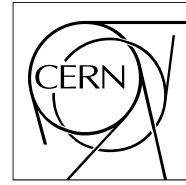


The Compact Muon Solenoid Experiment
Analysis Note

The content of this note is intended for CMS internal use and distribution only



17 January 2019 (v9, 25 March 2019)

Probing Effective Field Theory Models using Associated Top Quark Production in Multiple Lepton Final States at 13 TeV

Kevin Lannon, Andrew Wightman, Geoffrey Smith, Reza Goldouzian, Michael Hildreth, Brian Winer, Wuming Luo, Anthony Lefeld

Abstract

A search for new physics in the context of effective field theory models is presented. The analysis targets associated top quark production in final states with multiple leptons, utilizing the full pp collision dataset collected by the CMS experiment in 2017.

Contents

1	1	Introduction	2
2	2	Top Quark EFT	3
3	2.1	Oveview	3
4	2.2	Dim6TopEFT Model	4
5	2.3	Operator Selection	4
6	3	Data and MC Samples	5
7	3.1	Data Samples and Triggers	8
8	3.2	Signal Samples	8
9	3.3	Background Samples	11
10	4	Object reconstruction and identification	12
11	4.1	Lepton Identification	12
12	4.2	Jets and B-tagging	15
13	5	Event selection and categorization	18
14	5.1	$2\ell ss$ category	18
15	5.2	3ℓ category	18
16	5.3	4ℓ category	19
17	5.4	Separation into sub-categories	19
18	6	Data-to-MC corrections	19
19	6.1	Pileup reweighting	20
20	6.2	Trigger efficiency	20
21	6.3	Identification and isolation efficiency for e and μ	21
22	6.4	b-tag efficiency and mistag rate	21
23	6.5	ECAL prefiring correction	22
24	7	Background estimation	23
25	7.1	“Fake” background	23
26	7.2	Charge “flip” background	25
27	8	Signal extraction and EFT model parameters fitting	29
28	9	Systematic uncertainties	31
29	10	Results	37
30	10.1	Standard Model Signals	37
31	10.2	EFT Signals	38
32	A	Synchronization Plots	55
33	B	MC Validation	58
34	C	A comparison of the Limits	68
35	D	Additional plots requested during review	70
36			

1 Introduction

There are many motivations to search for new particles or interactions at the LHC: The existence of dark matter and dark energy implies that full list nature's constituents has not yet been discovered. Likewise, an explanation the observed asymmetry between matter and antimatter in the universe requires a new source of CP violation. Most solutions to the hierarchy problem also rely on new particles. Nonetheless, there is no guarantee that new particles exist in the mass range directly accessible at the LHC. To extend the discovery reach of the LHC, it is therefore prudent to consider not only direct searches for new particles, but also indirect means of probing higher energy scales.

One flexible framework for undertaking such indirect probes is that of effective field theory (EFT). In principle, an EFT is a low energy approximation for more fundamental theory involving particles of mass scale Λ . This would seem to imply that EFTs are not helpful until the higher energy theory is known. However, in practice, an EFT can be constructed by adding terms to the standard model (SM) lagrangian that have dimension higher than four, but otherwise respect the symmetries and conservation laws observed in nature. The additional terms are constructed from *operators*—products of fields—that involve only SM fields. The contribution of an operator of dimension- n to the lagrangian is suppressed by a factor of $1/\Lambda^n$ implying that the focus should be placed on operators of the lowest possible dimension. As dimension-five operators tend to produce lepton number violation, this analysis neglects these operators in favor of dimension-six operators as the leading contributors.

Although the impact of an EFT can in principle be detected in a large variety of experimental observables, it is interesting to consider the impact of EFT on the production of one or more top quarks in association with one or more W, Z, or H bosons. It has long been speculated that the large mass of the top quark, and hence its large coupling to the Higgs boson, might be a hint at a special relationship between the top quark and the physics of electroweak symmetry breaking. If so, the production of top quarks along with electroweak or Higgs bosons may shed some light on this mystery. Only recently have experimental measurements started to test directly the coupling of the top quark to Z[1, 2] and H [3] bosons. The current and future CMS datasets will thus provide an intriguing opportunity to study these processes in more detail.

