
CMS Physics Analysis Summary

Contact: cms-pag-conveners-higgs@cern.ch

2016/03/23

Search for $t\bar{t}H$ production in multilepton final states at $\sqrt{s}=13$ TeV

The CMS Collaboration

Abstract

A search for the standard model Higgs boson produced in association with a top quark pair is presented. Data collected by the CMS experiment in pp collisions at a center of mass energy of $\sqrt{s} = 13$ TeV and corresponding to an integrated luminosity of 2.3 fb^{-1} are used. The analysis targets the WW^* , ZZ^* and $\tau\tau$ decay channels of the Higgs boson by selecting final states with two same-sign leptons or more than three leptons, produced in association with b jets. The signal strength is measured to be $0.6^{+1.4}_{-1.1}$ times the standard model expectation. The observed 95% confidence level upper limit on the $t\bar{t}H$ production cross section is 3.3 times the standard model expectation, compared to the 2.6 expected in absence of a signal.

1 Introduction

The LHC Run I data have been exploited to measure nearly all the accessible properties of the newly-discovered Higgs boson [1, 2]. ATLAS and CMS have combined their data in order to reach a precise measurement of the boson mass, 125.09 ± 0.21 (stat.) ± 0.11 (syst.) GeV [3]. Measurements of the Higgs boson production and decay rates and constraints on its couplings to the particles of the standard model (SM) have been performed by both experiments, and combined [4–6]. In general, agreement with the SM predictions given the current uncertainties (10–30%) has been found. It is of great interest to use the 13 TeV LHC data to further constrain these measurements as any deviation from the expectation could be a sign of new physics.

The coupling of the Higgs boson to top quarks is particularly relevant. Given its large mass, the top quark could play a special role in the breaking of the electroweak symmetry. The largest contribution to the total SM Higgs production cross section at the LHC is due to gluon fusion, where virtual top quarks dominate this loop-induced process. The top-Higgs interaction vertex is only directly accessible when the Higgs boson is produced in association with one or more top quarks, since the Higgs boson is too light to decay to top quarks directly. Higgs boson production in association with a top quark pair (denoted $t\bar{t}H$ production, as shown in Fig. 1) offers the best opportunity for measurement of the top-Higgs coupling at the LHC. A comparison of the directly-measured top-Higgs coupling with the one inferred by other cross section measurements can be used to constrain contributions from new physics to the gluon fusion loop.

The $t\bar{t}H$ process has been used by both experiments to directly probe the top-Higgs coupling at tree level with the $5 + 20 \text{ fb}^{-1}$ of LHC Run I data at 7 and 8 TeV. The production cross section (130 fb at 8 TeV at next-to-leading order (NLO) [7]) is two orders of magnitude lower than that of gluon fusion. Nevertheless, both experiments reached a 30% accuracy on the top Yukawa coupling via this process, by exploring several topologies [8–11]. A first set of analyses targeted Higgs boson decays to $b\bar{b}$; the best fit value for the $t\bar{t}H$ signal strength obtained on Run I data by the CMS and ATLAS experiment are $0.7^{+1.9}_{-1.9}$ and $1.5^{+1.1}_{-1.1}$ respectively. A second set of analyses targeted Higgs boson decays to $\gamma\gamma$, and the corresponding Run I signal strengths are $2.7^{+2.6}_{-1.8}$ for CMS and $1.4^{+2.2}_{-1.4}$ for ATLAS. The third set of analyses targeted leptonic final states from Higgs boson decays to WW^* , ZZ^* , or $\tau\tau$, with at least one Z , W or τ decaying leptonically. Despite the small branching ratio, the presence of one or two additional leptons from top quark decays leads to clean experimental signatures: two same-sign leptons or at least three leptons (electrons or muons), plus b -tagged jets. The signal strengths measured by CMS and ATLAS on $\sqrt{s} = 8 \text{ TeV}$ data in these channels are $3.7^{+1.7}_{-1.4}$ and $2.1^{+1.4}_{-1.2}$ respectively.

In this note we report on the search for $t\bar{t}H$ production in the multilepton signature with the first 2.3 fb^{-1} of 13 TeV data collected by the CMS experiment during the 2015 LHC Run II. The $t\bar{t}H$ cross section increases by about a factor 4 with the higher center of mass energy compared to 8 TeV [12]. The cross sections of the main backgrounds ($t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$ +jets) increase by roughly a factor 3. The general strategy remains similar to the 8 TeV search. Prompt leptons are identified with high efficiency and distinguished from non-prompt backgrounds using multivariate analysis methods. Events are then split in categories and the signal is extracted by means of a fit to the output of multivariate discriminants aimed at separating $t\bar{t}H$ from other processes. The fit is performed simultaneously in all categories. The object selections have been adapted to the 13 TeV running conditions, the lepton identification technique has been improved, new event kinematic variables have been considered in the multivariate approach for the signal extraction, and the presence of hadronically-decaying τ leptons has been taken into account in the event categorization.

