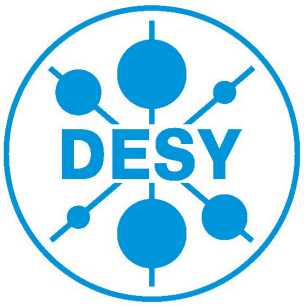

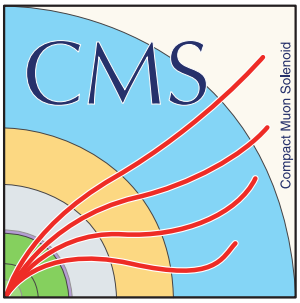

First look at $t\bar{t}H(b\bar{b})$ production in the dilepton channel in the boosted regime

Iván Esteban Muñoz, University of the Basque Country, Spain
Supervised by Carmen Diez Pardos, DESY

9th September 2015

Abstract

The kinematics and most immediate topological properties of the boosted Higgs regime, $p_{t,H} > 200$ GeV, in the dileptonic channel of $t\bar{t}H$ production with the Higgs boson decaying to a $b\bar{b}$ pair has been studied both at energies of $\sqrt{s} = 8$TeV and $\sqrt{s} = 13$TeV.

In particular, the angular distance between jets coming from a b-quark has been analysed. In addition, an algorithm to identify fat jets coming from the Higgs boson decay in the boosted regime has been implemented in the ntuples used by the DESY-CMS group. Its fundamental tools have been tested on a small MC sample of data and background, observing some promising signal-background separation variables.

Contents

1 Introduction

After the announcement in 2012 of the discovery of a Higgs-like particle by the ATLAS and CMS collaborations, its experimental study has become a fundamental research topic in high energy physics. In particular, it is essential to check its consistency with the Standard Model Higgs boson.

An important property of this particle is its coupling with fermions. This coupling, known as *Yukawa coupling*, is of especial interest for the top quark, the heaviest known elementary particle. Its Yukawa coupling is $\mathcal{O}(1)$, and some theories beyond the Standard Model give the top quark a role in electroweak symmetry breaking [1][1].

A process that directly tests the physics of the top-Higgs interactions is the production of a $t\bar{t}$ pair along with a Higgs boson. This phenomenon happens at LHC through a parton-parton interaction — 85% of the times a gluon-gluon process, 15% quark-quark — , whose leading-order Feynman diagram is depicted in Figure 1. The cross section for this process depends on the top-Higgs Yukawa coupling [2], and is sensitive to physics beyond the Standard Model.

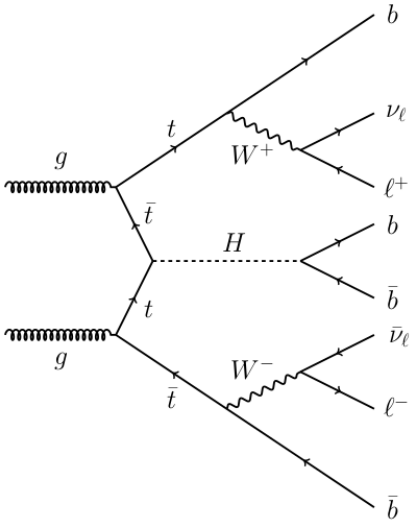

Figure 1: Leading-order Feynman diagram for $t\bar{t}H$ production via gluon fusion. Only the decay products considered in this report are shown. Source: Ref. [2].

Both the top quark and the Higgs boson are unstable particles; they decay before being detected and can only be identified by their decay products. Specifically, top quarks decay almost 100% of the times to a b quark and a W boson; the most frequent decay of the Higgs boson is into a $b\bar{b}$ pair, with a 57% branching ratio.

In this report, the topology of $t\bar{t}H$ events will be studied, comparing the results at centre of mass energies of $\sqrt{s}$=8 TeV and $\sqrt{s}$=13 TeV. To reduce the background and improve the purity of the signal, only the channel in which the final decay state of both W bosons only contains leptons, which has a 6.43% branching ratio, is considered. The analysis is also optimised for the decay of the Higgs boson into a $b\bar{b}$ pair; in particular, to the regime in which it has a high transverse momentum, referred to as the *boosted regime* in the following.

1.1 Experimental detection of the process: the background problem

Despite the physical interest of $t\bar{t}H$ production, any measurement of its cross section has proven to be highly difficult because of the overwhelming background: the final state can be mimicked

[1] Unless otherwise specified, most of the information in Section 1 has been extracted from Ref. [1]

Process	σ at $\sqrt{s}$ =8 TeV (pb)	σ at $\sqrt{s}$ =13 TeV (pb)[3–6]
$t\bar{t}H$ production	0.1293	0.5085
$t\bar{t}$ production	241.5	831.76
Drell Yan process	3503.71	6025.2
Single top/antitop production	11.1	35.6
Two gauge boson production	105.702	216.6

Table 1: Cross Section of $t\bar{t}H$ production and its main backgrounds. The Drell-Yan cross Section calculation assumes the invariant mass of the resulting dilepton pair to be ≥ 50 GeV.

by other Standard Model processes. Table 1 lists the most important backgrounds and their cross sections, as well as the cross Section of $t\bar{t}H$ production, at a centre of mass energy $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV.

In particular, at $\sqrt{s} = 8$ TeV $t\bar{t}$ production has a cross section about 1900 times larger than $t\bar{t}H$ production. The problem, though slightly less severe, remains at $\sqrt{s} = 13$ TeV, where the ratio $\sigma_{t\bar{t}}/\sigma_{t\bar{t}H}$ is about 1600. Furthermore, some of the $t\bar{t}$ processes, shown in Figure 2, are an irreducible background of $t\bar{t}H$ production.

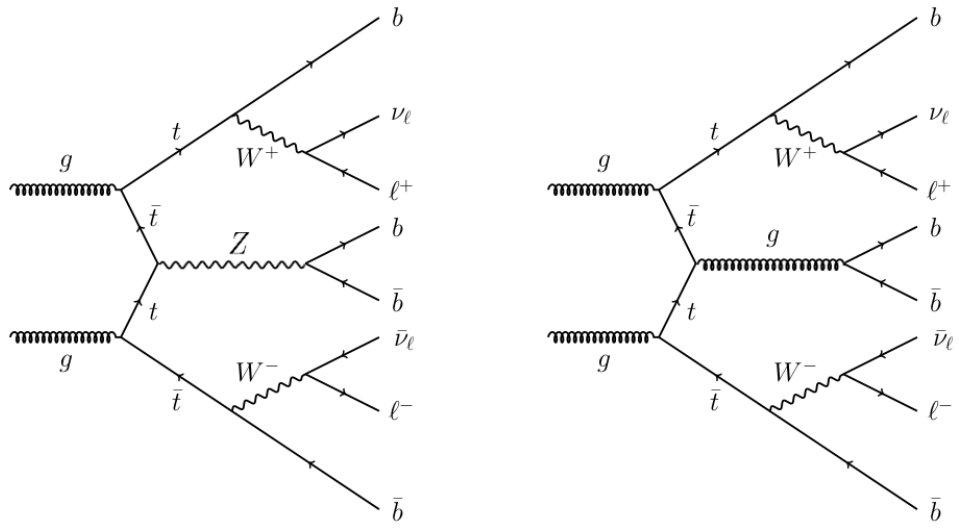

Figure 2: $t\bar{t}Z$ and $t\bar{t}g$ processes with the same decay products as the considered $t\bar{t}H$ production. With a high enough precision, the analysis of the invariant mass distribution of the $b\bar{b}$ system allows to distinguish them from $t\bar{t}H$ events.

Even though the Drell Yan process has the largest cross section amongst the processes listed in Table 1, it is largely rejected by the selection criteria described in Section 1.2. The contribution from $t\bar{t}$, however, has a similar topology and cannot be rejected by a simple selection.

