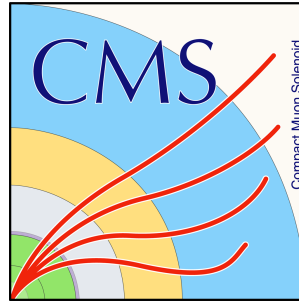


# Higgs production in association with a top anti-top quark pair with $H \rightarrow b\bar{b}$ in $\sqrt{s}=13\text{TeV}$

Sotiroulla Konstantinou

Supervised by Carmen Diez Pardos

September 10, 2016



## Abstract

A search for the standard model Higgs boson production in association with a pair of top, anti-top quarks is presented, using p-p collision data collected by the CMS detector, at  $\sqrt{s}=13\text{ TeV}$ , with luminosity  $2.7\text{ fb}^{-1}$ . The selected events contain a final state signature with four b-quark jets; a pair of b-quark jets produced from the Higgs-boson decay ( $H \rightarrow b\bar{b}$ ) and two b-quark jets, produced from the decay of the  $t\bar{t}$  pair. The transverse momentum of the  $t\bar{t}$  system on the selected events is required to be greater  $200\text{GeV}$ . The Combined Secondary Vertex (CSV) b-tagging algorithm is used to identify jets that are likely to originate from the hadronization of b quarks. Furthermore, two algorithms are used for the investigation of the substructure of the final state jets.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Final state . . . . .	3
1.2	Background processes . . . . .	4
1.3	Event reconstruction and selection . . . . .	5
1.3.1	B jets and b tagging . . . . .	5
1.3.2	Event Selection . . . . .	5
<b>2</b>	<b>Boosted Objects</b>	<b>7</b>
2.1	Fat jets . . . . .	7
2.2	Substructure of the fat jets . . . . .	8
<b>3</b>	<b>Results</b>	<b>13</b>
<b>4</b>	<b>Conclusion</b>	<b>14</b>
<b>5</b>	<b>References</b>	<b>15</b>

# 1 Introduction

Since the discovery of a new boson in July 2012 by the ATLAS and CMS collaborations, the experimental studies have focused on the investigation of its properties. All the measured properties of the new particle (spin, parity) are consistent with the Higgs boson that is predicted by the Standard Model. It is of particular interest to measure the coupling of the Higgs boson to the top quark because the top quark could play a special role in the context of electroweak symmetry breaking due to its large mass. Since the top quark has a very large mass, the Higgs boson does not decay to top quarks. However, the Yukawa coupling can be directly measured at the process of Higgs production in association with a top, anti-top quark pair ( $t\bar{t}H$ ). The cross section of the  $t\bar{t}H$  production in  $\sqrt{s} = 13$  TeV and  $M_H = 125\text{GeV}$  is  $\sigma = 0.5\text{pb}$  and the process has not been observed yet. The small cross section of the process makes the observation very challenging.

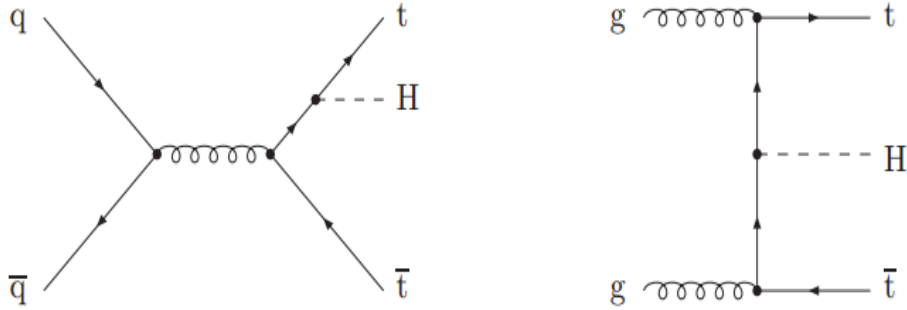


Figure 1: Higgs boson production in association with a  $t\bar{t}$  pair

## 1.1 Final state

The Higgs boson and the top quark are unstable particles, so they decay into other more stable particles very fast. Therefore, the detector can only give us information about the final state of the events that can be used for the identification of the signal. The figure 2 shows the decay modes of the Higgs boson. In the  $m_H = 125\text{GeV}$  region the dominant channel is the decay of the boson into a  $b\bar{b}$  pair with branching ratio  $\sim 57.8\%$ . The top quark decays with nearly 100% probability to a W boson and a bottom quark, so the decay of the top quark is determined by the decay of the W boson. The leptonic decay of the W has a low cross section; however it provides a clean final state. In this project the final state of the  $t\bar{t}H$  process with four b-quark jets and two opposite signed isolated leptons with high  $p_T$  will be investigated.