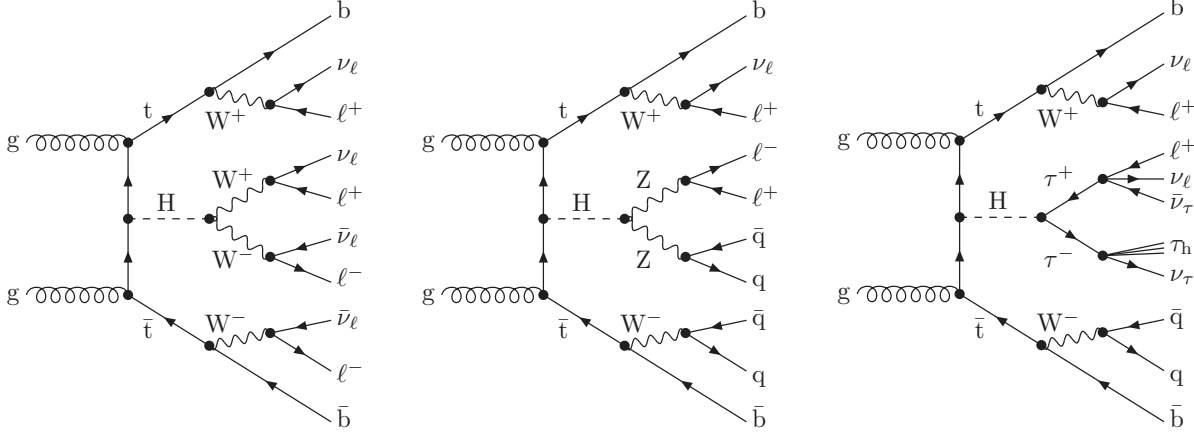


Figure 1: Feynman diagrams for $t\bar{t}H$ production at pp colliders, with the Higgs boson decaying to WW^* , ZZ^* , and $\tau\tau$ (from left to right). They represent examples of final states with four leptons, three leptons, and two same-sign leptons, respectively.

2 Data and simulated samples

The dataset used for this analysis has been collected in 2015 and corresponds to an integrated luminosity of 2.3 fb^{-1} . The events are selected by requiring the presence of either one, two, or three leptons (electrons or muons) at trigger level. The minimal transverse momentum (p_T) thresholds are 20 (23) GeV for muons (electrons) for the single lepton trigger, and 17 and 8 (12) GeV for muons (electrons) for the double lepton trigger. The three lepton triggers apply a threshold of 5 (9) GeV for muons (electrons) on the third lepton in order of p_T .

Simulated samples for the $t\bar{t}H$ signal and background processes are used for building the signal model, optimizing the event selection and estimating the systematic uncertainties. $t\bar{t}H$ events are simulated with Madgraph5_aMC@NLO [13], including up to one additional hadronic jet at NLO QCD accuracy. The same generator is used for the main backgrounds: $t\bar{t}W$, $t\bar{t}$ +jets, and $t\bar{t}\gamma$ +jets. Other minor backgrounds are simulated with different generators, such as POWHEG[14–19] and Madgraph at LO QCD accuracy. All generators are interfaced to PYTHIA8 [20] for the parton shower and hadronization steps. Pileup interactions are simulated with the multiplicity observed in data. All events are finally processed through a detailed simulation of the CMS detector based on GEANT4 [21], and reconstructed using the same algorithms used for the data. Moreover, they are required to satisfy the same trigger requirements as in the data.

3 Object reconstruction and identification

The CMS particle-flow (PF) algorithm [22, 23] provides a global interpretation of the event by combining the information from all sub-detectors to reconstruct and identify individual particles. They are classified into charged hadrons, neutral hadrons, photons, muons and electrons.

Particle-flow candidates are clustered into jets using the anti- k_T algorithm with a distance parameter of 0.4, as implemented in the FASTJET package [24, 25]. Charged hadrons that do not originate from the selected primary vertex are discarded. Jet energy corrections are then applied as a function of the p_T and pseudorapidity (η). Only jets with $p_T > 25 \text{ GeV}$, $|\eta| < 2.4$ and separated from any lepton candidate by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.4$ are retained.

Jets that are likely to originate from the hadronization of b quarks are identified by using a b-tagging algorithm [26]. Information about secondary vertices and track impact parameters are

combined in a likelihood discriminant (Combined Secondary Vertex, CSV). The efficiency to tag b jets and the probability to misidentify jets from light quarks or gluons are measured in data as a function of the jet p_T and η . We use two working points of the CSV output discriminant: a *loose* working point, with an efficiency of about 80% and a mistag rate of about 5%, and a *medium* working point with an efficiency of about 65% and a mistag rate of about 1% [27]. Simulated events are corrected for differences in the performance of the algorithm between data and simulation.

