PROCEEDINGS OF SCIENCE

Latest results on $\mathrm{t\bar{t}H}$ $(\mathrm{H} \rightarrow \mathrm{bb})$ production at CMS

Daniel Salerno[*][†]
for the CMS Collaboration
E-mail: daniel.salerno@cern.ch

Details of the search for the Higgs boson produced in association with top quarks and decaying to bottom quarks at CMS are presented. The data corresponds to the first 12.9 fb^{-1} of 2016 proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ delivered from the LHC and recorded by the CMS experiment. Candidate events are selected in the lepton+jets and dilepton decay channels of the $\mathrm{t\bar{t}}$ system. Selected events are categorised into several categories with varying amounts of signal and background. In each category the signal extraction is performed using multivariate techniques combining the matrix element method and boosted decision trees. The results are presented in terms of the signal strength, i.e. the observed $\mathrm{t\bar{t}H}$ cross section relative to the standard model prediction, $\mu = \sigma/\sigma_{\mathrm{SM}}$. A combined fit of the final discriminants across all categories results in an observed (expected) upper limit of $\mu < 1.5(1.7)$ at the 95% confidence level and a best-fit value of $\mu = -0.19^{+0.45}_{-0.44}(\mathrm{stat.})^{+0.66}_{-0.68}(\mathrm{syst.})$.

EPS-HEP 2017, European Physical Society conference on High Energy Physics
5-12 July 2017
Venice, Italy

[*]Speaker.
[†]University of Zurich.

PoS(EPS-HEP2017)335

1. Introduction

The Higgs boson (H) produced in association with top quarks ($t\bar{t}H$ production) provides a direct probe of the important top-Higgs Yukawa coupling, which, due to the large top quark mass, is expected to be of order one. The leading order (LO) Feynman diagram for $t\bar{t}H$ production is shown in Figure 1.

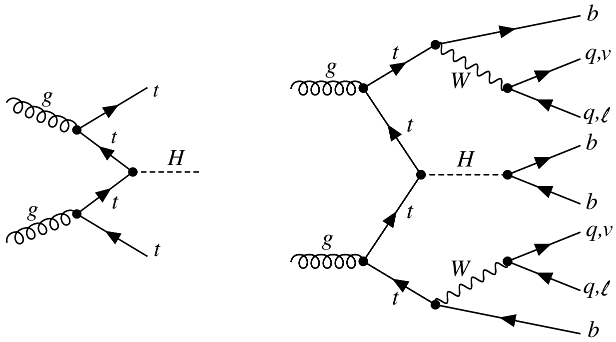

Figure 1: Feynman diagram of LO $t\bar{t}H$ production (left) and the $H \to b\bar{b}$ decay mode (right).

The search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay mode is performed in two separate analyses at CMS [1]. The leptonic analysis selects events with one or two leptons (muons or electrons) and then categorises events based on jet and b-tag multiplicity. It uses boosted decision trees (BDTs) and a matrix element method (MEM) as the final discriminants and has been completed with 12.9 fb^{-1} at 13 TeV [5]. The fully hadronic analysis selects events without leptons, giving access to a large fraction of $t\bar{t}H$, $H \to b\bar{b}$ decays. It includes the fully reconstructed final state of 8 jets including 4 b-jets, but suffers from a large QCD multijet background. First results for this analysis are currently in progress.

2. Analysis overview

Results on the $t\bar{t}H$, $H \to b\bar{b}$ search at CMS have been presented with three different datasets. The first dataset of 19.5 fb^{-1} of 2012 data at 8 TeV was analysed using a BDT discriminant [2] as well as the MEM [3]. The next dataset was 2.7 fb^{-1} of 2015 data at 13 TeV, which was analysed using a combination of BDT and MEM discriminants [4]. Most recently, results using the first 12.9 fb^{-1} of 2016 data at 13 TeV have been presented [5] and are discussed in these proceedings.

The data have been collected using single and double lepton triggers with varying p_{T} thresholds, where a lepton is defined as a muon or electron. The signal and background processes are estimated with Monte Carlo (MC) simulation and corrected for observed differences with respect to data. After an initial selection, events are divided into seven mutually exclusive categories based on jet and b-tagged multiplicity. The lepton+jets channel has four categories: 4 jets, 4 b-tags; 5 jets, ≥ 4 b-tags; ≥ 6 jets, 3 b-tags; and ≥ 6 jets, ≥ 4 b-tags. The dilepton channel has three categories: 3 jets, 3 b-tags; ≥ 4 jets, 3 b-tags; and ≥ 4 jets, ≥ 4 b-tags.