Collisions producing one or more top quarks plus an additional boson produce a variety of signatures involving multiples jets (including bottom quark jets) and also possibly leptons. Top quarks decay with almost 100% branching fraction to a bottom quark and a W boson, and W bosons (both those originating from top quarks and those produced in addition to top quarks), Z bosons, and H bosons decays in various ways involving quarks (including quarks, especially for the Higgs boson) or leptons. Ultimately, the final state signatures are primarily determined by how any present bosons decay, either hadronically or leptonically. Final states in which multiple bosons decay leptonically provide a number of experimental advantages. Multiple leptons provide an efficient trigger strategy which remains viable even at high instantaneous luminosity. Furthermore, final states involving bottom quark jets, other jets, and either a same-sign dilepton pair or three or more leptons have very few backgrounds. This analysis focuses on the same-sign dilepton and three or more lepton final states, which will be referred to for the rest of the paper as *multilepton* final states.

Although experimentally beneficial, multilepton final states do pose challenges for EFT-based searches. A common EFT analysis strategy is to measure a differential cross section for a specific physics process and unfold this measurement back to the particle level for comparison with EFT predictions. For the multilepton final states, this approach is complicated by the fact that multiple physics processes contribute, and it has so far proven impossible to separate these

contributions. For example, events in the three-lepton final state where one pair of leptons is a same-flavor, opposite-sign pair with an invariant mass within the Z peak may come from either the strong production of a top quark pair or the electroweak production of a single top quark in association with a Z boson. Likewise same-sign dilepton and trilepton final states outside the Z peak originate with equal probability from $t\bar{t}W$ or $t\bar{t}H$ production. Each of these physics processes can receive contributions from EFT operators. Sometimes the same operator will affect multiple processes in different ways. In other cases, different processes are subject to different sets of operators. To address this challenge, instead of unfolding a differential cross section, this analysis incorporates the effects of EFT operators on all relevant processes into the predictions for observed detector yields.

The basic strategy employed by this analysis is as follows: Multilepton events are divided into categories based on the number, flavor, and signs of the leptons present. The lepton categories are further subdivided according to the number of b-jets. Within each lepton and b-jet category, the event yields are characterized as a function of the number of jets. These event yields define the observables for the analysis. These event yields are compared against predictions that incorporate the effects of EFT operators. Contributions involving primarily prompt leptons—including the signal processes—are modeled using simulated samples. Where relevant, the predicted yields for processes impacted by EFT operators are parameterized in terms of the Wilson coefficients for those operators. Predictions for background processes involving primarily non-prompt leptons (e.g. leptons from bottom or charmed hadron decays or detector fakes) are based on extrapolations from control regions selected from the data. Wilson coefficients are varied in order to determine the best fit of the predictions to data, as well as to establish intervals in Wilson coefficients over which the predicted yields are consistent with observation at the two standard deviation level. Further details on each of the above aspects of the analysis are given in the subsequent sections.

2 Top Quark EFT

This section describes the specific top quark EFT model used in this analysis and explains some of the basic principles on which this analyses relies.

2.1 Overview

The basic idea of an EFT approach is to incorporate new physics (NP) interactions into the SM in a model-independent approach. This can be done by assuming the NP occurs at an energy scale Λ that is much greater than the energy scale that is being studied. In this case the NP interactions can be integrated out from the lagrangian to give an effective lagrangian of the form:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda} \mathcal{L}_1 + \frac{1}{\Lambda^2} \mathcal{L}_2 + \dots \quad (1)$$

where \mathcal{L}_{SM} is the SM lagrangian density of dimension four, \mathcal{L}_1 is the NP interactions of dimension five, \mathcal{L}_2 is of dimension six, etc. The higher dimension interactions can be written down in the form:

$$\mathcal{L} = \sum_i \frac{c_i}{\Lambda^{d-4}} \mathcal{O}_i \quad (2)$$

where \mathcal{O}_i are the effective operators constructed purely from SM fields, c_i are the dimensionless Wilson coefficients (WC) that parameterize the strength with which the NP couples to the SM particles, and d is the dimensionality of the corresponding operators.

The matrix element can then be written as the sum of SM and NP components:

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \sum_i c_i \mathcal{M}_i \quad (3)$$

the cross section is proportional to the square of the matrix element, and has the following form

$$\begin{aligned} \sigma(c_1, c_2) &\propto |\mathcal{M}|^2 \\ &\propto s_0 + s_{01}c_1 + s_{02}c_2 + s_{12}c_1c_2 + s_{11}c_1^2 + s_{22}c_2^2 \end{aligned} \quad (4)$$

the s_{ij} are structure constants of a multi-dimensional quadratic function, which can be solved for by evaluating the cross section at six values of (c_1, c_2) . These can then be used to parameterize a fit which can extrapolate to any combination of (c_1, c_2) . This procedure generalizes to any number of EFT operators and also applies to the differential cross section.