Therefore, it is crucial to develop analysis techniques that deal with this problem, both improving theoretical calculations of the background and exploiting topological differences between $t\bar{t}$ and $t\bar{t}H$ processes. This project aims to contribute to the latter approach, by analysing the topology of the boosted Higgs regime.

The analysis of this regime poses a challenge to the usual analysis techniques. Because of energy and momentum conservation, the jets resulting from Higgs decay are emitted collinear with respect to the trajectory of the parent particle. Therefore, they may be misidentified as just one jet and the event would fail a selection criteria that requires at least four jets, a usual

approach for the analysis of $t\bar{t}H$ events. In order to keep the highest selection efficiency in this regime, new analysis strategies and algorithms are needed to disentangle signal and background. This is discussed in Sections 3 and 4. In particular, the approach will be first to look at closely emitted b-jets, and then to analyse structural properties of jets with a high angular size that might contain the two b-jets from Higgs decay; always comparing signal and background.

1.2 Event reconstruction and selection

When analysing the experimental data, the acceptance and performance of the detector has to be taken into account to select events that minimise misidentification. For that, several selection steps are applied both to data and simulations:

1. The reconstructed interaction vertex is required to have good quality, i.e., a significant amount of tracks associated with it and a small distance to the proton beam.

   Additionally, several cuts are applied for the decay particles to ensure that they can be reliably reconstructed:

   - Leptons are required to have $|\eta| > 2.4$, $p_t > 20$ GeV. $\eta \equiv -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$ is the pseudorapidity, where θ is the polar angle, and p_t is the transverse linear momentum.
   - Jets are required to have $|\eta| > 2.4$ and $p_t > 20$ GeV.

   The event is also required to have reconstructable jets and leptons. Further details about algorithms and isolation criteria for this can be found in Ref. [1].

2. A dilepton pair of opposite charge is required to be present.

3. The dilepton pair is required to have an invariant mass $m_{l\bar{l}} > 20$ GeV to reduce the Drell-Yan and QCD-multijet background.

4. If both leptons have the same flavour, the events with 76 GeV$<m_{l\bar{l}}<$106 GeV are rejected to dismiss the Drell-Yan background contribution.

5. At least 2 jets are required to be present.

6. If both leptons have the same flavour, the missing transverse energy, corresponding to neutrinos, is required to be greater than 40 GeV. This further reduces the Drell-Yan background.

7. At least one b-tagged jet is required to be present. To reconstruct it, an algorithm known as *Combined Secondary Vertex* (CSV) is used. This algorithm exploits the long lifetime of b hadrons, $\mathcal{O}(ps)$, that allows them to travel a distance $\mathcal{O}(mm)$ before decaying. Consequently, the identification of a secondary origin vertex for the particles in the jet close to the primary interaction vertex serves to identify b-jets. More details about the algorithm can be found in Ref. [7].

Finally, the kinematic properties of the top and antitop quarks can also be reconstructed based on their decay products. To do so, 4-momentum conservation is demanded and the physically most probable solution is chosen. Further details can be found in Ref. [8].

The analysis is performed using data and simulations at centre of mass energies of $\sqrt{s}$=8 TeV and $\sqrt{s}$=13 TeV. Unless otherwise specified, the integrated luminosity, $\mathcal{L}$, is of 19.7fb^{-1} at $\sqrt{s}$=8 TeV and of 40.0pb^{-1} at $\sqrt{s}$=13 TeV.

2 Kinematics of the boosted Higgs regime

A first way of approaching the boosted Higgs regime comes from the analysis of the transverse momentum p_t of the $t\bar{t}$ system. Figure 3 shows its Monte Carlo-generated momentum distribution in different channels, which, by momentum conservation, is directly correlated to the transverse momentum of the Higgs boson in $t\bar{t}H$ events. This correlation is explicitly shown in Figure 4.

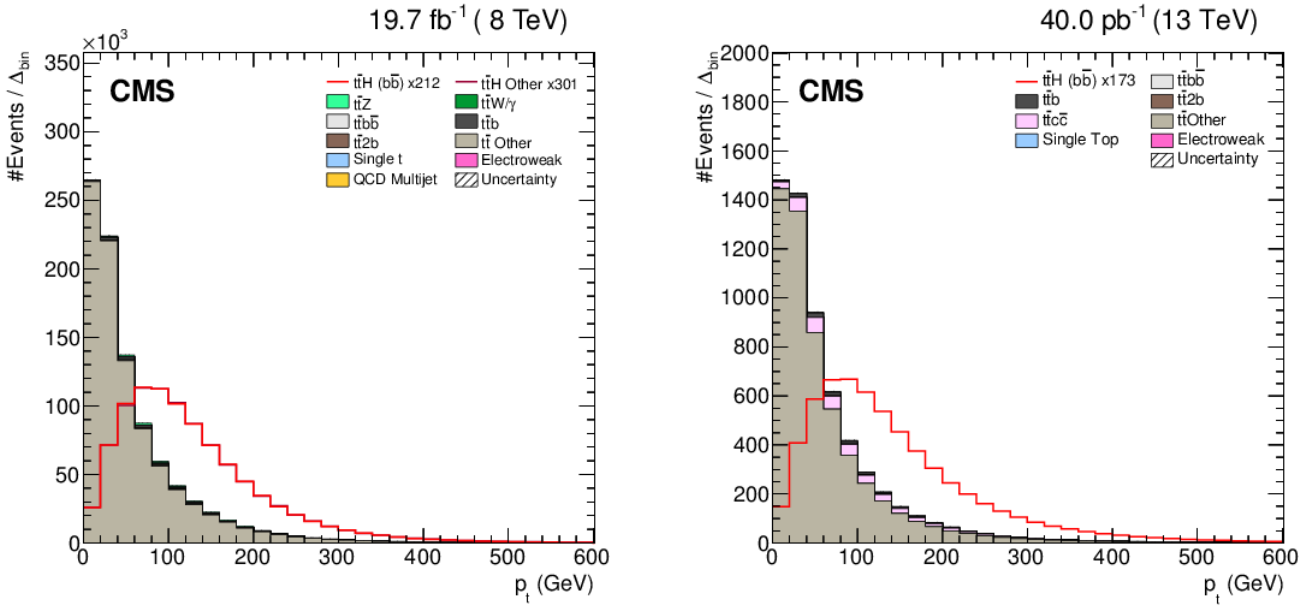

Figure 3: Generated p_t distribution of the $t\bar{t}$ system. Left: $\sqrt{s}$=8 TeV. Right: $\sqrt{s}$=13 TeV.

As the figures show, the $t\bar{t}H$ signal is more prominent in the boosted regime. Quantitatively, after all the selection steps described in Section 1.2, at $\sqrt{s}$=8 TeV the signal-to-background ratio increases from $(4.59 \pm 0.04) \cdot 10^{-4}$ to $(1.52 \pm 0.03) \cdot 10^{-3}$ if the $t\bar{t}$ system is boosted. At $\sqrt{s}$=13 TeV a $(5.73 \pm 0.03) \cdot 10^{-4}$ signal-to-background ratio increases to $(1.76 \pm 0.02) \cdot 10^{-3}$. For all the calculations the channel in which the Higgs boson decays to a $b\bar{b}$ pair has been considered as signal and the other $t\bar{t}$ events as background.

Comparing the transverse momentum at the generated level with the reconstructed one, it was found that in the boosted regime the reconstruction efficiency $\varepsilon \equiv 1 - \frac{p_{t\text{gen}} - p_{t\text{reco}}}{p_{t\text{gen}}} = 0.993 \pm 0.001$. Deviations mostly come from the high $p_t \gtrsim 500$ GeV region, where the statistics are anyway very low. For completeness, Figure 5 shows the reconstructed momentum distribution of the $t\bar{t}$ system after all the selection steps described in Section 1.2.