The missing transverse energy (E_T^{miss}) is calculated as the magnitude of the negative vector sum of transverse momenta of all reconstructed PF particles. On the other hand, the H_T^{miss} variable is defined as the magnitude of the negative vector sum of transverse momenta of all selected leptons and jets in the event. It has a worse resolution than E_T^{miss} , but it is also more resilient to energy depositions from pileup and the underlying event. Following the same strategy used in [8], we use a combination of the two variables in a linear discriminant, $E_T^{\text{miss}}\text{LD} \equiv 0.6E_T^{\text{miss}} + 0.4H_T^{\text{miss}}$. The requirement of $E_T^{\text{miss}}\text{LD} > 30\text{ GeV}$ has a similar signal efficiency to a requirement of $E_T^{\text{miss}} > 25\text{ GeV}$, but rejects about a factor of two more Drell-Yan background.

Muon candidates are reconstructed combining the information from both the silicon tracker and the muon spectrometer in a global fit [28]. An identification selection is performed using the quality of the geometrical matching between the tracker and the muon system measurements. Only muons within the muon system acceptance $|\eta| < 2.4$ and a minimum p_T of 5 GeV are considered.

Electrons are reconstructed using tracking and electromagnetic calorimeter information by combining ECAL superclusters and Gaussian sum filter (GSF) tracks [29]. We require electrons to have $|\eta| < 2.5$ to ensure that they are within the tracking volume and a minimum p_T of 7 GeV. The electron identification is performed using a multivariate discriminant built with shower-shape variables and track quality variables.

Further requirements are applied to reject electrons from photon conversions, by rejecting candidates with missing hits in the innermost layers of the tracking system or matched to a conversion secondary vertex candidate [29]. Moreover, we ensure that the lepton charge measurement is of good quality by requiring that the relative uncertainty on the muon p_T is less than 20%, and that the measurements of the electron charge from different observables based on the ECAL energy deposit and the reconstructed track are consistent [29]. We define *loose leptons* as those that pass the requirements described so far.

Throughout this note, we define *background leptons* as those originating from b hadron decays, from the misidentification of hadrons in jets, and from photon conversions. We define *signal leptons* as the isolated leptons coming from W , Z , and prompt τ decays. We developed advanced identification criteria to retain the highest possible efficiency for signal leptons while rejecting leptons from background processes. In particular, we use the following variables as an input to a multivariate discriminator based on a boosted decision tree (BDT):

- vertexing variables: the impact parameter in the transverse plane d_0 , the impact parameter along the z axis d_z , the three-dimensional impact parameter significance SIP_{3D} ,
- a particle flow isolation with variable cone size 0.05–0.2, dependent on the lepton p_T [30, 31],
- variables related to the closest jet to the lepton, such as the ratio between the p_T of the lepton and the p_T of the jet, the CSV b -tagging discriminator value of the jet, the

number of charged particles in the jet, and the p_T^{rel} variable:

$$p_T^{\text{rel}} = \frac{||(\vec{p}(\text{jet}) - \vec{p}(\ell)) \times \vec{p}(\ell)||}{||\vec{p}(\text{jet}) - \vec{p}(\ell)||},$$

- variables used in the identification of the electron and muon candidates: the muon segment compatibility, and the electron ID multivariate discriminant.

We define *tight leptons* as the loose leptons that satisfy a requirement on the discriminator value.

Hadronically decaying taus (τ_h) are reconstructed using the hadron-plus-strips algorithm [32]. τ_h candidates are required to pass the “decay mode finding” discriminator, either being reconstructed in 1- or 3-prong decay modes with or without additional π^0 s. In addition, they have to fulfill $p_T > 20 \text{ GeV}$, $|\eta| < 2.3$, and isolation requirements.

4 Event selection

The event selection aims at rejecting final state signatures that do not match that of the signal. Each event is required to have at least two leptons with $p_T > 10 \text{ GeV}$ passing the tight selection requirements, and the leading lepton must have $p_T > 20 \text{ GeV}$. Moreover, events with a pair of leptons passing the loose selection requirement and with an invariant mass of less than 12 GeV are rejected, as they are not accurately modeled by the simulation. As signal events contain a pair of top quarks, we require that at least two selected jets are present in the final state. At least two of them are required to pass the loose working point of the CSV b-tagging discriminator, or at least one to pass the medium working point.

If no additional lepton with a p_T greater than 10 GeV passes the tight selection requirements, the event is tagged as a candidate for the two lepton same-sign (2lss) category of the analysis. It is accepted in that category only if the two tight leptons have the same charge, and if it contains at least four jets. If the sub-leading tight lepton is an electron, the p_T threshold is tightened to 15 GeV . The two selected leptons are required to pass criteria aimed at rejecting leptons from conversions, and on the quality of the charge measurement. To suppress further backgrounds from $Z \rightarrow ee$ with a misidentified charge, in the dielectron final states events are also rejected if the dilepton invariant mass is within 10 GeV of the Z boson mass, and a requirement of $E_T^{\text{miss}} \text{LD} > 30 \text{ GeV}$ is applied.

