Search for the standard model Higgs boson produced through vector boson fusion and decaying to bottom quarks with the CMS experiment at 13 TeV.

Nadezda Chernyavskaya[1,2,2]

[1] Institute for Theoretical and Experimental Physics, Russia

[2]Moscow Institute of Physics and Technology, Russia

[3] National Research Nuclear University MEPhI, Russia

Supervisors:

Paolo Azzurri

August 18, 2015

Abstract

This report discusses the preparation for the search for the standard model Higgs boson produced through vector boson fusion and decaying to bottom quarks with the CMS experiment at 13 TeV in the LHC Run II. The analysis strategy, preselection and the set of discriminating variables are discussed. A new discriminating variable $\Delta\eta_{bq}$ with a good separation power is proposed. The efficiency of correctly finding the b-jets and q-jets in signal events is increased by developing and implementing a jet b-likelihood and q-likelihood.

Contents

1 Motivation

The Higgs boson predicted by the standard model was discovered at the LHC in 2012 at the mass of about 125 GeV. The main efforts are now to measure its properties. At the LHC the Higgs boson can be produced by several mechanisms one of which is Vector Boson Fusion(VBF) that has the second largest predicted cross-section following gluon-gluon fusion production. Investigation of this mechanism of production is crucial to check the unitarity of the standard model. In the standard model couplings of the Higgs boson to the vector bosons like W and Z are fully specified thus they are needed to be checked in experiment. Next, the branching ratio of the decay of the Higgs boson to a pair of bottom quarks is expected to be the largest at the boson mass of 125 GeV and very important to be measured because it allows to establish the couplings to fermions and down-type quarks. Fig. 1 shown the Feynman diagram for this process. Although the search for the VBF $H \to b\bar{b}$ is the fully hadronic search and thus very challenging, the CMS collaboration has performed such a search at LHC Run 1[1]. The signal strength obtained in the CMS Run I analysis is $\mu = \sigma/\sigma_{SM} = 2.8 \pm 1.5$ with significance of 2.2σ. The Run I analysis shows that such search is possible and the significance of observation might be reached in the LHC Run II.

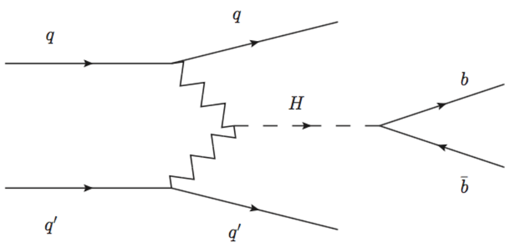


Figure 1: The Feynman diagram of the Vector Boson Fusion process with the Higgs boson decaying to bottom quarks.

In the VBF $H \to b\bar{b}$ process four energetic jets are present : two q-jets from the colliding protons and thus being forward jets and two central b-jets originating from the Higgs boson decay. The essential feature is the rapidity gap in the central region due to no exchange of a color singlet at leading order. This means that there are almost no other particles besides signal four jets. This feature is used to discriminate signal from background(BG). The main background is QCD, although there are some other backgrounds as hadronic decays of Z, W bosons + 2 jets and hadronic decays of top-quarks that have to be considered.

2 Analysis strategy

The analysis is data driven. Since the BG is huge one needs a special dedicated trigger. Then an offline selection that follows the trigger requirements is to be applied. Among all the variables that do not exploit b-jets kinematics the ones with the best signal from BG separation power are chosen to train a Multi Variate Analysis(MVA) to separate signal and BG. Events are classified with the MVA output and fits of the invariant mass of two b-jets are performed to extract the signal presence.
For the Run II VBF analysis two dedicated HLT trigger paths were designed : DoubleBTag and SingleBTag. Both of the trigger paths require 4 jets. The DoubleBTag trigger requires two online b-jets, whilst the SingleBTag trigger only one. Both of the triggers have different cuts on the invariant mass of two q-jets, the threshold is higher for the SingleBTag trigger. The events passing these two triggers are splitted in two different datasets and each of the datasets are treated separately. The table 1 summarizes the preselection applied for the two datasets.