Both a BDT and a MEM have been used to discriminate signal from background in this analysis. The BDT discriminant uses the stochastic gradient boost method and is trained separately in each of the seven event categories. The training is performed on $t\bar{t}H$ versus $t\bar{t}+$jets using 50% of all available MC statistics and different input variables for each category. The MEM provides

PoS(EPS-HEP2017)335

an optimal separation of the signal and background. It also avoids the uncertainty in matching the detected objects with the underlying particles, by summing over all possible object-particle combinations. Specifically, it calculates the likelihood that the measured observables come from a $t\bar{t}H$ process (signal), w_S, and a $t\bar{t}+b\bar{b}$ process (background), w_B, and then combines the two into a final discriminant $P_{SB} = w_S/(w_S + \kappa w_B)$, where κ is a scale factor optimised to provide the best discrimination in each category. The method is similar to that described in Ref. [3].

Compared to the BDT, the MEM requires no training and can therefore be used with low MC statistics. It also has an easier interpretation, however it has a long computation time per event. In general the MEM performs better at the right number of jets and b-tags (high multiplicity), while the BDT performs better with higher statistics (low multiplicity).

In the lepton+jets channel, events in each of the four categories are subdivided into high and low BDT regions based on the median BDT score of the signal and then the MEM discriminant is fitted in each region. The low BDT regions are useful to constrain backgrounds and systematics, while the high BDT regions are signal enriched. Figure 2 shows the BDT output and subsequent MEM discriminants in the 5 jet, ≥ 4 b-tag category. In the dilepton channel, the 3 jet, 3 b-tag category uses the BDT as the final discriminant since this has a large number of events which is better for BDT training. The 4-jet categories use the same high/low BDT subdivisions with the MEM as in the lepton+jets channel.

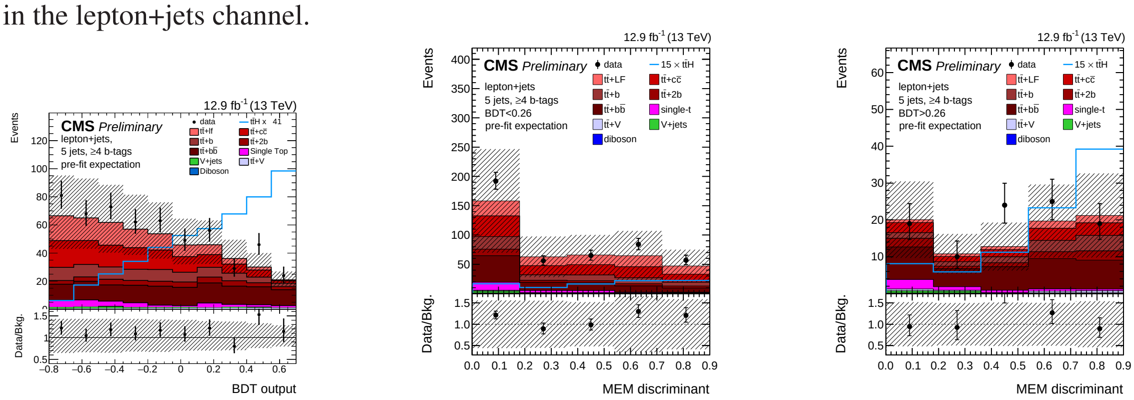

Figure 2: BDT output distribution (left) and the MEM discriminant in the low BDT region (centre) and high BDT region (right) in the 5 jet, ≥ 4 b-tag category of the lepton+jets channel. From Ref. [5].

A binned maximum likelihood fit is performed to the final discriminant across all categories considering all systematic uncertainties. The dominant systematic uncertainties in terms of yield are the jet energy scale uncertainty, b-tagging scale factor uncertainties, and tt+hf normalisation uncertainty (50%), which has the largest impact on sensitivity. Figure 3 shows examples of the pre and post-fit discriminants in the high BDT region of one category of each channel. As can be seen, the fit constrains the systematic uncertainties (dashed band) very well, indicating that they are quite conservative.