If, on the other hand, at least three tight leptons with p_T greater than 10 GeV are found, the event is tagged as a candidate for the three-lepton category. Background processes with Z bosons in the final state are reduced by applying the same invariant mass veto used for the 2lss category and by requiring that $E_T^{\text{miss}} \text{LD} > 30 \text{ GeV}$. The $E_T^{\text{miss}} \text{LD}$ threshold is tightened to 45 GeV if the event has a pair of opposite-sign and same-flavor leptons, but is not applied if the event has at least four jets. Finally, the event is rejected if any of the first three leptons do not pass the conversion veto requirements, or if the sum of their charges is not equal to $+1$ or -1 . Table 1 summarizes the event yields observed in data in each category, and the corresponding expectations from simulated events and from data-driven predictions for non-prompt and charge mis-identified leptons, discussed in Section 5.4.

After the selection described above, the yields are still dominated by background. The selected events are thus further categorized. The 2lss events are divided according to the presence of a τ_h in the final state. Furthermore, events where no τ_h is found are split according to lepton flavor into dielectron, dimuon and electron-muon sets. These sets, with the exception of the dielectron one, are then further split according to the presence of at least two jets passing the

Table 1: Expected and observed yields after the selection in 2lss and three-lepton final states. The rare SM backgrounds include ZZ , $W^\pm W^\pm qq$, WW produced in double-parton interactions, and triboson production. Uncertainties are statistical only. The backgrounds from non-prompt leptons and charge flips are extracted from data.

	$\mu\mu$	ee	$e\mu$	3ℓ
$t\bar{t}W$	3.22 ± 0.16	1.47 ± 0.11	4.95 ± 0.19	2.56 ± 0.14
$t\bar{t}Z/\gamma^*$	0.82 ± 0.03	1.14 ± 0.14	2.42 ± 0.17	3.75 ± 0.18
WZ	0.09 ± 0.05	0.06 ± 0.06	0.25 ± 0.11	0.33 ± 0.11
$tttt$	0.19 ± 0.03	0.11 ± 0.02	0.28 ± 0.03	0.22 ± 0.03
tZq	0.10 ± 0.06	0.00 ± 0.00	0.12 ± 0.13	0.44 ± 0.17
rare SM bkg.	0.06 ± 0.03	0.04 ± 0.04	0.13 ± 0.06	0.16 ± 0.59
non-prompt (data)	3.99 ± 0.38	3.58 ± 0.38	10.10 ± 0.65	8.08 ± 0.67
charge mis-ID (data)		1.11 ± 0.05	1.65 ± 0.05	
all backgrounds	8.47 ± 0.42	7.52 ± 0.44	19.90 ± 0.73	15.55 ± 0.95
$t\bar{t}H$ signal	1.53 ± 0.08	0.69 ± 0.05	2.27 ± 0.10	2.12 ± 0.09
data	9	11	11	28

medium CSV b-tagger working point (b-tight) or just passing the minimum CSV requirements of the event selection (b-loose). On the other hand, three-lepton events are only separated into b-tight and b-loose categories. Finally, all resulting categories are split by positive or negative sum of charges of the selected leptons, with the exception of the 2lss category with τ_h which is not further split. This is motivated by the charge asymmetry in several background processes.

In order to further improve the sensitivity of the analysis, topological and kinematic differences between $t\bar{t}H$ signal and background events are exploited by means of two BDTs, aimed at discriminating the signal from the $t\bar{t}$ and $t\bar{t}V$ processes respectively. The BDTs are trained separately for 2lss and three-lepton events. In the 2lss category, for the training against the $t\bar{t}$ background the input variables are the following: the maximum $|\eta|$ of the two leading leptons, the multiplicity of hadronic jets, the minimum distance between the leading lepton and closest jet, the minimum distance of the trailing lepton and closest jet, the missing transverse energy, the average separation between the two jets, and the transverse mass of the leading lepton and missing transverse energy. For the training against the $t\bar{t}V$ process, the missing transverse energy and the average jet-jet separation are replaced in the above list by the p_T of the leading and trailing among the selected leptons. The training against $t\bar{t}$ in the three-lepton category differs from that used in 2lss by the usage of H_T^{miss} in the place of E_T^{miss} , while the one against $t\bar{t}V$ uses the same input variables as in 2lss.

The output of the BDT discriminators is used simultaneously to divide each category in bins of different S/B. The signal extraction is performed by fitting its normalization from the distribution of events among these bins. Figures 2 and 3 show the distributions of the BDT outputs and the number of jets for selected events in data and simulation.