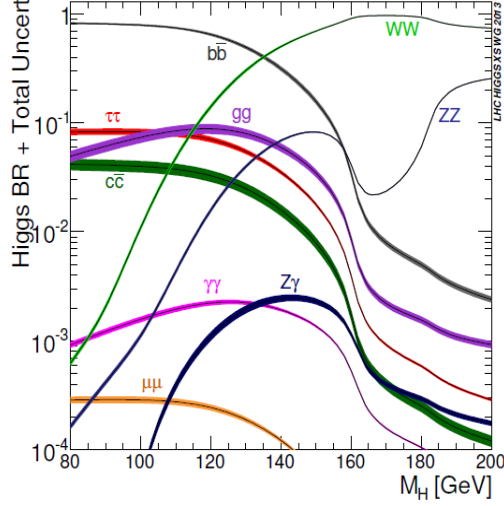


Figure 2: Higgs boson decay modes

## 1.2 Background processes

Apart from the very low cross section of  $t\bar{t}H$ , the observation of the process becomes very difficult because of the large cross section of some background processes that mimic the final state of the signal. The most important background is the production of  $t\bar{t}$  with jet radiation which has a cross section about 2000 times larger. The Figure 3 shows the signal process and the production of a  $t\bar{t}$  pair with the emission of a gluon that splits into a  $b\bar{b}$  pair. As can be shown, the final state of this process is the same as the final state of the signal but the process of  $t\bar{t} + b\bar{b}$  has a very large cross section compared to the  $t\bar{t}H$  production.

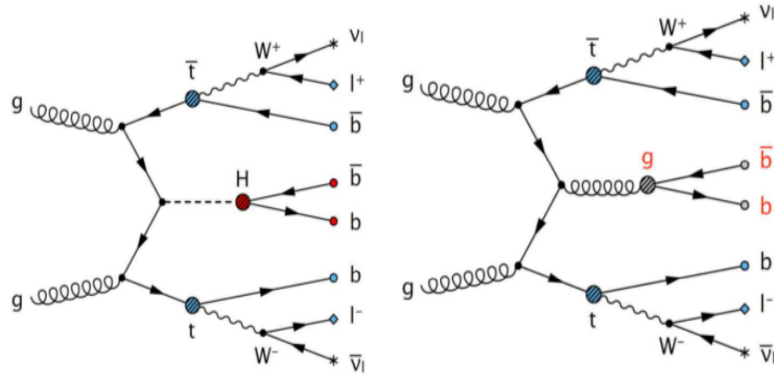


Figure 3: Feynman diagrams of  $t\bar{t}H$  and  $t\bar{t} + b\bar{b}$  production. The two processes have the same final state

## 1.3 Event reconstruction and selection

### 1.3.1 B jets and b tagging

The identification of the events with  $t\bar{t}H$  production requires the discrimination of jets that are produced by the hadronization of the b hadrons, the so called b-jets, from jets that come from light quarks and gluons [3]. The discrimination can be achieved with the use of b-tagging algorithms. B tagging is a reconstruction technique that takes advantage of the b hadron properties and assigns to each jet a "likelihood" that it contains a b hadron. The algorithm used in this project is the Combined Secondary Vertex (CSV) algorithm that utilizes information about the impact parameter of tracks and reconstructed secondary vertices within the jets in a multivariate algorithm. The CSV algorithm provides a continuous output ranging from 0 to 1; the higher the discriminator value, the more likely it is that the jet comes from the decay of a B hadron.

### 1.3.2 Event Selection

The selected events that are considered suitable for the analysis, should fulfill the criteria that are listed below:

- The event should pass the dilepton trigger specified by the analysis channel
- The presence of two opposite signed leptons is required
- The invariant mass of the dilepton system should be greater than 20GeV
- The invariant mass of the dilepton system should not be between 76 and 106 GeV. This cut excludes the Z window for same-flavour channels
- The presence of at least two jets with  $p_T$  greater than a given threshold
- The missing transverse energy of the event should be greater than 40GeV
- At least one b-tagged jet is required in the events

The figure 4 shows the distribution of the transverse momentum  $p_T$  of the  $t\bar{t}$ . As can be shown in the plot, the distribution for the signal is shifted to the region of highest values of  $p_T$  compared to the distribution of the background processes. In the figure 4, the signal is multiplied by a factor of 1985 – so the background is  $\sim 2000$  times larger than the signal.

The correlation between the  $p_T$  of the  $t\bar{t}$  system and the Higgs boson is shown in the figure 5. The plot shows that the two variables are linear correlated. This correlation is a consequence of the energy and momentum conservation.