The study of the boosted regime is specially interesting at $\sqrt{s}$=13 TeV, where the fraction of events in which the Higgs boson is boosted increases by about 30% compared with $\sqrt{s}$=8 TeV. This behaviour can be seen in Figure 6, where the p_t distribution of the $t\bar{t}$ system in $t\bar{t}H$ events is compared for $\sqrt{s}$=8 TeV and $\sqrt{s}$=13 TeV.

3 Identification of b-jets from boosted Higgs decay

The study of boosted topologies, apart from increasing the signal fraction, requires new analysis techniques. As has been mentioned, when a boosted Higgs decays to a $b\bar{b}$ pair, the resulting jets are emitted closely to each other; an analysis of the angular separation of b-jets may lead to the

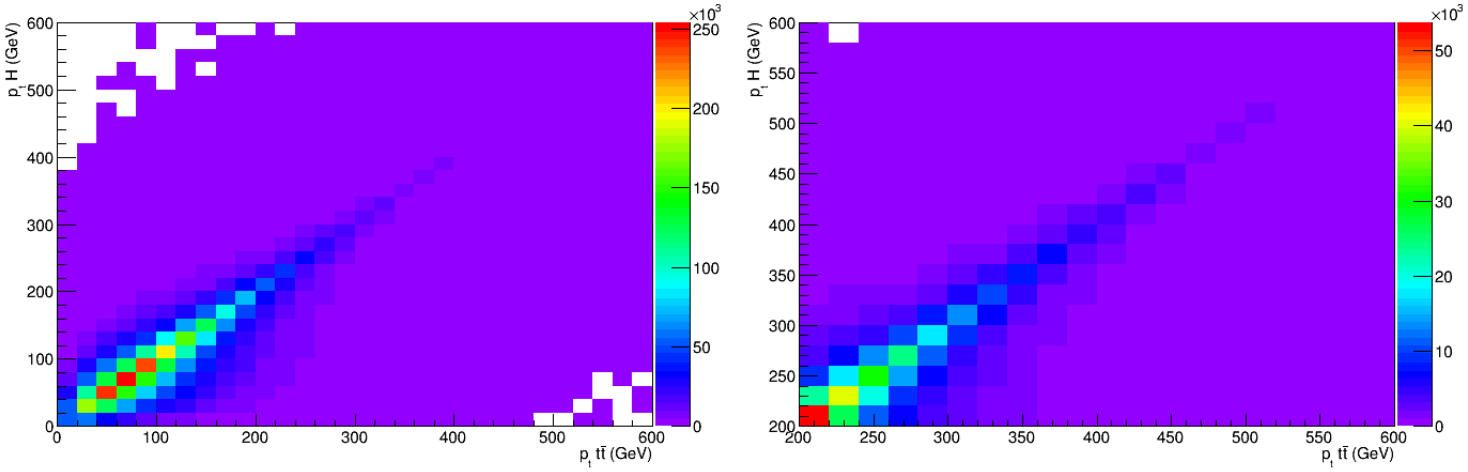

Figure 4: Correlation between the generated p_t of the $t\bar{t}$ system and of the Higgs boson at $\sqrt{s}$=8 TeV. The right picture zooms in the p_t>200 GeV region. The linear correlation factor has been found to be R=0.81.

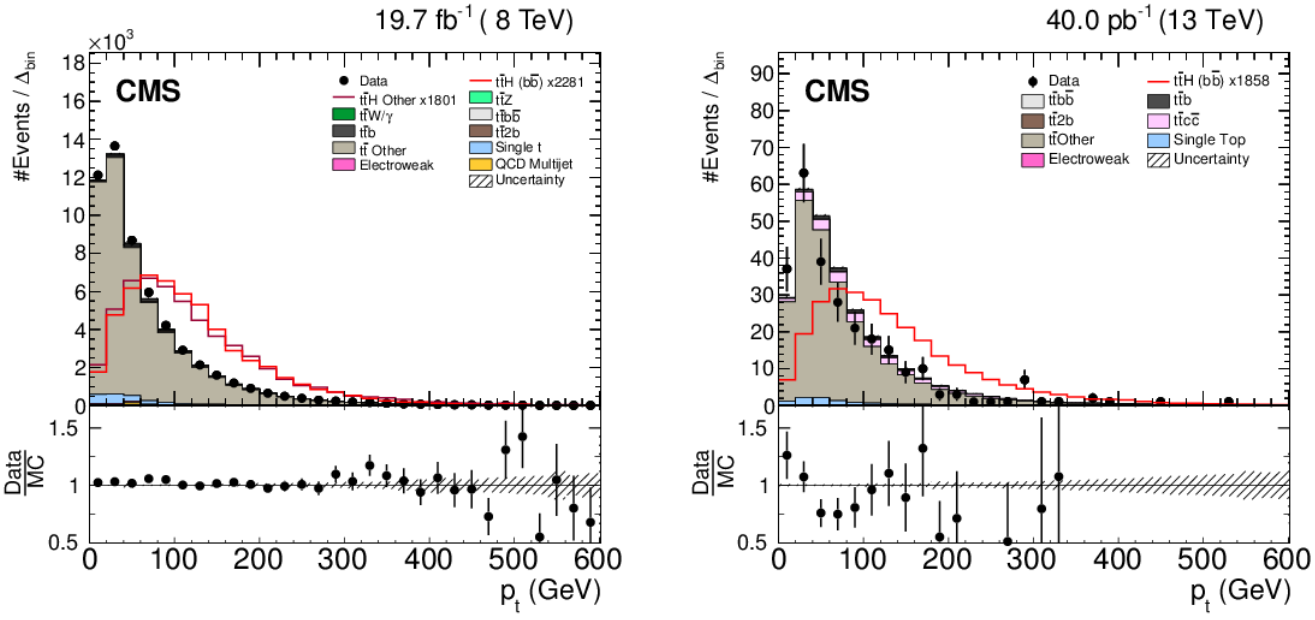

Figure 5: Reconstructed p_t distribution of the $t\bar{t}$ system after the selection steps. Left: $\sqrt{s}$=8 TeV. Right: $\sqrt{s}$=13 TeV.

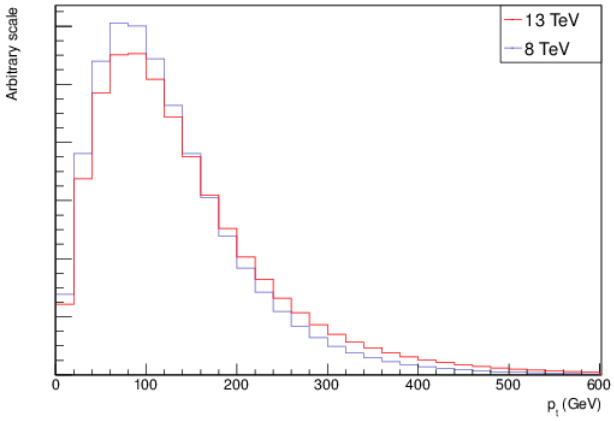

Figure 6: Comparison of the p_t distribution for the $t\bar{t}$ system in $t\bar{t}H$ events at different centre of mass energies.

identification of $t\bar{t}H$ processes. To do so, a measure of the angular distance between two jets[2] ΔR is introduced as

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

where ϕ is the azimuthal angle, measured in the plane perpendicular to the beam direction.