5 Signal and background modeling and systematic uncertainties

5.1 Signal model

Simulated samples are used to understand the event selection efficiency and build the discriminants ultimately used for signal extraction. Systematic uncertainties associated with the correction factors applied to better match the detector performance in data and with the theoretical model are taken into account for the signal simulation.

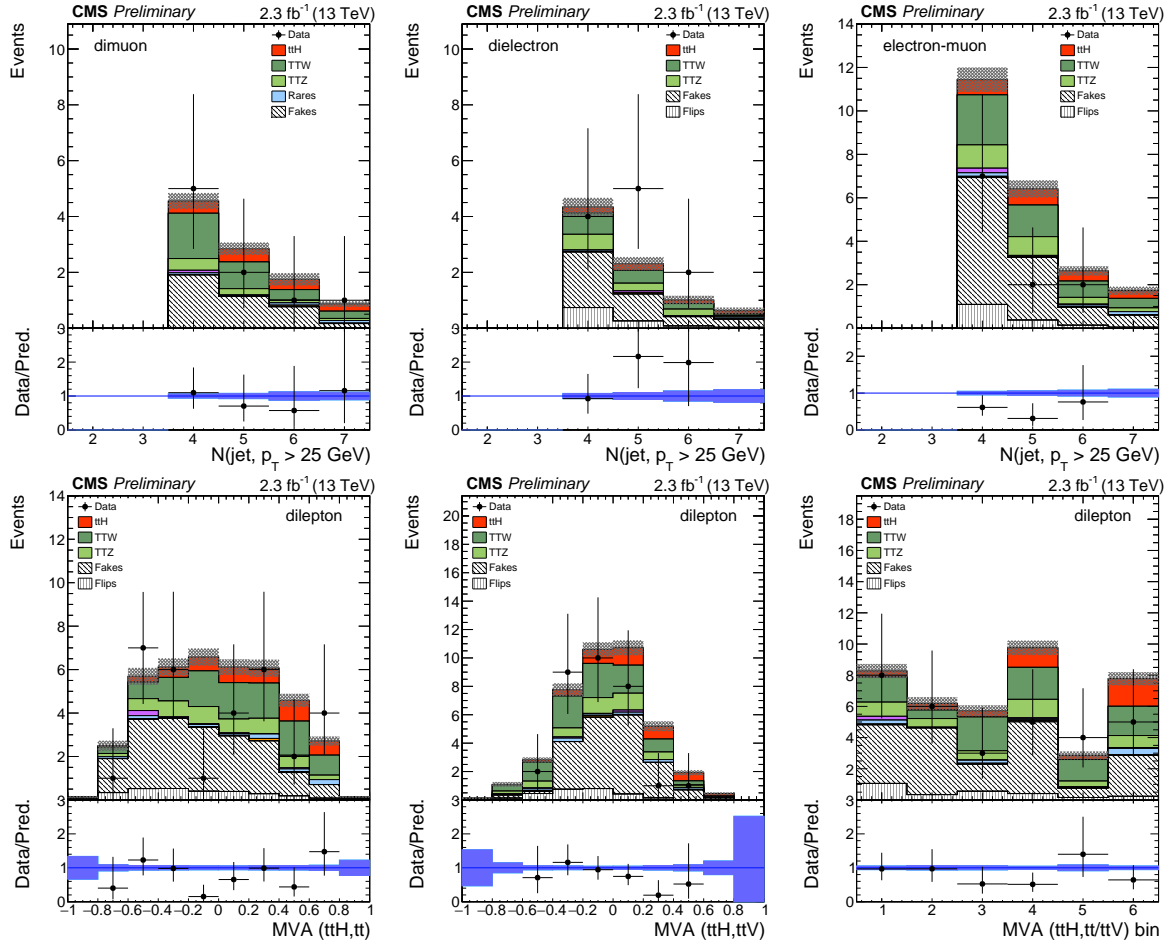


Figure 2: Top row: number of selected jets in the same-sign $\mu\mu$, ee , and $e\mu$ channels. Bottom row: distributions of the BDT kinematic discriminants and result of their combination in the bins used for signal extraction, for the two lepton same-sign selection inclusive in lepton flavor. Pre-fit distributions are shown. Uncertainties are statistical only.

We use scale factors to correct for differences in the lepton trigger, reconstruction and identification performance between data and simulation. The magnitude of the systematic uncertainty on the lepton corrections is about 5%. We assess the impact of the uncertainties associated with the jet energy corrections [33] by shifting the correction factors up and down by 1σ and re-calculating all kinematic quantities. Systematic effects both on normalization and shape are taken into account in extracting the results. The uncertainty on the jet energy resolution has a negligible impact on the analysis. The uncertainties on the correction for the data to simulation differences in the b-tagging performance are parametrized as a function of p_T , η , and flavor of the jet. We assess their effect on the analysis by shifting the correction factor for each jet up and down by 1σ of the appropriate uncertainty, and recalculating the overall event scale factor.

