will give the opportunity in the LHCb experiment to observe the $H \rightarrow b\bar{b}$ in the forward region, complementary to ATLAS and CMS collaborations. In this summer student project, two of the main background sources for Higgs production via *Higgs-Strahlung*, or associated production, are studied through the use of Monte Carlo POWHEG and PHYTIA, for 8 and 13 TeV, and the expected yield is calculated. The properties of the tool *Particle Flow*, present in the analyses, are investigated with respect to the definition of the jet algorithm radius.

1 Introduction

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. It first started up on 10th September 2008 to test the prediction of the Standard Model of particle physics and look for new phenomena that go beyond this theoretical framework. Since the begin of Run II on 3rd June 2015 it is running with an increased energy in the centre of mass frame of 13 TeV.

The Large Hadron Collider beauty (LHCb) is one of the seven particle physics detector experiments at LHC. It is a single arm forward spectrometer optimized for precision measurements of CP violation and rare decays of beauty and charm hadrons; the excellent performance of the detector has allowed the LHCb collaboration to publish a wide range of physics results, demonstrating LHCb's unique role, both as a heavy flavour experiment and for other general measurements in the forward region.

2 Higgs Boson

Since the final development of the Standard Model in the '70s, incorporating electroweak theory, Higgs mechanism and QCD, many efforts have been made in order to confirm its predictions. The observation of a new particle with mass of about 125 GeV and with properties consistent with those of the SM Higgs boson in July 2012, however, was mainly performed in the bosonic decay mode of the new particle; it is then of a great importance to verify its decay into fermions as well, as studied from CMS collaboration for the $H \rightarrow \tau\tau$ decay mode [1], and for the $H \rightarrow b\bar{b}$ in the ATLAS collaboration [2]. The $H \rightarrow b\bar{b}$ decay mode could be observed in LHCb in the forward region ($2 < \eta < 4.5$) hence complementary to ATLAS and CMS.

2.1 Higgs-Strahlung

Higgs boson production through gluon fusion and its subsequent decay $H \rightarrow b\bar{b}$ suffers overwhelming background if searched for in the inclusive production of $pp \rightarrow b\bar{b} + X$ via strong interaction; the usual channel considered for this decay mode is then $pp \rightarrow H + W/Z$ with $H \rightarrow b\bar{b}$ and $W^\pm \rightarrow \ell^\pm + \nu_\ell$ or $Z \rightarrow \ell^+ \ell^-$, called Higgs-Strahlung or Higgs associated production; the Feynman diagram is shown in Figure 1.

This channel has several features that allow to perform measurements for this particular decay mode of the Higgs boson, but also few drawbacks. The advantages are related to the topology of the process: the associated production of the W/Z assure a purification of the signal and rejection of the QCD background;

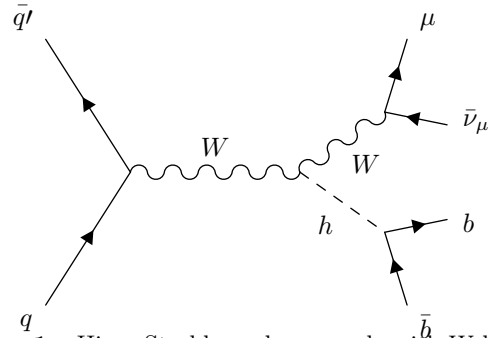


Figure 1 – Higgs-Strahlung decay mode with W boson decaying into μ and $\bar{\nu}_\mu$