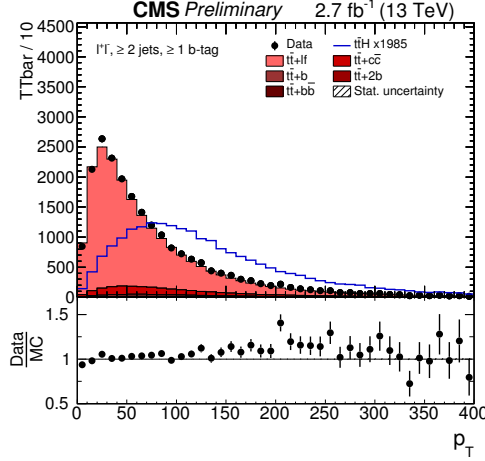


Figure 4: Transverse momentum of the  $t\bar{t}$  system

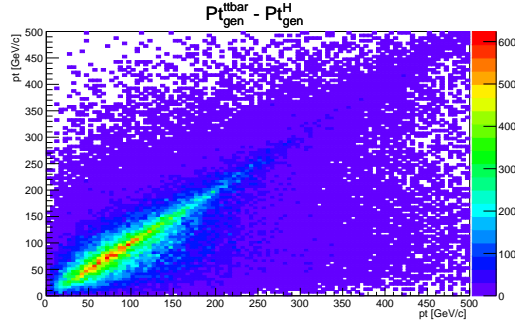


Figure 5: Correlation between the transverse momentum of the  $t\bar{t}$  system and the Higgs boson in generator level

The events that fulfill the criteria that have been listed above can be divided in different categories based on the number of jets and bjets. The table 1 shows the number of events in Data, signal and background for the different categories

Table 1: Events in categories of jets and b-tagged jets

Sample	2J, 1bTag	3J, 2bTags	3J, 3bTags	$\geq 4$ J, 2bTags	$\geq 4$ J, 3bTags	$\geq 4$ J, 4bTags	Boosted
<i>Data</i>	21768	3017	110	2852	308	27	1634
$t\bar{t}H$	26.1	1.3	0.4	8.8	4.0	1.2	5.7
Total Backgr	25374.0	3257.6	91.3	3647.9	310.2	26.3	1767.3
$signal/bckg(x10^{-2})$	0.10	0.04	0.44	0.24	1.29	4.10	0.32

As the table shows, the signal-over-background ratio increases for the categories with tighter requirements on the number of jets and b-jets. The last column on the table

shows the number of events for the boosted regime; the events in which the transverse momentum of the  $t\bar{t}$  system is greater than 200 GeV which indicates that the Higgs boson has also large  $p_T$ . The significance of the category is not very high, although it could be helpful to the discrimination between signal and background. Further studies on this category will be studied in this project.

## 2 Boosted Objects

The boosted objects, produced in the detector, pass their momentum to their decay products. Therefore, the jets are emitted with large  $p_T$  and in a small distance  $\Delta R$  in the  $\eta - \phi$  plane in the case of the decay of a Higgs boson. The distance  $\Delta R$  is defined as:

$$\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (1)$$

Therefore, the standard algorithms are not valid to reconstruct them, so jets are reconstructed with a larger  $\Delta R$

### 2.1 Fat jets

Because of their small distance, the jets sometimes are reconstructed to one big jet, the so called fat jet, in a cone of  $\Delta R = 1.5$ . The figure 6 shows the number of the reconstructed fat jets in the events.

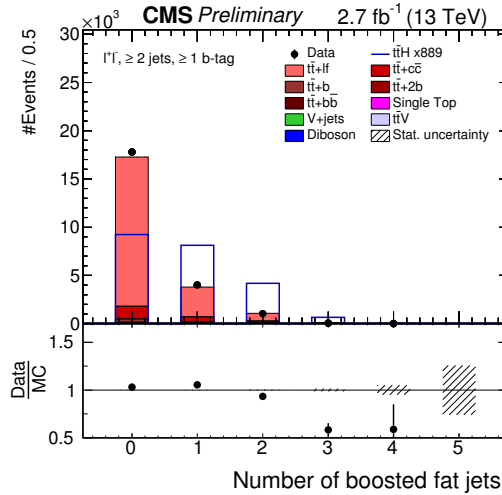


Figure 6: Number of reconstructed "fat jets"

The signal-to-background ratio increases in the events where there is at least one fat jet. The distribution of the signal is multiplied by a factor of 889. The figures 7 and 8 show the distribution of transverse momentum  $p_T$  and pseudorapidity  $\eta$  of the fat jets respectively. As shown in the plots, the fat jets have high  $p_T$  and in the signal events, they tend to be emitted in the region of small values of  $\eta$ .