The theoretical uncertainties on the NLO prediction for the inclusive $t\bar{t}H$ production cross section amount to $+5.8/\,-9.2\%$ from unknown higher orders in the perturbative series and 3.6% from the knowledge of the parton distribution functions (PDFs) and α_s [12]. These uncertainties are propagated to the final normalization of the signal yields. In addition to the overall normalisation, a systematic uncertainty on the shape of the final discriminating variables is estimated by varying the normalisation and factorisation scales in generated events, and has an effect of 2–3% on the normalisation.

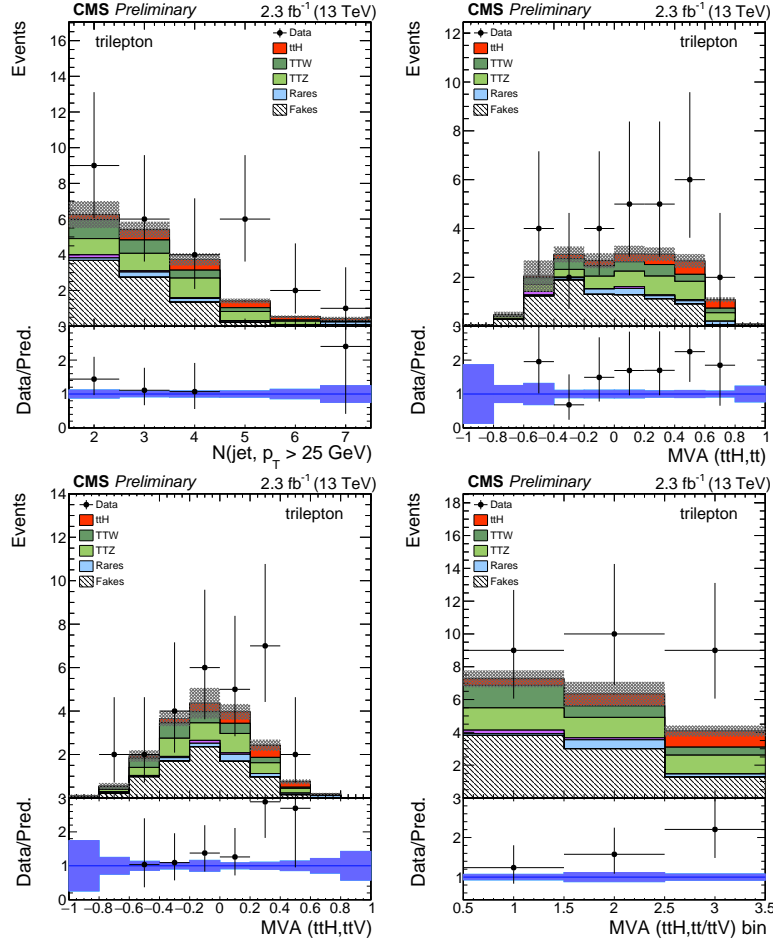


Figure 3: Number of selected jets, distributions of the BDT kinematic discriminants and result of their combination in the bins used for signal extraction, for the three-lepton channel. Pre-fit distributions are shown. Uncertainties are statistical only.

5.2 Irreducible backgrounds

Irreducible backgrounds from $t\bar{t}W$ and $t\bar{t}Z$ are estimated from simulated events. The same corrections used for the signal are applied to account for the different performances between data and simulation. The inclusive production cross sections for the $t\bar{t}W$ and $t\bar{t}Z$ processes are calculated at NLO QCD and NLO EWK [12, 13, 34], with theoretical uncertainties from unknown higher orders of 12% and 10% respectively, and uncertainties from the knowledge of the PDFs and α_s of 2 to 4%. In addition to the overall normalisation, systematic uncertainties of theoretical origin on the distribution of the events in the final discriminating variables are considered. They are estimated by varying the normalisation and factorisation scales up and down, and their amplitude is found to be of 2% to 4%.

5.3 Diboson processes

The WZ and ZZ production processes with the gauge bosons decaying to leptons can yield the same final states as the signal. Out of the two, WZ is the one giving the largest contribution. The inclusive production cross sections for these processes have been measured at the LHC and are found to be in agreement with the NLO predictions. However, the agreement does not translate directly to the phase space of the signal region used in this analysis, as we require the

presence of at least one additional b jet.

It is possible to obtain a clean sample of WZ events with three leptons and at least two jets in the final state by requiring the presence of a lepton pair compatible with the Z decay and vetoing events with a loose b-tagged jet, as dibosons are preferentially produced in association with jets from light quarks or gluons. A scale factor is extracted by normalizing the yield of simulated events in this control region to the data, and then applied to the diboson prediction in the signal region. The majority of diboson events in the signal region passes the event selection because they contain jets from light quarks or gluons that are mistagged as b jets. Therefore, this estimate is sensitive to the experimental uncertainty on the mistag rate, while reducing the impact from theoretical uncertainties in the jet flavor composition due to higher-order QCD terms.