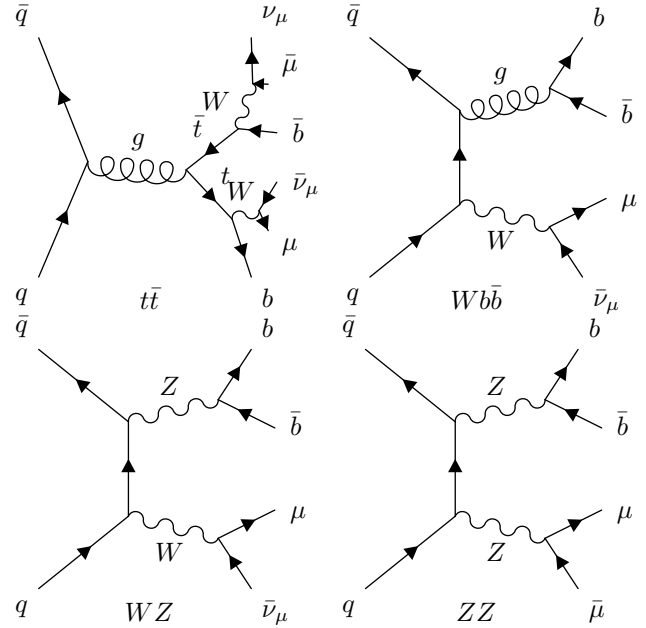


Figure 2 – Feynman diagrams for the major background sources

triggering on the lepton p_t assures high efficiency ($\sim 75\%$). Moreover, the $H \rightarrow b\bar{b}$ decay represents the most probable one, with a branching ratio of around 56%. From the detector point of view, in the forward region the LHCb detector has an excellent tracking capability and secondary vertex reconstruction alongside with the very good b-jet identification and reconstruction, essential feature of the channel taken into account.

The disadvantages are related to the decreased number of events that could be identified: the cross section for the Higgs-Strahlung is an order of magnitude lower than the gluon fusion channel. Additionally, the detector geometry is here a limiting factor, with a coverage of the solid angle which is just $\sim 16\%$ of 4π ; the luminosity delivered to the LHCb experiment is lower than the LHC designed one, in order to reduce pile-up, have less radiation damage and lower occupancy.

3 Background

3.1 Background sources

There are four major contributions to the background of the channel $pp \rightarrow W/Z + H \rightarrow \ell \bar{\nu}_\ell / \ell \bar{\ell} + b \bar{b}$; Feynman diagrams are shown in Figure 2 :

- $t\bar{t}$, with one of the W s from t decays into ℓ , reducible taking into account the extra jet activity in the event;
- $Wb\bar{b}$, with the W decays into ℓ and the $b\bar{b}$ from QCD, reducible through the constrain on the invariant mass of the system $b\bar{b}$ and the p_t in the final state;
- WZ and ZZ , with ℓ from W/Z and $b\bar{b}$ from Z decay, irreducible unless excellent mass resolution of the $b\bar{b}$ system can be achieved.

3.2 Background estimation

In order to evaluate the contribution of the various background sources, Monte Carlo samples have been used: the $t\bar{t}$ POWHEG Next To Leading Order (NLO) generator was used at 8 and 13 TeV and PYTHIA Leading Order (LO) generator was used at 8 TeV.

For the $Wb\bar{b}$ POWHEG NLO for both 13 and 8 TeV were used.

An event reconstruction was then performed with the MC samples; events were required to fulfil:

- At least two secondary vertexes (SV) for the jet algorithm and 6 charged tracks
- High transverse momentum muon: $p_t^\mu \geq 20 \text{ GeV}$ and compatibility with primary vertex
- High transverse momentum of the clustered jet: $p_t^{\text{jet}} > 25 \text{ GeV}$
- Boosted Decision Tree Gradient

Distributions of the generated MC samples were in addition studied, they are shown in Figure 3

3.3 Boosted Decision Tree Gradient

A decision tree is a machine-learning algorithm technique which combines several classifiers to create a more powerful multivariate discriminant. It is organized into several layers of nodes, the decision point in the tree in which a variable and cut value are provided and the candidate is determined to either pass or fail it; the pass or failure determines which node the candidate will encounter next. The decision tree training is a fundamental part of the multivariate analysis, using