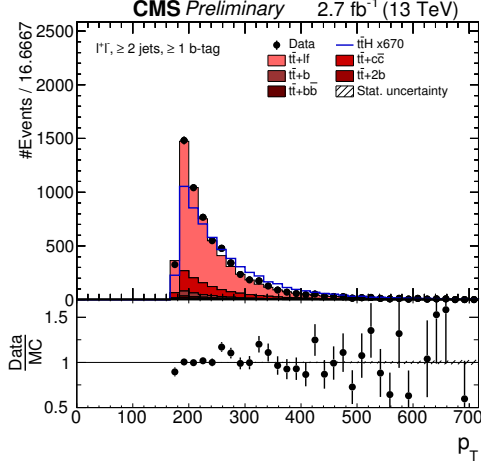


Figure 7: Transverse momentum of Fat jets

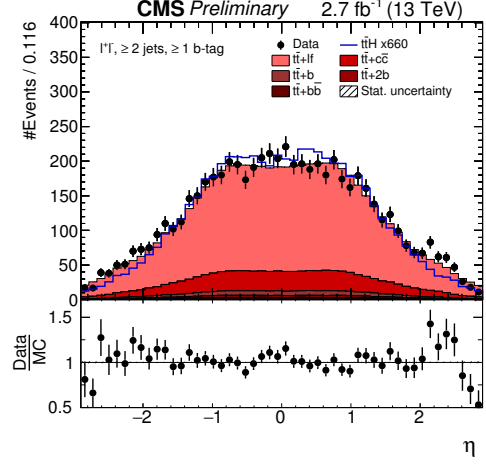


Figure 8: Pseudorapidity  $\eta$  of Fat jets

## 2.2 Substructure of the fat jets

In order to identify if the fat jet corresponds to the decay product of a Higgs boson, it is important to study the substructure of the fat jets. In this project, two algorithms have been used which look for the hard jets inside the main fat jet and remove soft radiation from pile up events and initial state radiation.

### Subjet-Filter jet algorithm

The Subjet-Filter Jet algorithm declusters the fat jet with mass  $m$  into two subjets with masses  $m_1$  and  $m_2$ ,  $m_1 > m_2$  [1]. If the fraction  $m_1/m$  is greater than 0.67 then the subjet is divided into two subjets etc, otherwise the declustering stops. The declustering stops when the distance of two jets becomes less than a threshold  $r_{cut}$  or when the softest subjet fulfills the condition

$$z = \frac{\min(p_T^i, p_T^j)}{p_T^{i+j}} = z_{cut} \quad (2)$$

### Soft Drop jets declustering

The Soft Drop jets algorithm divides the fat jet  $j$  with radius  $R_o$  into two subjets  $j_1$  and



$j_2$  with transverse momentum  $p_T^1$  and  $p_T^2$  respectively, where  $p_T^1 > p_T^2$  [1]. If the  $p_T$  of the two subjets fulfill the condition

$$\frac{\min(P_T^1, P_T^2)}{P_T^1 + P_T^2} > z_{cut} \left( \frac{\Delta R_{12}}{R_o} \right)^\beta \quad (3)$$

where  $\Delta R_{ij}$  is the distance between the two subjets, then the  $j$  jet is the final soft drop jet, otherwise the  $j_1$  is defined as the fat jet and the procedure is iterated. The Soft Drop algorithm depends on two parameters – a soft threshold  $z_{cut}$  and the angular exponent  $\beta$ . For the algorithm used in this project,  $z_{cut}$  is equal to 0.1 and  $\beta$  is equal to 0.0. More information about the algorithm can be found in [2].

The study of the kinematics of the subjets can be helpful for the discrimination between the signal and the background processes. The figures 9 and 10 show the distribution of the  $\Delta R$  distance between the axis of the fat jet and the axis of the highest b-tagged jet for the Filter jet and Soft Drop algorithm respectively. The  $\Delta R$  distance between the axis of the fat jet and the axis of the second highest b-tagged jet for the two algorithms are shown at the figures 11 and 12. As can be shown from the figure 12, the distributions of the signal and background for the Soft Drop algorithm have different shape, so they can give us information to discriminate the signal from the background. In general all distributions are well described by Monte Carlo.

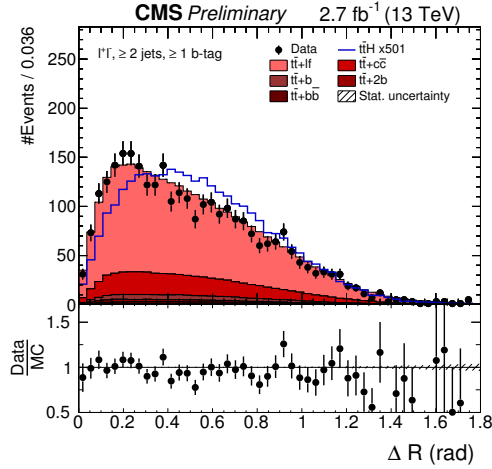


Figure 9: Distribution of  $\Delta R$  between the highest b-tagged jet and the fat jet - Filter Jets