The overall uncertainty is composed of the statistical uncertainty on the control region (30%), the residual backgrounds in the control region (20%), the uncertainties on the b-tagging rate (ranging from 10 to 40%), and from the knowledge of the PDFs and the theoretical uncertainties on the extrapolation (up to 10%).

5.4 Non-prompt and charge mis-identified leptons

The yield of events where one of the selected leptons does not originate from the decay of a W, Z or H boson is estimated from data using a fake-rate method. The probability for a non-prompt candidate to pass the tight selection requirements is first measured in a control sample where the candidate population is dominated by background leptons (as defined in Section 3), and parametrized as a function of its p_T and $|\eta|$ separately for muons and electrons. For muon candidates and electron candidates above 30 GeV QCD multijet events are used, while lower p_T electrons are taken from Z+jets events to overcome limitations posed by the trigger used to collect the data for this fake-rate determination. In both cases, the residual contamination from prompt leptons originating from W and Z decays is subtracted using the transverse mass as the discriminating variable and vetoing the presence of additional leptons.

A control region is then defined relaxing the requirement applied on the BDT used for lepton identification. Events in this control region where only one lepton does not pass the tight selection are then weighted by $f/(1-f)$, where f is the mis-identification probability measured above, while those where two leptons fail the tight requirements are weighted by $-f_1 f_2 / [(1-f_1)(1-f_2)]$. The resulting yield is used as the background prediction in the signal region, with an uncertainty of 30 to 50% arising from the statistical and systematic uncertainties in the applied weights. The latter are derived from comparing alternative techniques for subtracting prompt leptons and from closure tests of the method performed in simulated background events.

Following a similar logic, the probability for mis-reconstructing the electron charge is evaluated in a sample of electrons from Z decays. They are split in observed opposite- and same-sign pairs, and from the relative abundance of these sets a charge mis-reconstruction probability is measured as a function of the electron p_T and $|\eta|$. The probability ranges from 0.03% in the barrel to about 0.4% in the endcap regions. A systematic uncertainty of 30% is evaluated from closure tests conducted in the simulation.

6 Results

The event yields are compared with the expectation from the background processes and a 125 GeV SM Higgs boson. A signal strength parameter $\mu = \sigma/\sigma_{\text{SM}}$ is introduced, and the

expected yields from $t\bar{t}H$ are scaled by its value without any modification to the branching fractions or to the kinematic properties of the events.

Results in terms of the asymptotic 95% CL upper limit [35–38] on μ are presented in Table 2. The observed limits are compared with the expected ones under the background-only hypothesis. The best fit of the signal strength is presented in Table 3, and the content of the tables is summarized in Fig. 4. The measured signal strengths in the same-sign dilepton and three-lepton categories are compatible with each other at the level of 1.8σ .

Table 2: Asymptotic 95% CL upper limits on the signal strength parameter.

Category	Observed limit	Expected limit $\pm 1\sigma$
same-sign dileptons	2.1	2.7 (+1.4) (−0.9)
trileptons	11.7	5.4 (+2.9) (−1.8)
combined	3.3	2.6 (+1.3) (−0.8)

Table 3: Best fit of the signal strength parameter.

Category	μ best fit $\pm 1\sigma$
same-sign dileptons	−0.5 (+1.0) (−0.7)
trileptons	5.8 (+3.3) (−2.7)
combined	0.6 (−1.1) (+1.4)