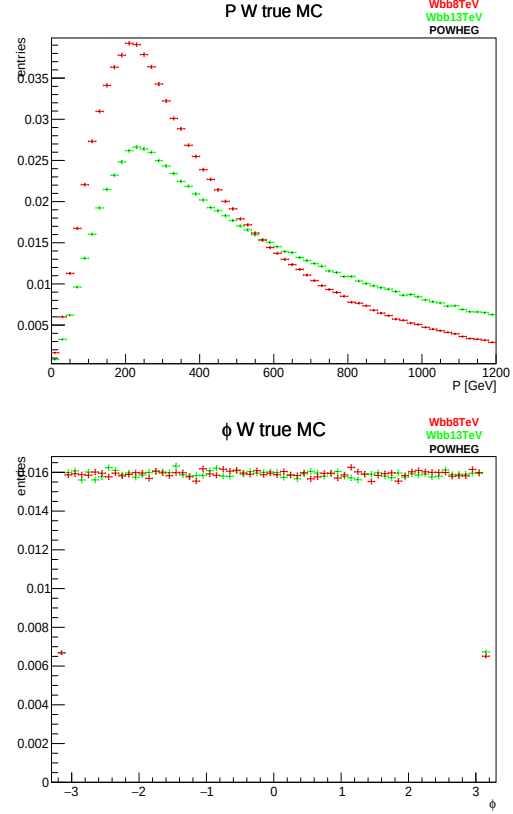


Figure 3 – Distribution of the momentum and ϕ of the Monte Carlo W boson, in red points from 13 TeV $Wb\bar{b}$ sample, in red from 8 TeV.

a set of known signal and background training events to built the structure of nodes. The boosting process uses the training results of the first tree to change the weight of the candidates that were misclassified, and the training is performed again with these new weights.

In the context of this project, the weights for the BDTG were taken from a training done using Higgs samples as signal and top-antitop as background, both generated using PYTHIA (LO) @ 8 TeV; correspondingly, the response variable is shown in Figure 4 for the Higgs and background. The input variables from the new POWHEG (NLO) ntuples, were then compared to the previous ones for both $t\bar{t}$ & $Wb\bar{b}$, and the analogue behaviour of the kinematic distributions taken into account was used to justify the use of the weights from the PYTHIA training. The classifier is hence performing in a suboptimal way, and a new training is needed to take into account the new energy, and differences in the Monte Carlo.

Several kinematic distributions were taken into account from the BDTG, some of which are shown in Figure 5:

- p_t of the lepton of highest transverse momentum,
- p_t of the Underlying Event (UE),
- Mass of the recoiling jets,

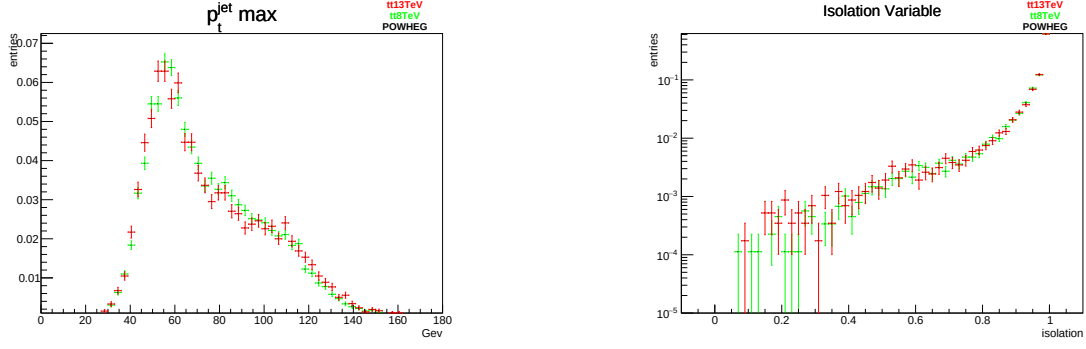


Figure 5 – Example of kinematic distributions taken into account from the BDTG, in red 13 TeV and in green 8 TeV, for the $t\bar{t}$ samples