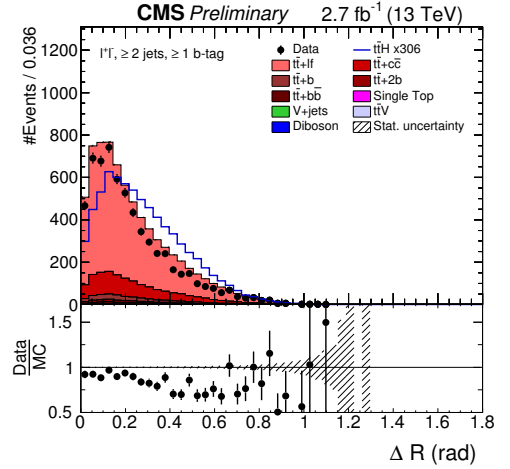


Figure 10: Distribution of  $\Delta R$  between the highest b-tagged jet and the fat jet - SoftDrop Jets

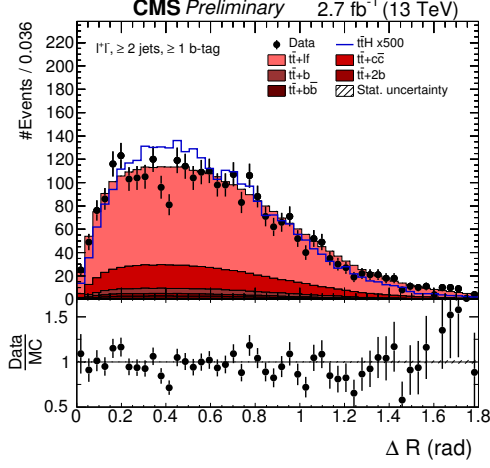


Figure 11: Distribution of  $\Delta R$  between the  $2^{nd}$  highest b-tagged jet and the fat jet - Filter Jets

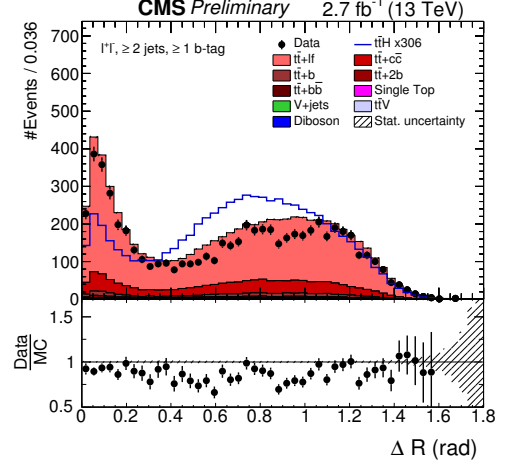


Figure 12: Distribution of  $\Delta R$  between the  $2^{nd}$  highest b-tagged jet and the fat jet - SoftDrop Jets

The figures 13 and 14 show the  $p_T$  of the highest and the second highest b-tagged subjet for the two algorithms. In the distributions, there is a peak at the 200GeV region and the signal events tend to have larger transverse momentum. Monte Carlo seems not to describe data well in the low  $p_T$  region.

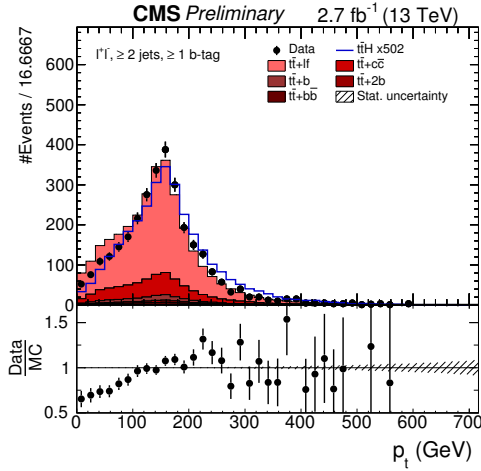


Figure 13: Distribution of the  $p_T$  of the highest and the  $2^{nd}$  highest b-tagged jet - Filter Jets

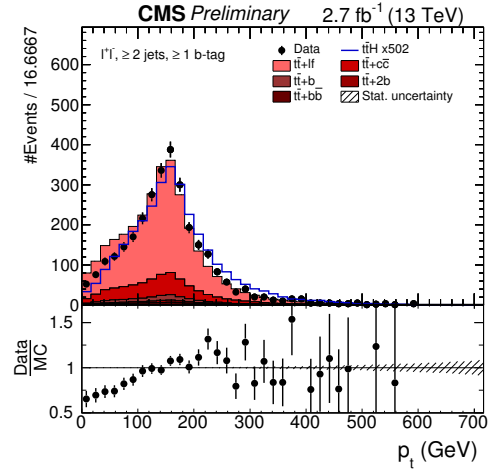


Figure 14: Distribution of the  $p_T$  of the highest and the  $2^{nd}$  highest b-tagged jet - SoftDrop Jets

The distributions of the mass of two subjects out of which one is b-tagged are shown in the figures 15 and 16 for the two algorithms. The plots 17 and 18 present the same variable for the boosted events ( $p_T^{t\bar{t}} > 200\text{GeV}$ ). While for the Soft Drop algorithm there are not large differences between the distributions of the signal and background, for the filter jet algorithm there is a peak at the 125GeV region in the signal distribution – as it is expected since the two jets are the decay products of the Higgs boson.