Table 1: Preselection

cuts	**DoubleBTag**	**SingleBTag**
# b-jets	2	1
b-tag	> 0.4	> 0.7
m_{qq}	200 GeV	460 GeV
$\Delta\eta_{qq}$	> 1.2	> 4.1
$\Delta\phi_{bb}$	< 2.4	< 1.6

Monte Carlo samples are produced with the assumption of 20 pile-up interactions and a 25 ns bunch crossing. The MC samples of the VBF signal are produced by Pythia8 together with two MC generators : AMC@NLO and powheg with the Higgs boson mass at the value of 125 GeV and including the branching ratio of $H \to bb$(BR$= 0.58$). The table 2 summarizes the number of events left for signal and the QCD BG samples after the preselection and trigger requirements. The signal efficiency for DoubleBTag is 2.5 % and for SingleBTag is 0.8%. For 1 fb^{-1}we expect to collect $\sim$ 50 signal events after DoubleBtag trigger requirement and an offline preselection.

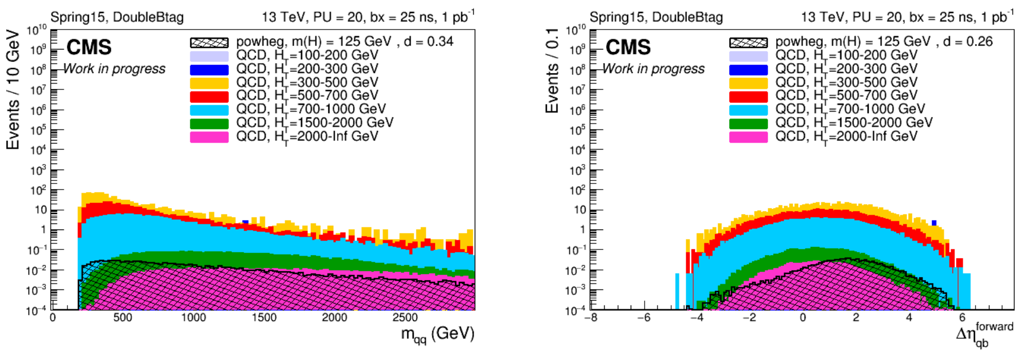
As mentioned at the beginning of Section 2 we need to find the most discriminating variables that we can further use to train the MVA to separate signal from the BG. The separation power can be evaluated using the shape differences of the distributions, as

$$d = \frac{1}{2}\Sigma|s - b|$$

Table 2: Summary for 40 pb^{-1}

Sample	yield DoubleBTag	yield SingleBTag
AMC@NLO, m_H = 125 GeV	2 .14	0.64
Powheg, m_H = 125 GeV	2 .12	0.70
QCD, H_T=100-200 GeV	2689	0
QCD, H_T=200-300 GeV	2700	624
QCD, H_T=300-500 GeV	33885	6121
QCD, H_T=500-700 GeV	14975	1607
QCD, H_T=700-1000 GeV	5570	392
QCD, H_T=1500-2000 GeV	158	5
QCD, H_T=2000-Inf GeV	31	0
Total BG	60008	8749

where s and b are signal and BG distributions. If the distributions are completely separated d = 1, if they are identical d = 0. When $d > 0.2$ the variable is considered to have a good separation power. Fig. 2 (a) shows the invariant mass of two q-jets, the separation power is very good for this variable and equals d = 0.34. I propose to use a new discriminating variable that hasn't been used in the Run I analysis. Since VBF $H \to b\bar{b}$ signature has two forward q-jets and two central b-jets, one can plot difference in pseudorapidity between the forward q-jet and the closest b-jet $\Delta\eta_{qb}^{forward}$ and the same for the backward q-jet $\Delta\eta_{qb}^{backward}$. The new discriminating variable together with signal and BG is shown in Fig. 2 (b). As one can see due to asymmetric behaviour for the signal and rather symmetric for BG this variable has a good separation power and can be used for further signal/BG separation.