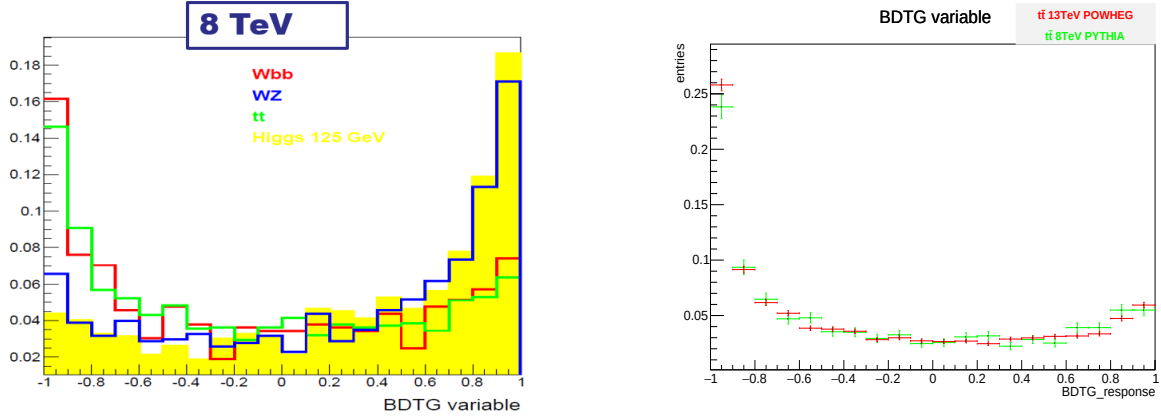


Figure 4 – Response of the BDTG for the H and $t\bar{t}$ samples with PYTHIA @ 8 TeV

Figure 6 – Response of the BDTG to the kinematic input, in red 13 TeV and in green 8 TeV, for the $t\bar{t}$ samples

- Difference of the p_t of the two leading jets,
- Difference of the ϕ of the two leading jets,
- Calibrated p_t of the jet,
- Mass of the combined b-jet and UE,
- Rapidity of the jet,
- Total p_t of the two-jet system,
- p_t of the leading jet,
- Separation in θ of the two leading jets,
- Isolation value of the lepton.

For some of them the logarithm of the distribution was taken.

The response of the variable is shown in Figure 6. The Higgs *signal region* was define for values bigger than 0.4 in the BDTG variable response (see Figure 4), where the decision tree is able to distinguish the Higgs event, and the *background region*, with BDTG variable response smaller than 0, in which the the events selected are dominated by the $t\bar{t}$ or $Wb\bar{b}$ events. Applying the requirement for the Higgs signal region for $t\bar{t}$ ($Wb\bar{b}$) samples will give us the indication of the events misidentified as Higgs events and hence representing

the background expected yield. Selecting events in the background region for the $t\bar{t}$ ($Wb\bar{b}$) samples can also be used for a measurement of the $\sigma(t\bar{t})$ ($\sigma(Wb\bar{b})$) and thus validate the estimated background for the Higgs.

3.4 Efficiencies & Expected Yield

Starting from the various selection, an efficiency table was produced taking into account the requirement imposed on each event, as listed in the previous section. The efficiency table, together with the expected yield, is listed on Table 7 for the signal region and for the background region of $t\bar{t}$ sample and in Table 8 for signal and background region of $Wb\bar{b}$ sample. In order to calculate the expected yield per fb^{-1} , the following formula was used:

$$N^{exp}/fb^{-1} = \epsilon_{MC\ gen} \times \epsilon_{sel} \times \sigma$$

Where $\epsilon_{MC\ gen}$ is the efficiency from the Monte Carlo generator, which also takes into account the detector geometry and ϵ_{sel} is the efficiency of the selection; the cross section, taken from [3] and [4] was multiplied for the probability that at least one muon is present in the