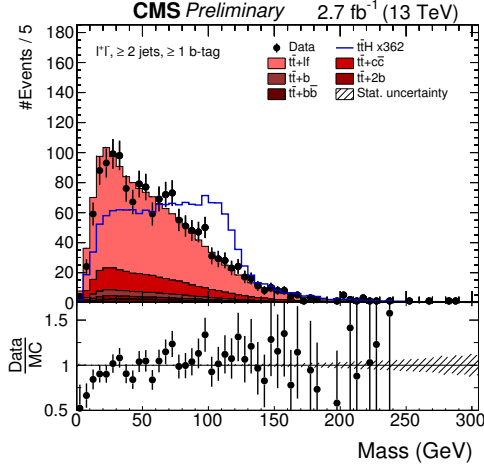


Figure 15: Distribution of the mass of two jets, one b-tagged - Filter Jets

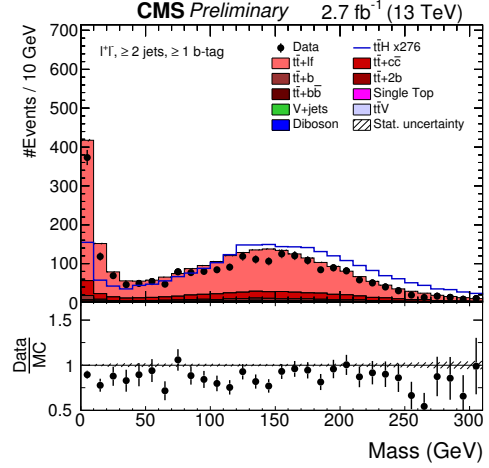


Figure 16: Distribution of the mass of two jets, one b-tagged - SoftDrop Jets

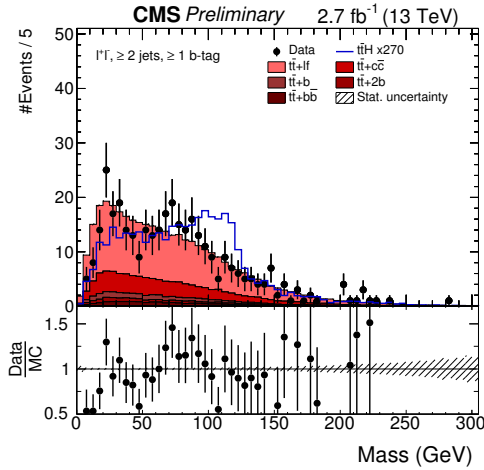


Figure 17: Distribution of the mass of two jets, one b-tagged in the boosted regime - Filter Jets

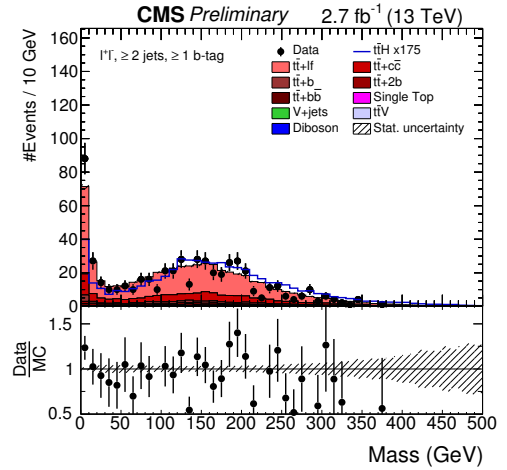


Figure 18: Distribution of the mass of two jets, one b-tagged in the boosted regime - SoftDrop Jets

An additional handle in the identification of the events with Higgs boson production is the discrimination between b-jets and jets that come from light quarks and gluons. The

figures 19 and 20 show the distribution of the CSV discriminator of the highest b-tagged jet for the two methods. For the both algorithms, there are not large differences in the distribution of the signal and background.