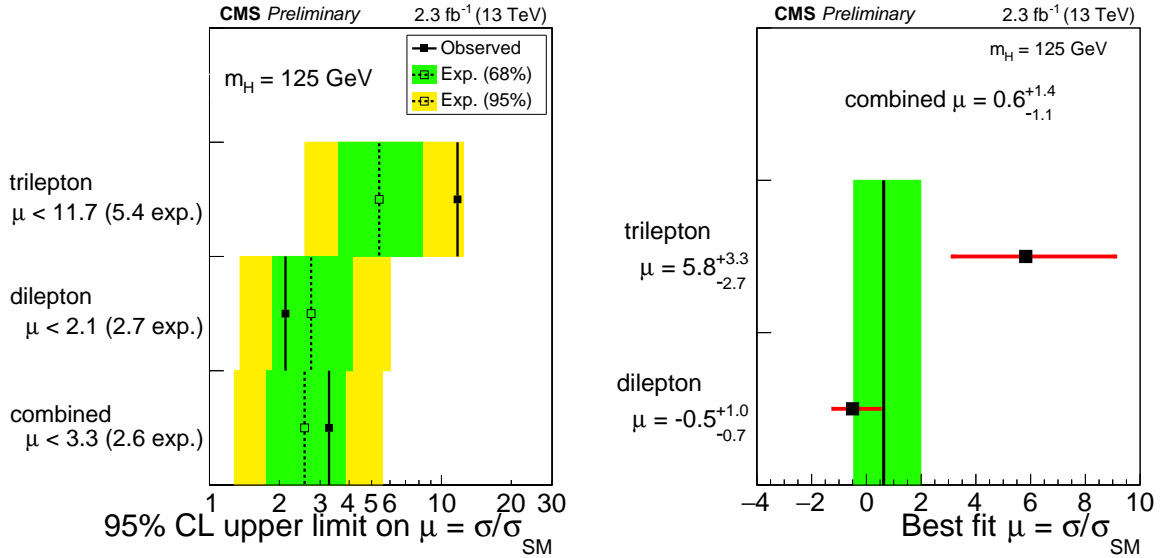


Figure 4: Left: asymptotic 95% CL upper limits on the signal strength parameter. Right: best fit of the signal strength parameter.

7 Conclusions

A search for SM $t\bar{t}H$ production has been conducted using pp collision data collected by the CMS experiment in 2015 at a center of mass energy of $\sqrt{s} = 13$ TeV, and corresponding to an integrated luminosity of 2.3 fb^{-1} . The analysis targets the WW^* , ZZ^* and $\tau\tau$ decay channels of the Higgs boson by selecting final states with two same-sign leptons or more than three leptons, produced in association with b jets. Multivariate techniques are used to distinguish prompt leptons from mis-identified jets and to separate the signal from other physics processes. The signal strength is measured to be $0.6^{+1.4}_{-1.1}$ times the SM expectation. The observed 95%

confidence level upper limit on the $t\bar{t}H$ production cross section is 3.3 times the SM expectation, compared to the 2.6 expected in absence of a signal.

References

- [1] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [2] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [3] ATLAS and CMS Collaborations, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.
- [4] ATLAS Collaboration, “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”, *Eur. Phys. J.* **C76** (2016), no. 1, 6, doi:10.1140/epjc/s10052-015-3769-y, arXiv:1507.04548.
- [5] CMS Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, *Eur. Phys. J.* **C75** (2015), no. 5, 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.
- [6] ATLAS and CMS Collaborations, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”, ATLAS-CONF-2015-044, CMS-PAS-HIG-15-002, 2015.
- [7] LHC Higgs Cross Section Working Group Collaboration, “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, doi:10.5170/CERN-2011-002, arXiv:1101.0593.
- [8] CMS Collaboration, “Search for the associated production of the Higgs boson with a top-quark pair”, *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106, arXiv:1408.1682. [Erratum: JHEP10,106(2014)].
- [9] ATLAS Collaboration, “Search for $H \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector”, *Phys. Lett.* **B740** (2015) 222–242, doi:10.1016/j.physletb.2014.11.049, arXiv:1409.3122.
- [10] ATLAS Collaboration, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J.* **C75** (2015), no. 7, 349, doi:10.1140/epjc/s10052-015-3543-1, arXiv:1503.05066.
- [11] ATLAS Collaboration, “Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector”, *Phys. Lett.* **B749** (2015) 519–541, doi:10.1016/j.physletb.2015.07.079, arXiv:1506.05988.

- [12] LHC Higgs Cross Section Working Group. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV?rev=15>.
- [13] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [14] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **0411** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [15] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **0711** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [16] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **1006** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [17] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J.* **C71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.
- [18] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”, *JHEP* **0909** (2009) 111, doi:10.1007/JHEP02(2010)011, 10.1088/1126-6708/2009/09/111, arXiv:0907.4076.
- [19] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, “ W^+W^- , WZ and ZZ production in the POWHEG BOX”, *JHEP* **1111** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.
- [20] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [21] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, doi:10.1109/TNS.2006.869826.
- [22] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- [23] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-002, 2010.
- [24] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [25] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys. Lett.* **B641** (2006) 57–61, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- [26] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

- [27] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.
- [28] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.
- [29] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [30] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic Top Quarks at the LHC”, *JHEP* **03** (2011) 059, doi:10.1007/JHEP03(2011)059, arXiv:1007.2221.
- [31] CMS Collaboration, “Search for SUSY in same-sign dilepton events at $\sqrt{s}=13$ TeV”, CMS Physics Analysis Summary CMS-PAS-SUS-15-008, 2015.
- [32] CMS Collaboration, “Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS”, *JINST* **11** (2016), no. 01, P01019, doi:10.1088/1748-0221/11/01/P01019, arXiv:1510.07488.
- [33] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [34] S. Frixione et al., “Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons”, *JHEP* **06** (2015) 184, doi:10.1007/JHEP06(2015)184, arXiv:1504.03446.
- [35] A. L. Read, “Presentation of search results: The CL(s) technique”, *J. Phys.* **G28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.
- [36] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [37] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: *Eur. Phys. J.*C73,2501(2013)].
- [38] ATLAS and CMS Collaborations, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in summer 2011”, ATL-PHYS-PUB-2011-011, CMS-NOTE-2011-005, 2011.