Efficiency table	13TeV [%]		8TeV [%]	
Region	Higgs	$t\bar{t}$	Higgs	$t\bar{t}$
$p_t^\mu \geq 20\text{GeV}$ Secondary Vertexes conditions	22.9		5.0	
Jet selection criteria	7.3		15.8	
$p_t^{jet} \geq 25\text{GeV}$	84.6		85.7	
BDTG variable cut	14.11	41.16	14.9	38.3
Total efficiency	0.2	0.6	0.1	0.3
$\sigma_{t\bar{t}}^{\geq 1\mu} [pb]$	$163.38^{+9.11}_{-9.98}$		$55.01^{+3.01}_{-3.19}$	
$N^{expected}$ per fb^{-1}	$13.2^{+0.7}_{-0.8}$	$38.5^{+2.1}_{-2.4}$	$1.3^{+0.1}_{-0.1}$	$3.4^{+0.2}_{-0.2}$

Figure 7 – Efficiency Table from the selection for $t\bar{t}$ samples in the Higgs signal region and $t\bar{t}$ signal region, for 13 and 8 TeV; trigger efficiency not included

final state.

Hence, from the last line of Table 7, we can see that the expected yield of the $t\bar{t}$ events in the Higgs signal region increases of an order of magnitude for 13 TeV with respect to 8 TeV; this is caused mainly by the great increase in the production cross section of top and antitop as well as an overall decreased efficiency of the selection. Similar behaviour is observed in the background region. For what concerns the $Wb\bar{b}$ background, the increase in the expected yield is not so extreme as in the $t\bar{t}$ case, this because the cross section increase is below a factor of two, and also because the efficiency of the selection remains much more similar at higher energies. Errors on the expected yields are obtained from the cross section only, and the efficiency of the trigger ($\sim 75\%$) is not considered.

4 Particle Flow

The Particle-Flow event reconstruction is an important tool used by CMS in the first place and officially released for LHCb too; it aims at reconstructing and identifying the neutrals in the event, i.e., π^0 , K^0 , Λ , Σ , with a combination of the LHCb sub-detectors, in order to optimally determine their direction, energy and type.

4.1 Invariant two-jet mass

To validate the behaviour of this tool, WZ (with $W \rightarrow \mu\nu_\mu$ and $Z \rightarrow b\bar{b}$) Monte Carlo samples @ 7 TeV from POWHEG (NLO) generator were analysed. The events were required to fulfil similar requirements as the ones described in the subsection 3.1; the calibration for the jets was, on the contrary, skipped.

Several variables were taken into account for the study, in particular the *two-jet mass*, the invariant mass of the system $b\bar{b}$, was analysed as a function of the radius of the jet algorithm (cone algorithm was chosen), with the radius R defined as follows:

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

Seven definitions of the radius were taken into account: from 0.2 to 0.8 with steps of 0.1. For each of these values the distribution of the invariant mass was fitted to a Gaussian function, the mean of the fit was plotted against the R ; the plots are shown in Figure 9. On the right plot the value of the true Z boson mass is represented by the red horizontal line at $\simeq 91$ GeV; hence the correct value of the radius definition required in order to reconstruct the mass of the Z boson lies between $R = 0.3$ and $R = 0.4$. Further studies are needed then to validate the use of the Particle Flow tool, especially also taking into account the calibration of the jets which was skipped.

5 Conclusions and outlook

In conclusion: the expected yield of the background for the $t\bar{t}$ and $Wb\bar{b}$ and its increase with the energy from 8 TeV to 13 TeV was studied in POWHEG Monte Carlo samples;

for the $t\bar{t}$ the number of expected event per inverse femtobarn increases from $1.3^{+0.1}_{-0.1}$ to $13.2^{+0.7}_{-0.8}$ in the Higgs signal region and from $3.4^{+0.2}_{-0.2}$ to $38.5^{+2.1}_{-2.4}$ for the background region.

For the $Wb\bar{b}$, the number of expected event per inverse femtobarn increases from $2.6^{+0.6}_{-0.5}$ to $5.6^{+1.3}_{-1.1}$ for the Higgs signal region and from $3.5^{+0.8}_{-0.7}$ to $12.7^{+2.9}_{-2.5}$ for the background region. A new training is needed for the BDTG in order to correct for the differences at the Monte Carlo generator and to correct for the higher energy; for a quantitative statement on the Higgs expected yield at 13 TeV, new Monte Carlo simulations are needed.