However, the same distribution for the second highest b-tagged jet, using the filter jet algorithm has different shape between the signal and background events especially in the case of the filter jets. Therefore, this distribution will be used to estimate limits on the production cross section

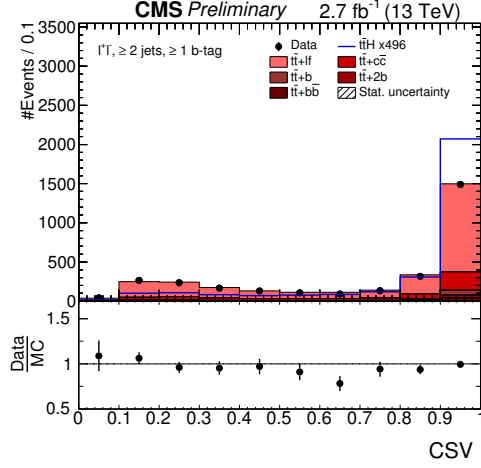


Figure 19: Distribution of the CSV discriminator of the highest b-tagged jet - Filter Jets

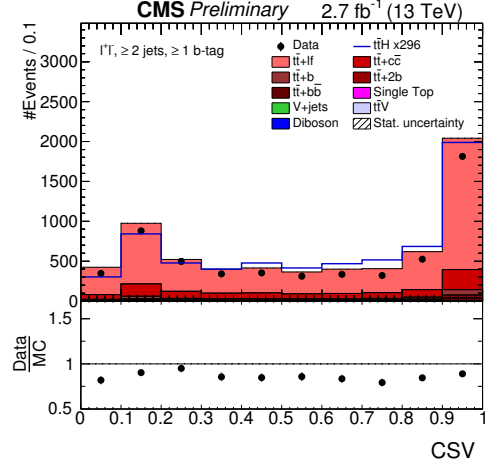


Figure 20: Distribution of the CSV discriminator of the highest b-tagged jet - SoftDrop Jets

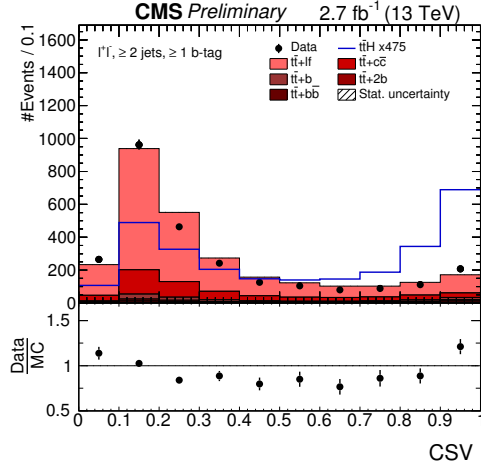


Figure 21: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet - Filter Jets

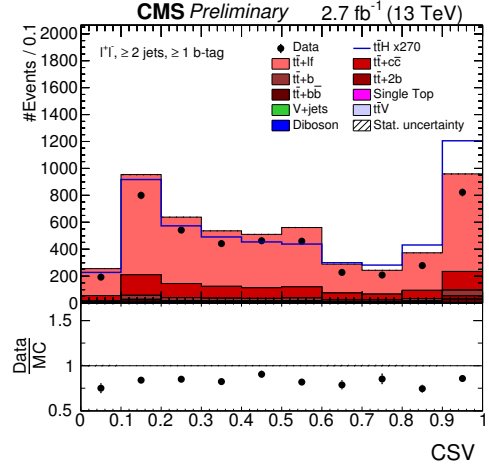


Figure 22: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet - SoftDrop Jets

### 3 Results

The distribution of the second highest b-tagged jet using the filter jet algorithm can be used for the calculation of the expected limit.

For the calculation of the expected limit, the shape and rate of the data, signal and total background processes have been used. The expected limit describes the sensitivity of a method with respect to the Standard Model. In this analysis, the expected limit has been calculated for all the events where a fat jet is reconstructed and for the boosted category only. The systematic uncertainties have not been included in the calculations.

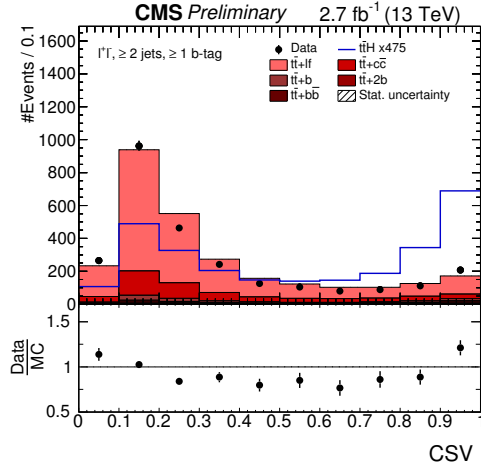


Figure 23: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet - Filter Jets

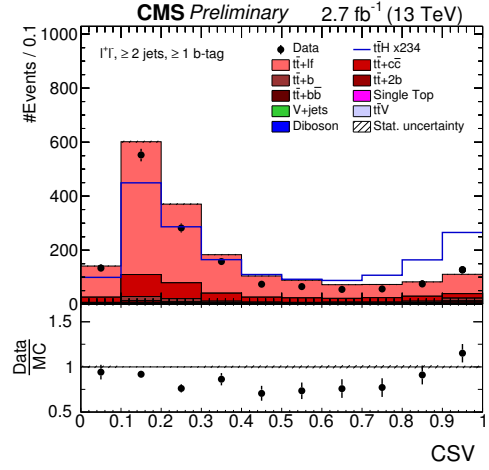


Figure 24: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet, boosted regime - Filter Jets