Figure 7 shows the smallest value of ΔR between b-jets both for all momenta of the $t\bar{t}$ system as well as for the boosted regime at $\sqrt{s}$=8 TeV. As can be seen, the $t\bar{t}H$ signal shows a peak around $\Delta R \sim 0.1$, particularly in the boosted regime, where the signal-to-background ratio reaches a value $\mathcal{O}(5 \cdot 10^{-3})$. Since the plots also deal with experimental data, both the momenta of the top and antitop quark as well as the identification of b-jets have been kinematically reconstructed.

Figure 8 shows the same plot for $\sqrt{s}$=13 TeV. Because of the low statistics of the boosted regime, however, to better visualise its topology simulation data has been normalised there to an integrated luminosity $\mathcal{L} = 2$ fb^{-1}. This luminosity corresponds to the expected integrated luminosity collected by the end of the year. The result is shown in Figure 9. There, the signal-to-background ratio is $\mathcal{O}(6 \cdot 10^{-3})$.

The signal-to-background ratio can be improved even more if only events with at least four jets and at least 3 b-jets are considered. The result is show in Figure 10 taking $\mathcal{L} = 2$ fb^{-1} for $\sqrt{s} = 13$ TeV. Even though the statistics is very low for both centre of mass energies and simulation uncertainties grow, the signal-to-background ratio is $\mathcal{O}(10^{-2})$; interestingly, the $t\bar{t}$Other background is highly reduced as a result of tightening the selection.

4 Fat jets

The peak at small values of ΔR in the b-jet distribution shows that, specially for boosted topologies, it is likely to reconstruct the b-jets coming from the Higgs decay as a single jet, usually known as *fat jet* because of its bigger size.

In this work, an algorithm to reconstruct fat jets and study their substructure has been implemented in the ntuples used by the DESY CMS group, constituting the first analysis of fat jets in the dilepton channel. Further information about the algorithm can be found in Ref. [9]. Overall, it looks for boosted jets — $p_t > 200$ GeV — with an angular aperture $\Delta R = 1.2$, rather than the usual aperture used for jets of $\Delta R = 0.5$. Then, subjets inside the main jet are reconstructed. Some useful kinematic properties of fat jets coming from $t\bar{t}H$ events run on a small

[2]The angular distances are measured between the sum for each jet of the linear momenta of its constituents.

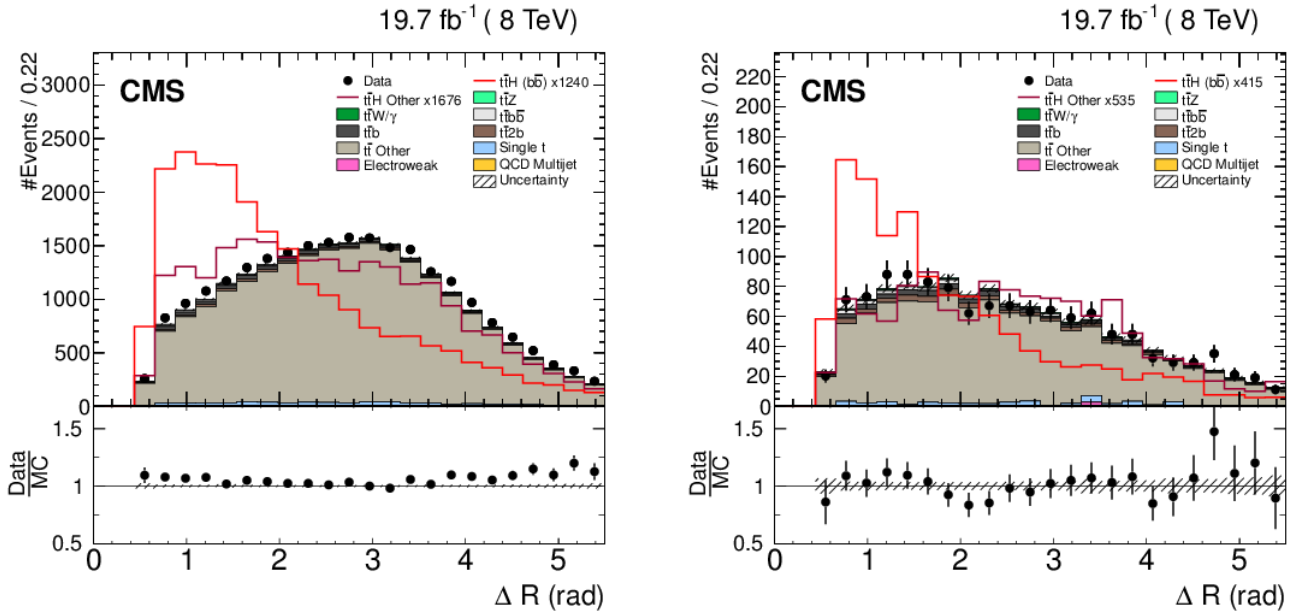

Figure 7: Minimum ΔR between b-jets. At right, boosted regime.

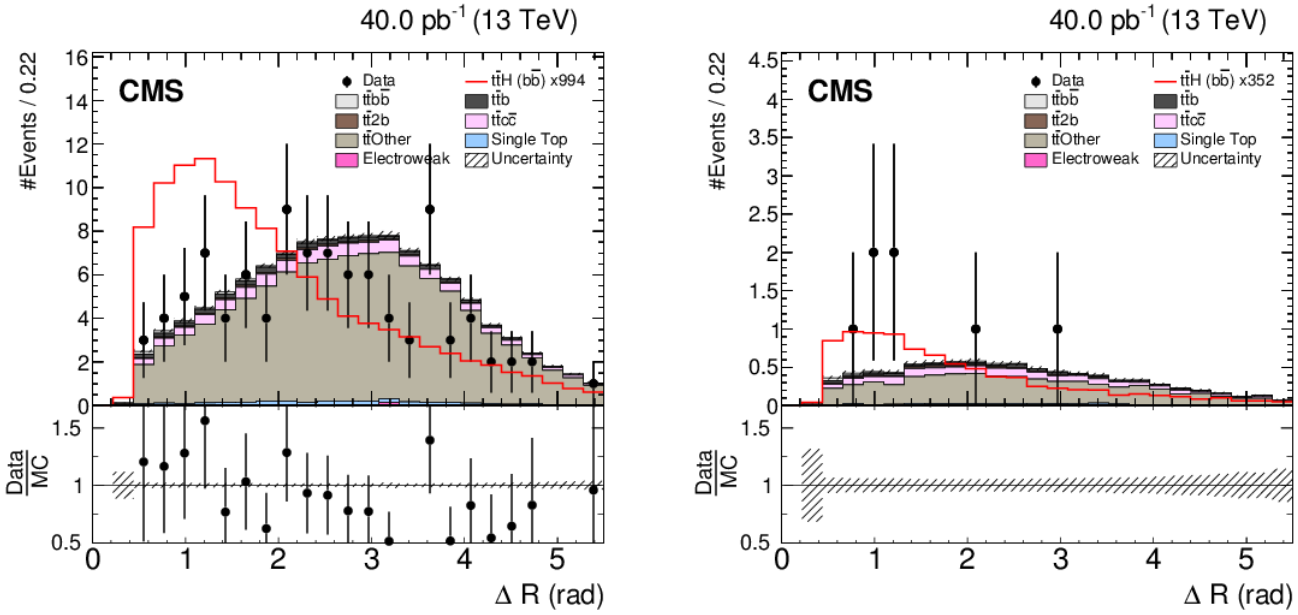

Figure 8: Minimum ΔR between b-jets. At right, boosted regime. Notice the difference in the scale of the $t\bar{t}H$ signal.