For the Particle Flow tool, the dependency of the two-jet mass variable (b-jet coming from the decay of the Z boson) was studied as a function of the definition of the radius of the jet algorithm; the mass of the Z boson mass was found to be reconstructed between $R = 0.3$ and $R = 0.4$, although a new study is needed to take into account the jet calibration in order to find the optimal definition of R .

Efficiency table	13TeV [%]		8TeV [%]	
Region	Higgs	$Wb\bar{b}$	Higgs	$Wb\bar{b}$
$p_t^\mu \geq 20\text{GeV}$ Secondary Vertexes conditions	25.0		22.3	
Jet selection criteria	1.8		1.7	
$p_t^{jet} \geq 25\text{GeV}$	59.0		53.7	
BDTG variable cut	22.0	48.7	30.5	40.3
Total efficiency	0.06	0.13	0.06	0.08
$\sigma_{Wb\bar{b}}^{>1\mu}[pb]$	$50.31^{+11.73}_{-9.94}$		$28.22^{+6.55}_{-5.50}$	
$N^{expected}$ per fb^{-1}	$5.6^{+1.3}_{-1.1}$	$12.7^{+2.9}_{-2.5}$	$2.6^{+0.6}_{-0.5}$	$3.5^{+0.8}_{-0.7}$

Figure 8 – Efficiency Table from the selection for $Wb\bar{b}$ samples in the Higgs signal region and $Wb\bar{b}$ signal region, for 13 and 8 TeV; trigger efficiency not included

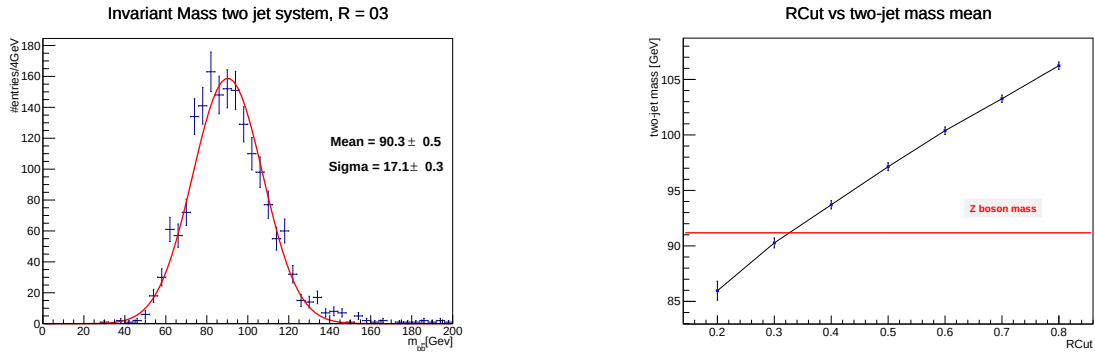


Figure 9 – The left plot shows the Invariant mass distribution of the two-jet system for the radius definition of the jet algorithm $R = 0.3$; the mean and the sigma for the fit are shown on the plot. The right plot shows the dependency of the mean value from the fit of the two-jet mass distribution with respect to seven definitions of the radius R .

References

- [1] CMS Collaboration, *Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons*, JHEP 05 (2014) 104, [arXiv:1401.5041].
- [2] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector* [arXiv:1409.6212v2].
- [3] Twiki page, *NNLO+NNLL top-quark-pair cross sections*, [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO].
- [4] CERN-THESIS, *Recherche directe à LHCb du boson de Higgs en production associée et contribution au développement du nouveau trajectographe à fibres scintillantes* Pierre Jaton, Lausanne, EPFL, 2013.
- [5] LHCb Collaboration, *Study of W boson production in association with beauty and charm*, [arXiv:1505.04051v1].
- [6] LHCb Collaboration, *First observation of top quark production in the forward region*, [arXiv:1506.00903v1].