In order to reduce the total number of samples needed to parameterize the WC fits, MADGRAPH event reweighting is used. A set of hard events are generated and for each event a set of rescaling factors are calculated and stored along with the normal event information. Since each individual event represents an infinitesimal section of the inclusive cross section, the rescaling factors also follow a quadratic parameterization based on the WC strengths. The same fitting procedure as outlined above can then be used to generate a fit for every event in the sample. Using these fits, every distribution described in the analysis can be consistently reweighted to an arbitrary point in the WC phase space. This is implemented by using a custom `TH1EFT` histogram class, which stores and keeps track of each event fit on a bin by bin basis.

One important limitation of the MADGRAPH reweighting procedure is that it is unable to reweight a sample to a section of phase space that is not already occupied by the original set of hard events. This is particularly important in the case of an EFT analysis, where the inclusion of new EFT operators leads to Feynman diagrams not present in the SM. Each signal sample is therefore generated starting at a non-SM point, so as to ensure that as much of the relevant phase space is populated as possible.

2.2 Dim6TopEFT Model

The EFT model used by MADGRAPH to generate the signal samples was the dim6TopEFT model [4]. Some of the important aspects of this implementation are:

- The Warsaw basis of dimension-six operators is used.
- Λ is conventionally fixed to 1 TeV.
- The CKM matrix is assumed to be a unit matrix
- The masses of u, d, s, c, e, μ fermions are set to zero by default
- The unitary gauge is used and Goldstone bosons are removed
- Baryon and lepton number violating operators are not included
- Only tree-level simulation is possible.

2.3 Operator Selection

There are 59 independent dimension-six operators that conserve baryon and lepton number. In order to reduce the phase space of all possible NP coupling strengths, we consider only a subset of the dimension-six operators. The excluded operators have their couplings fixed to SM values.

In this analysis only operators which directly appear in signal processes at tree level and with at most one insertion of a NP vertex are considered. Operators which have a disproportionately large impact on background processes are excluded or had the range of considered values restricted. To determine an operator's impact on a given process, a set of reference samples that only vary a single WC at a time are used.

The operators are selected based on their overall potential impact to the signal processes considered. The reference samples are used to determine the degree to which each operator individually scales a particular process cross section. The range of values for the operators is always required to be in the range $[-16\pi^2, 16\pi^2]$. Additionally, each operator is not allowed to modify the $t\bar{t}$ inclusive cross section by more than 15%. For a given operator, the maximum scaling for each process is summed together, weighted by the expected sensitivity to that process. The sum is used as a figure of merit (FOM) to rank each operator:

$$\text{FOM} = \sum \frac{|\mu_i - 1|^2}{\sigma_i^2} \quad (5)$$

Where μ_i is the expected scaling of a given signal process for a particular WC and σ_i is the expected sensitivity. The FOM is then used to sort and pick the operators that are most relevant to the analysis. Table 1 shows the range of values considered along with the corresponding FOM. Only the top 16 WCs were chosen, with cQq83 and cQq13 excluded in favor of ctW and ctZ.

The set of selected operators can be divided into two groups. The first nine are the so-called 'two heavy + boson', whose WC include: ctp, cpQM, cpQ3, cpt, cptb, ctW, ctZ, cbW, and ctG. The other group consists of the 'two heavy + two lepton', which includes: cQl3, cQlM, cQe, ctl, cte, ctlS, and ctlT. For the second set, there is a different WC for each lepton flavor for a total of 21 degrees of freedom. We make the assumption that these NP operators couple to each lepton flavor equally and thus require the WC strengths to be equal for each lepton flavor. This reduces the second set of operators to seven.

3 Data and MC Samples

In the following we list as reference the set of samples of the RunIIFall17MiniAOD (CMSSW 94X) campaign used for the studies and results presented in this analysis.

Because of the new techniques used in this analysis, a strategic decision had to be made about which datasets to include. As described in detail elsewhere in this note, this analysis introduces a number of techniques not previously used in CMS analysis, in particular the technique of parameterizing the weights as a function of Wilson coefficients event-by-event so that we can extract the dependence of the observed yields on the Wilson coefficients. Ultimately, we decided to focus the analysis solely on the 2017 dataset and associated MC samples. There were two primary concerns that drove this decision.