(a) (b)

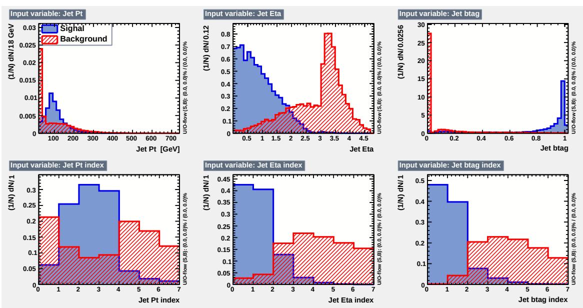
Figure 2: (a) - The invariant mass of two q-jets. m_{qq} has a good separation power d = 0.34 ; (b) - The difference in pseudorapidity between the forward q-jet and the closest to it b-jet $\Delta\eta_{qb}^{forward}$, new discriminating variable with a separation power d = 0.26

3 Jet b-likelihood, q-likelihood

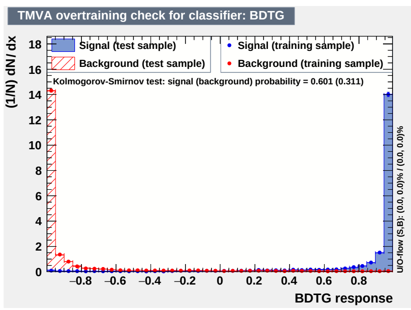
Since the goal of the analysis is to reconstruct the Higgs boson from two b-jets, it is essential to make sure that we choose the b-jets originating from the Higgs boson decay. One way to match the b-jets is to choose the b-jets with the highest value of combined secondary vertex(CSV) b-tag. However, for the SingleBTag dataset such method of choosing is not the optimal and the efficiency of matching b-jets can be increased using a jet MVA. The idea behind it is to train a Boosted Decision Tree(BDT) to distinguish between b-jets, q-jets and other jets using the following information : b-tag, b-tag order, η, η order, p_T, p_T order. The result of such a training is a new discriminator - b-likelihood that we can use to find the b-jets. In Fig. 3(a) input BDT[2] variables are shown. The final signal and BG separation of b-likelihood is shown in Fig. 3(b).

Table 3: b-likelihood, q-likelihood

b-likelihood		
efficiency	**DoubleBTag**	**SingleBTag**
to find bb	97 %	96 %
to match bb by CSV	89 %	80 %
to match by b-likelihood	89 %	86 %
q-likelihood		
efficiency	**DoubleBTag**	**SingleBTag**
to find qq	82 %	91 %
to match qq by p_T	75 %	84 %
to match by q-likelihood	75 %	85 %



(a)



(b)

Figure 3: (a) - b-likelihood input variables in BDT for the SingleBTag dataset. (b) - Final signal and BG separation of b-likelihood for the SingleBTag dataset.

Since we have to choose correctly not only b-jets but also q-jets one can try to increase the efficiency of matching q-jets using BDT as well. However, the efficiency to reconstruct q-jets is smaller in comparison to the one for b-jets and even the BDT cannot increase it considerably. The results of the efficiency studies for b-likelihood and q-likelihood are presented in the table 3. For the SingleBTag dataset the efficiency of matching right b-jets increased from 80% to 86% and thus b-likelihood will be used in the Run II analysis to tag b-jets for SingleBTag dataset.

4 Conclusion

The preparation for the search for the standard model Higgs boson produced through vector boson fusion and decaying to bottom quarks for Run II has been started. The new discriminating variable $\Delta\eta_{bq}$ was studied and found to give a good separation power. A new b-likelihood to increase the efficiency of choosing the correct b-jets in signal events was developed and implemented.

References

[1] V. Khachatryan *et al.* [CMS Collaboration], arXiv:1506.01010 [hep-ex].

[2] A. Hocker *et al.*, PoS ACAT (2007) 040 [physics/0703039 [PHYSICS]].