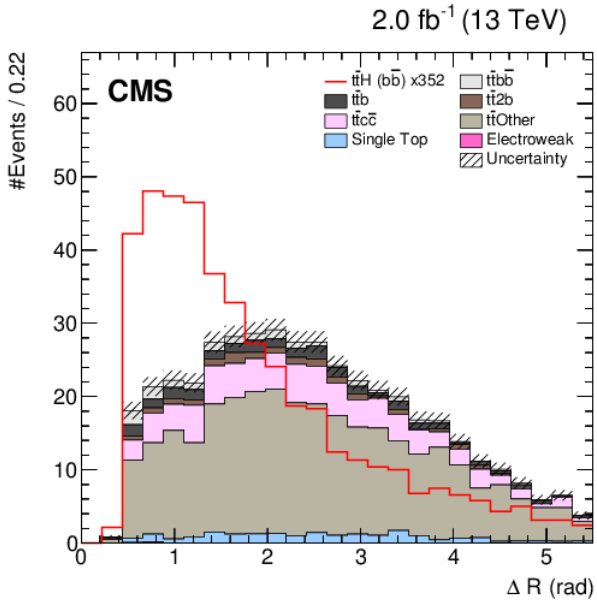

Figure 9: Simulated minimum ΔR between b-jets at $\sqrt{s}$=13 TeV and $\mathcal{L} = 2$ fb $^{-1}$.

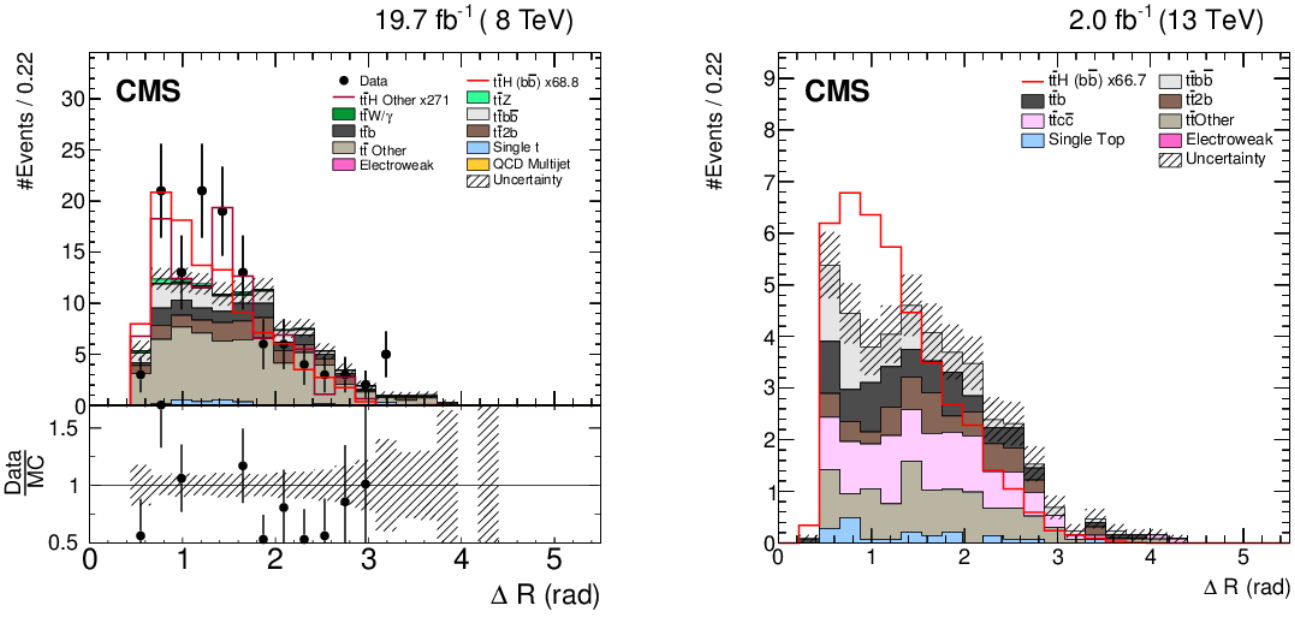

Figure 10: Minimum ΔR between b-jets for the boosted regime, requiring at least four jets and at least 3 b-jets.

MC sample of signal and background will be analysed in the next subsections. For completeness, Appendix A shows a comparison with experimental data in the eμ channel after all the cuts described in Section 1.2.

4.1 Angular separation

Considering that a fat jet resulting from a Higgs boson decay is expected to have two b-jets equally spaced from the fat jet centre, angular distances of subjets may be interesting taggers.

In particular, Figure 11 shows the angular distance ΔR between the subjet with the highest b-tagging discriminant and the fat jet, as well as between all the subjets and the parent fat jet. As can be seen, the $t\bar{t}H$ signal has these distributions shifted towards values relatively far from the fat jet centre.

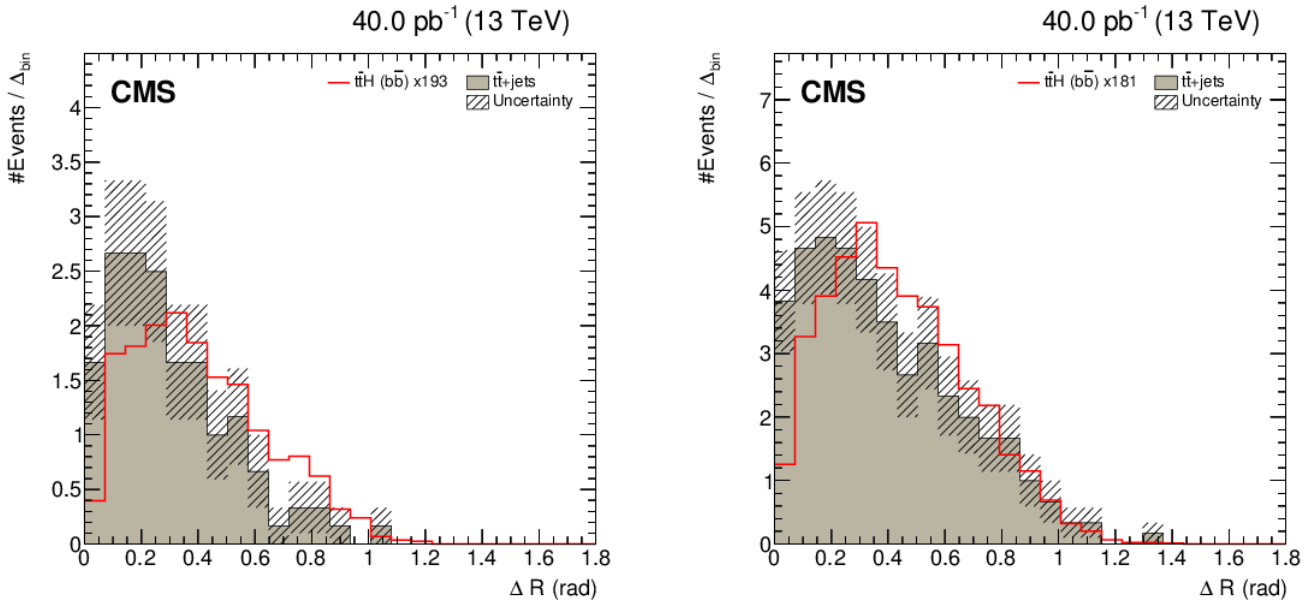

Figure 11: ΔR between the highest b-tagged subjet and the parent fat jet (left) and between the subjets and the parent fat jet (right).

4.2 b-tagging discriminant

The b-tagging discriminant of the subjets can also help to distinguish $t\bar{t}H$ events by looking for fat jets with two b-subjets. Figure 12 represents the highest and second-highest CSV discriminant values, respectively.

Even though the statistics of the background for the second highest CSV discriminant is not high enough to make definite assertions, the topology of the high CSV region favours in both cases the $t\bar{t}H$ signal.