The results of the final fit are presented in terms of the signal strength, $\mu = \sigma_{t\bar{t}H}/\sigma_{SM}$. The 95% confidence level (CL) upper limit on μ and the best-fit signal strength are shown in Figure 4. The observed (expected) limit is 1.5 (1.7) times the standard model prediction. The best fit signal strength is -0.2 ± 0.8, which is dominated by systemic uncertainties as seen in Figure 4. For comparison, the expected upper limits at 8 TeV were 3.3 and 4.1 for the MEM and BDT respectively, and at 13 TeV with 2015 data, the expected limit was 3.6 and the best-fit μ was -2.0 ± 1.8.

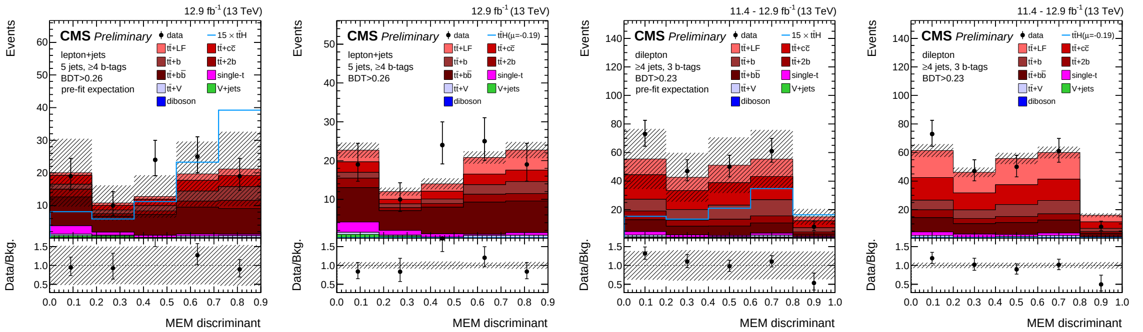

Figure 3: MEM distribution in the high BDT region of the 5 jet, ≥ 4 b-tag category of the lepton+jets channel, pre-fit (far left) and post-fit (centre left). MEM distribution in the high BDT region of the ≥ 4 jet, 3 b-tag category of the dilepton channel, pre-fit (centre right) and post-fit (far right). From Ref. [5].

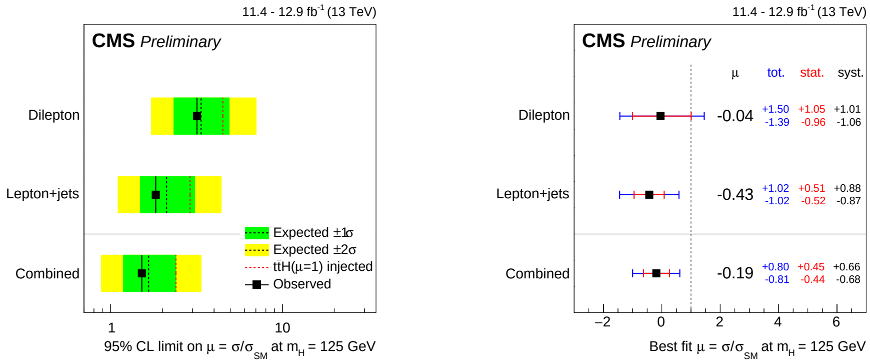

Figure 4: 95% CL upper limit on μ (left) and the best-fit μ (right). From Ref. [5].

3. Conclusion and outlook

In summary, the leptonic analysis has achieved strong results with only a partial 2016 dataset. Most recently, it has an observed limit of 1.5, an expected limit of 1.7 and a best-fit signal strength of $\mu = -0.2 \pm 0.8$. It shows a solid trend of improving performance, despite the increasing dominance of the systematic uncertainties, which currently contribute 60% to the total uncertainty. However, there are several long term efforts underway to constrain the dominant systematics. Looking ahead, new results using the full 2016 dataset for both the leptonic and hadronic analyses are expected soon. This will be followed by a combination with other $t\bar{t}H$ searches at CMS. In addition, efforts have already begun on updating these searches for the data being collected in 2017.

References

[1] CMS Collaboration, JINST 3 S08004 (2008).

[2] CMS Collaboration, CMS-PAS-HIG-13-019. https://cds.cern.ch/record/1564682

[3] CMS Collaboration, EPJC 75 (2015) 251 [hep-ex/1502.02485].

[4] CMS Collaboration, CMS-PAS-HIG-16-004. https://cds.cern.ch/record/2139578

[5] CMS Collaboration, CMS-PAS-HIG-16-038. https://cds.cern.ch/record/2231510