The results are presented in the table 2. The observed limit can be defined as the level of agreement between the measured data and the expectation (Monte Carlo).

Table 2: Observed and expected value with luminosity  $2.7 \text{ fb}^{-1}$

	Observed	Expected	$1\sigma$	$2\sigma$
	10.1	13.3	[9.5,18.6]	[7.1,25.0]
<i>Boosted</i>	13.1	23.0	[16.4,32.4]	[12.2,43.6]

The results show that the method used in this project would only be sensitive to the  $t\bar{t}H$  production, if the process had 13 times larger cross section. In the boosted category better performance of the method was expected. The reason why this is not the case would be the lower number of events. In this analysis, we also estimate the sensitivity for  $100 \text{ fb}^{-1}$  that correspond to the expected luminosity of the LHC after 2018 (Run II). Additionally, improvements in the definition of the boosted regime could improve the limit.

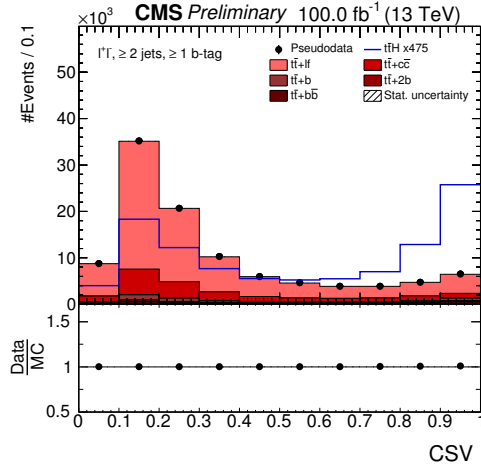


Figure 25: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet, 100  $fb^{-1}$  - Filter Jets

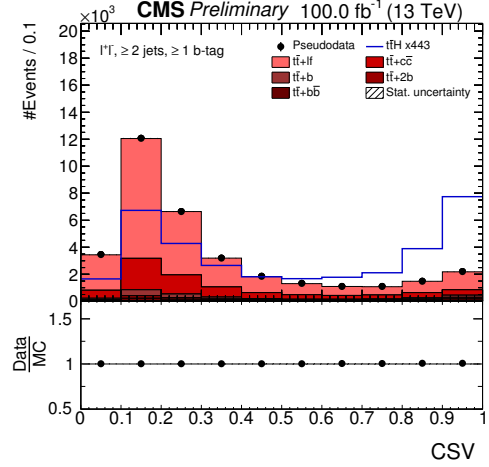


Figure 26: Distribution of the CSV discriminator of the 2<sup>nd</sup> highest b-tagged jet, 100  $fb^{-1}$ , boosted regime - Filter Jets

For this calculation, we used pseudodata and the systematic uncertainties have not been included. The expected limit provides a better sensitivity by a factor of 6, a consequence of the larger statistics.

Table 3: Expected value with luminosity 100  $fb^{-1}$

	Expected	$1\sigma$	$2\sigma$
	2.1	[1.5,3.0]	[1.1,3.9]
<i>Boosted</i>	3.6	[2.6,5.0]	[2.0,6.7]

## 4 Conclusion

The kinematics and topology of the boosted Higgs, produced in association with a  $t\bar{t}$  have been studied. The comparison of the two algorithm that investigate the substructure of the fat jets can be used for the identification of the decay of the Higgs boson to a  $b\bar{b}$  pair. The study of the two algorithms identifies which variables have discriminant power. As a next step, the systematic uncertainties will be included in the analysis.

## 5 References

### References

- [1] Felix Riese *Substructure algorithm studies for Higgs reconstruction*, 2016
- [2] Andrew J. Larkoski et al. “Soft Drop” In: *JHEP 05 (2014)*, p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].
- [3] CMS B-tagging Data Analysis school, 2015  
**b-tagging Exercise**
- [4] Jasone Garay Garcia. *Search for Higgs-Boson Production in Association with Top-Quark Pairs in  $H \rightarrow b\bar{b}$  Decays at  $\sqrt{s} = 8$  and 13 TeV with the CMS Experiment*, University of Hamburg, 2016