4.3 Number of fat jets and subjets

Another useful tagging variable is the amount of subjets present in a certain fat jet. Figure shows it for $t\bar{t}H$ and $t\bar{t}$ events.

As is expected, $t\bar{t}H$ events contain more subjets than $t\bar{t}$ events. The difference, though, is not particularly high. A more intelligent approach employs a variable known as N-subjettiness.

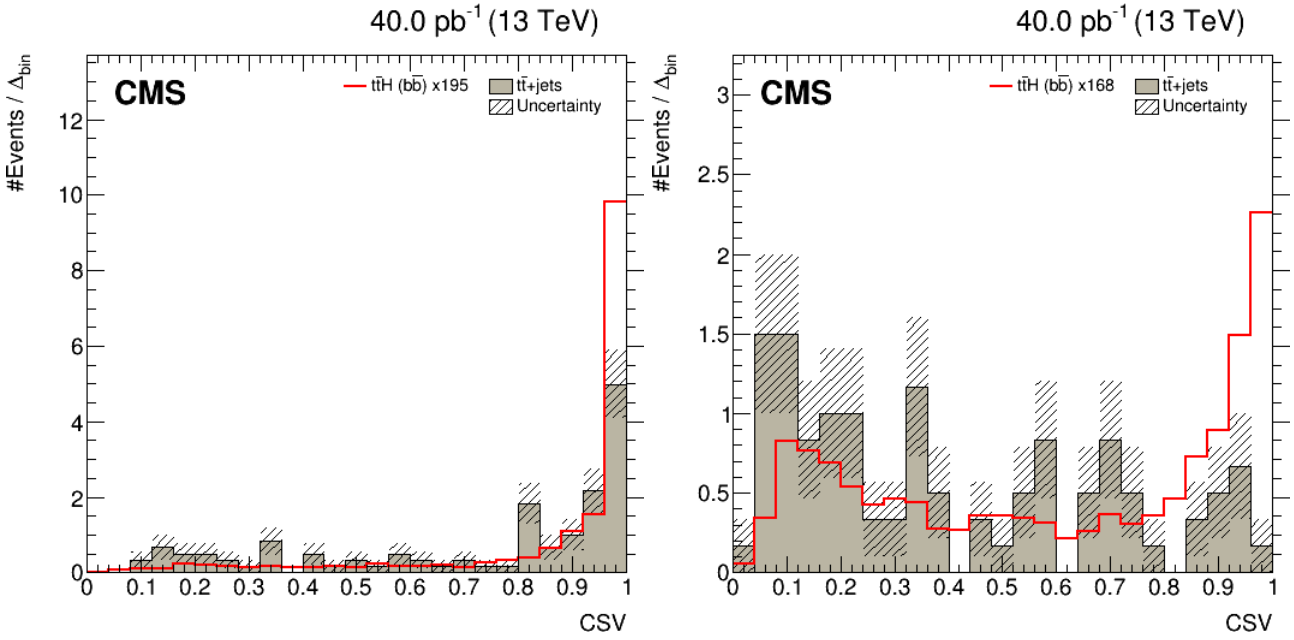

Figure 12: Highest (left) and second highest (right) CSV discriminant.

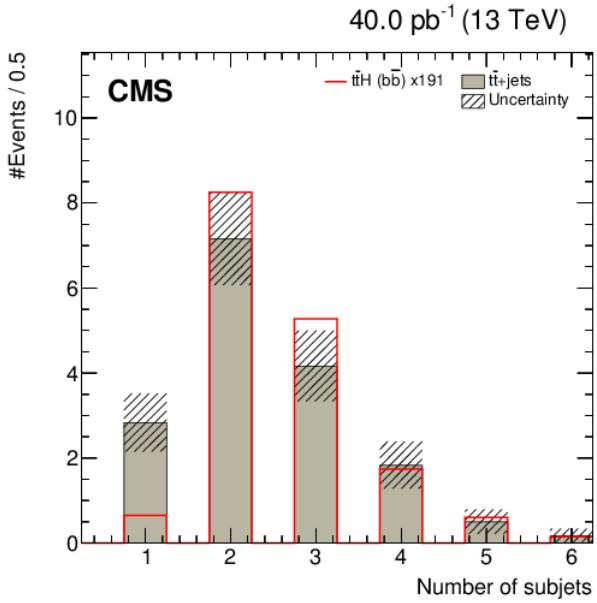

Figure 13: Number of subjets in a fat jet.

Usually denoted τ_N, it is defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{t,k} \Delta R_k \,. \tag{2}$$

It is a variable that measures the probability of the fat jet *not* being formed by more than N subjets by analysing the energy distribution inside the fat jet [10]. The sum runs over all the constituent particles k of a fat jet, where $p_{t,k}$ is their transverse momenta and, once the fat jet has been divided in N subjets, ΔR_k is the angular distance, as defined in Equation (1), of the k-th particle to the closest subjet. Therefore, N-subjettiness is smaller if the constituents are aligned with N reconstructed subjets. d_0 is a normalisation factor given by

$$d_0 = \sum_k p_{t,k} \Delta R_0 \tag{3}$$

where ΔR_0 is the angular aperture of the fat jet.

τ_1, τ_2, as well as its ratio τ_2/τ_1 are in principle useful tagging variables; larger, smaller and smaller, respectively, for $t\bar{t}H$ events compared with $t\bar{t}$ processes. Because of the limited statistics in the $t\bar{t}$ sample (see Figure 14), however, a study of their discriminant power has proven to be inconclusive with the studied precision.

5 Conclusions

The first studies of the kinematic variables and topologies of the boosted Higgs regime in $t\bar{t}H$ production have been presented. Its physical interest as a region that increases the cleanliness of the signal and adds new analysis techniques has been shown.

In particular, the study of fat jets brings a whole new set of variables that, appropriately exploited, open up new analysis perspectives. With more MC studies and the combination with other techniques, the technical implementation of the fat jet algorithms that has been done in this project might help to the discovery of the $t\bar{t}H$ process at $\sqrt{s} = 13$ TeV.

6 Acknowledgements

I would like to acknowledge several people that have backed the development of this project. Firstly, I feel grateful to my supervisor, Carmen Diez Pardos, for her help and support. I would like to thank as well Jasone Garay García, for her scientific and personal help through my stay here. And, in general, I feel grateful to all the people that organise the summer student programme for giving us this once-in-a-lifetime opportunity.

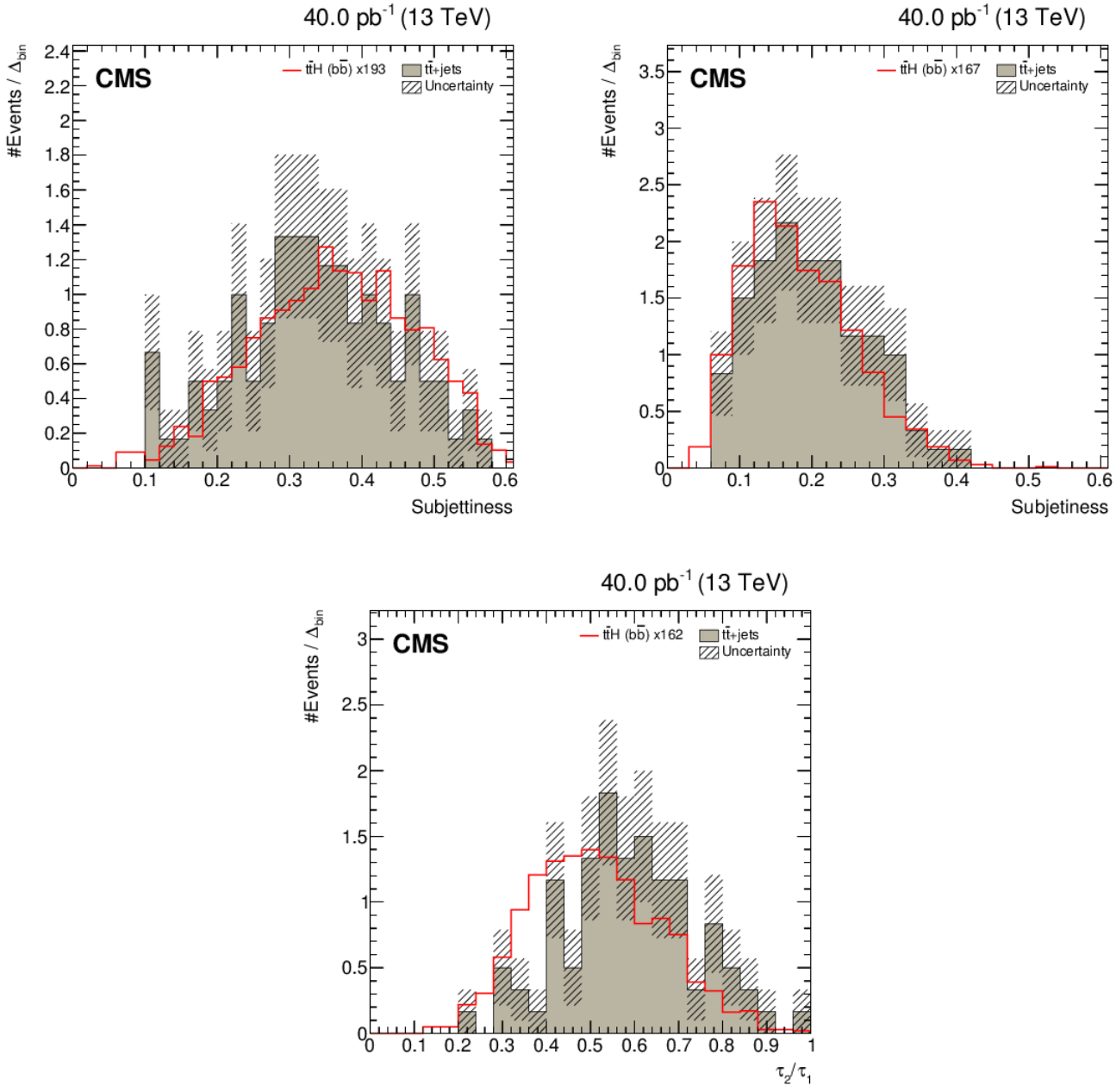

Figure 14: From left to right and top to bottom: τ_1,τ_2,τ_1/τ_2. Even though the results are inconclusive, some qualitative differences between signal and background can be seen.

A Fat jet plots with eμ data

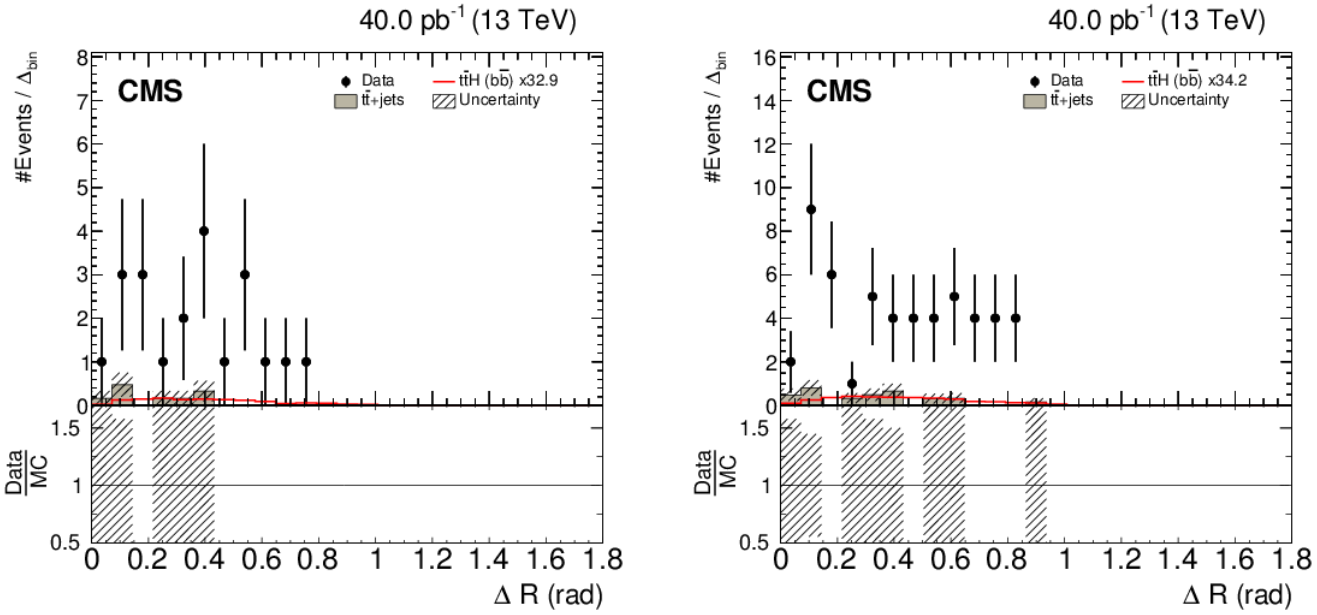

Figure 15: ΔR between the highest b-tagged subjet and the parent fat jet (left) and between the subjets and the parent fat jet (right).

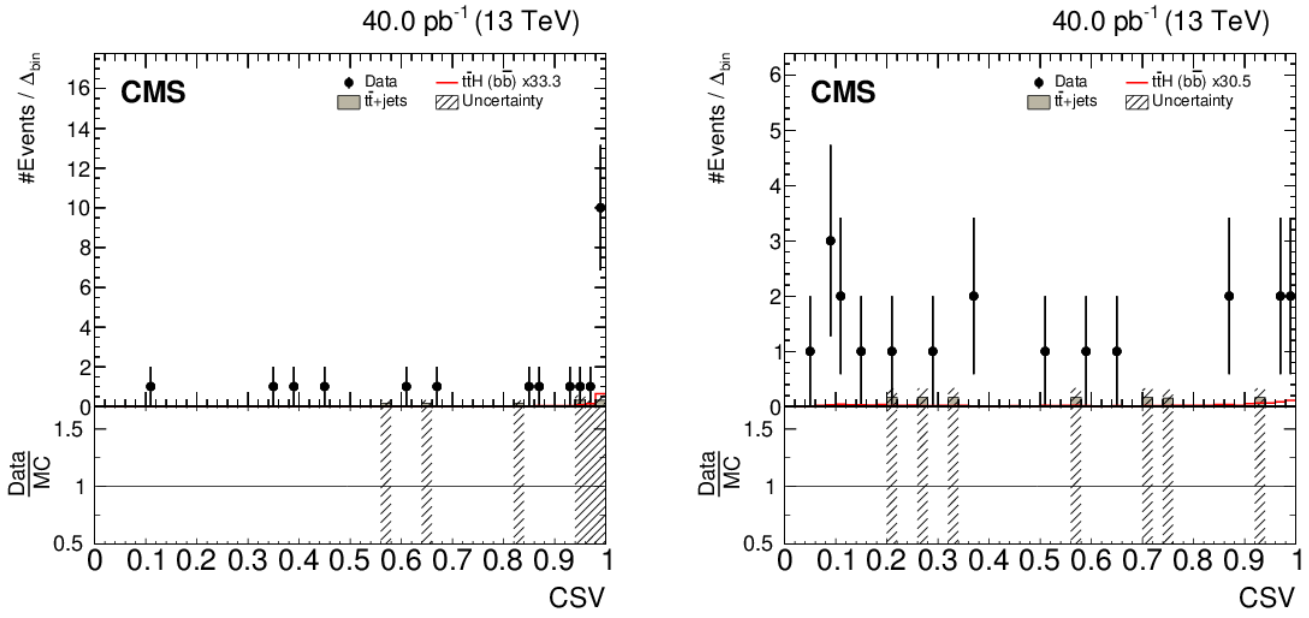

Figure 16: Highest (left) and second highest (right) CSV discriminant.

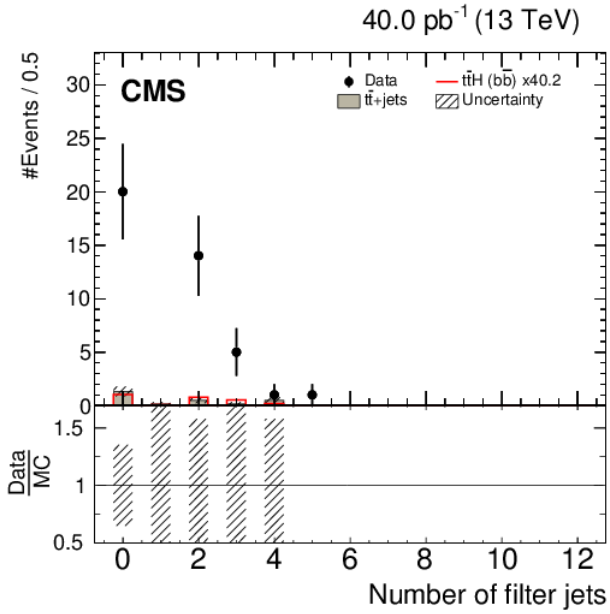

Figure 17: Number of subjets in a fat jet.

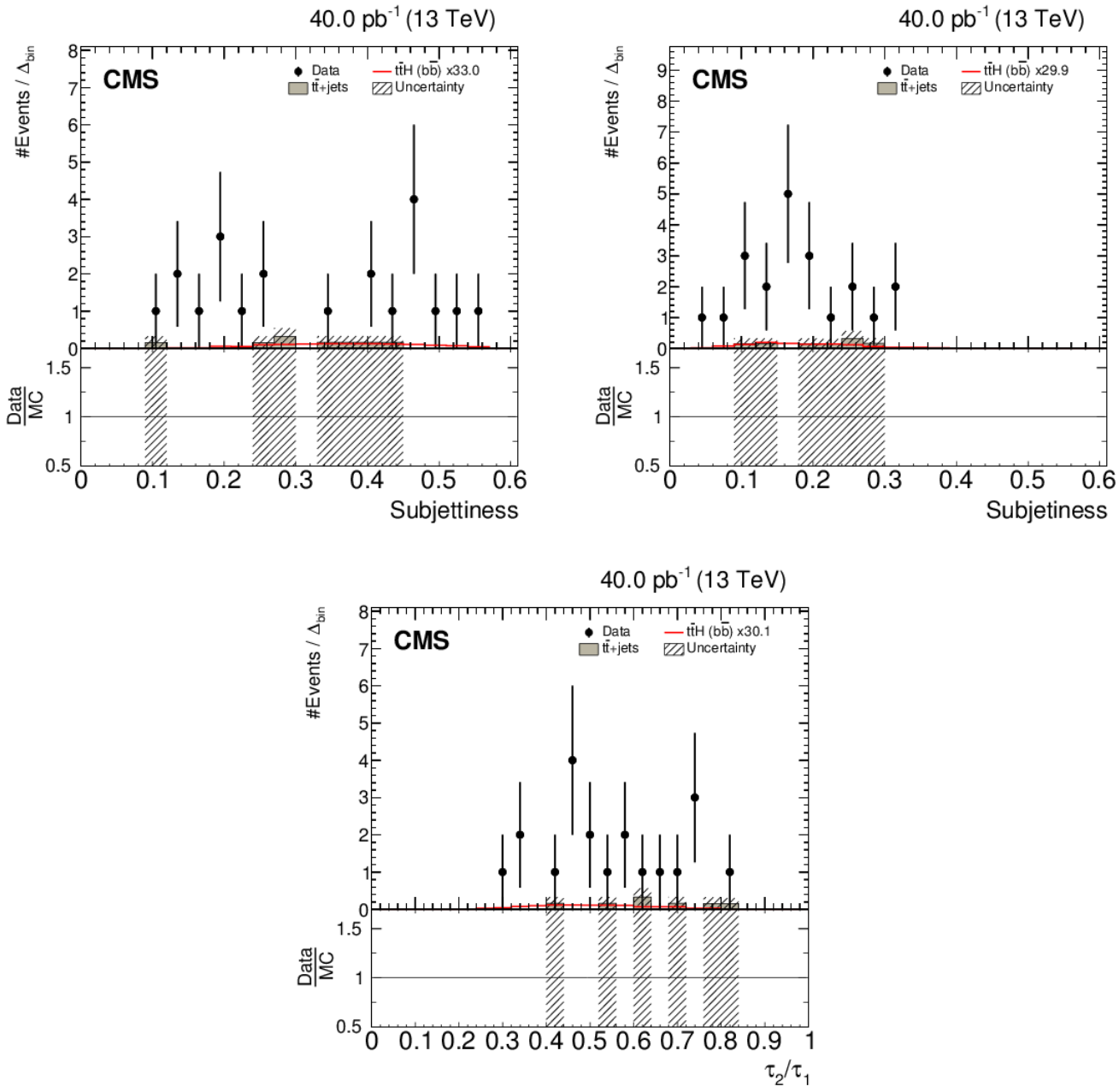

Figure 18: From left to right and top to bottom: τ_1,τ_2,τ_1/τ_2.

References

[1] Nazar Bartosik. 'Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as a test of QCD and background to tt+Higgs production'. PhD thesis. Hamburg: Universität Hamburg, July 2015.

[2] Serguei Chatrchyan et al., CMS Collaboration. 'Search for the standard model Higgs boson produced incassociation with a top-quark pair in pp collisions at the LHC'. In: *JHEP* 05 (2013), p. 145. DOI: 10.1007/JHEP05(2013)145. arXiv: 1303.0763 [hep-ex].

[3] LHC Higgs Cross Section Working Group. *SM Higgs production cross sections at* $\sqrt{s} = 13 - 14$ *TeV*. 2015. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV (visited on 31/08/2015).

[4] LHC Higgs Cross Section Working Group. *NNLO+NNLL top-quark-pair cross sections*. 2013. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO (visited on 31/08/2015).

[5] Tomislav Seva and Henry Yee-Shian Tong. *Standard Model Cross Sections for CMS at 13 TeV*. 2015. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV (visited on 31/08/2015).

[6] Nikolaos Kidonakis. 'Top Quark Production'. In: *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013)*. 2014, pp. 139–168. DOI: 10.3204/DESY-PROC-2013-03/Kidonakis. arXiv: 1311.0283 [hep-ph]. URL: http://inspirehep.net/record/1263209/files/arXiv:1311.0283.pdf.

[7] CMS Collaboration. 'Performance of b tagging at $\sqrt{s}$=8 TeV in multijet, ttbar and boosted topology events'. In: (2013).

[8] Ivan Asin Cruz. 'Measurement of Top-Quark-Pair Differential Cross Setions in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Experiment'. PhD thesis. U. Hamburg, Dept. Phys., 2014. URL: http://www-library.desy.de/cgi-bin/showprep.pl?thesis14-045.

[9] Tilman Plehn et al. 'Stop Reconstruction with Tagged Tops'. In: *JHEP* (2010). arXiv: 1006.2833 [hep-ex].

[10] Jesse Thaler and Ken Van Tilburg. 'Identifying Boosted Objects with N-subjettiness'. In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].