On the technical side, we wanted to focus on having a single dataset that we could work with to minimize the challenges of implementing this analysis. A particular challenge is the need to produce privately the signal MC samples. Each of the 2016, 2017, and 2018 datasets require different MC samples to model the changing detector conditions, so to analyze the full Run 2 dataset would require that we produce three separate MC samples for our signal. On the time scale of this result, it was only feasible to produce a single sample. When we started this analysis, the only feasible choices were the 2016 and the 2017 datasets, and we chose the 2017

WC	FOM	ttH	ttll	ttl ν	ttlq	tHq
ctTi	4.3e+07	-	[-8.2,8.2]	[-157.9,157.9]	[-22.9,22.9]	-
cbW	4.1e+05	[-38.4,38.4]	[-38.4,38.4]	-	[-6.7,6.7]	[-2.6,2.6]
cptb	3.0e+05	[-108.2,108.2]	[-128.6,128.6]	-	[-25.4,25.4]	[-9.2,9.2]
cpQM	1.3e+05	[-157.9,157.9]	[-27.8,52.6]	[-157.9,157.9]	[-70.4,51.0]	-
ctp	1.2e+05	[-20.1,52.1]	[-47.3,79.2]	[-157.9,157.9]	[-78.2,80.6]	[-17.5,19.4]
cpQ3	1.2e+05	[-41.3,39.9]	[-50.4,39.9]	[-50.4,39.9]	[-13.6,8.1]	[-4.5,4.7]
cpt	1.1e+05	[-157.9,157.9]	[-47.4,30.7]	[-157.9,157.9]	[-97.0,83.8]	-
cQei	5.5e+04	-	[-43.5,43.9]	-	[-93.1,93.3]	-
cQlMi	5.5e+04	-	[-43.0,44.8]	[-157.9,157.9]	[-94.5,91.9]	-
ctli	5.4e+04	-	[-43.5,44.3]	[-157.9,157.9]	[-157.9,157.9]	-
ctei	5.3e+04	-	[-43.4,44.7]	-	[-157.9,157.9]	-
cQl3i	4.4e+04	-	[-157.9,157.9]	[-157.9,157.9]	[-28.4,28.0]	-
ctG	1.5e+04	[-2.5,0.4]	[-3.8,0.4]	[-7.1,0.4]	-	-
ctlSi	1.4e+04	-	[-61.5,61.5]	[-157.9,157.9]	[-157.9,157.9]	-
cQq83	4.4e+03	[-7.4,6.7]	[-7.4,6.7]	[-3.9,4.4]	[-4.5,4.5]	[-2.8,2.8]
ctW	4.3e+03	[-10.6,9.6]	[-9.9,9.6]	[-10.6,9.6]	[-5.9,5.5]	[-2.5,2.0]
ctZ	4.0e+03	[-12.3,13.0]	[-6.9,7.0]	[-12.3,13.0]	[-12.3,13.0]	-
cQq13	3.5e+03	[-3.4,3.1]	[-3.4,3.1]	[-2.0,1.9]	[-1.9,2.5]	[-1.2,1.2]
cQq81	1.1e+03	[-9.8,5.0]	[-9.8,5.0]	[-8.0,5.0]	-	-
ctq8	9.2e+02	[-9.7,4.4]	[-9.7,4.4]	[-8.0,4.4]	-	-
cQq11	5.3e+02	[-3.3,3.2]	[-3.3,3.2]	[-3.0,3.0]	-	-
ctq1	4.8e+02	[-3.3,3.2]	[-3.3,3.2]	[-3.0,3.0]	-	-
cQd8	9.6e+01	[-14.2,8.9]	[-14.2,8.9]	-	-	-
cQu8	9.1e+01	[-11.5,6.7]	[-11.5,6.7]	-	-	-

Table 1: Range of Wilson Coefficient values considered. The ranges represent the WC values where $\sigma_{\text{NP}}/\sigma_{\text{SM}} < 5$ for each signal process considered. A hard cut-off at $[-16\pi^2, 16\pi^2]$ is placed on all coefficients. These ranges are also used to determine resonable WC values for the initial set of hard events and also the set of reweight points. Blank entries indicate the process is unaffected by the WC.

Operator	definition	Wilson coefficient	comments
$\dagger O_{u\phi}^{(ij)}$	$\bar{q}_i u_j \tilde{\phi} (\phi^\dagger \phi)$	ctp + i ctpI	Non-Hermitian, ctpI is set to 0
$O_{\phi q}^{1(ij)}$	$(\phi^\dagger i \overleftrightarrow{D}_\mu \phi) (\bar{q}_i \gamma^\mu q_j)$	cpQM + cpQ3	-
$O_{\phi q}^{3(ij)}$	$(\phi^\dagger i \overleftrightarrow{D}_\mu^L \phi) (\bar{q}_i \gamma^\mu \tau^L q_j)$	cpQ3	-
$O_{\phi u}^{(ij)}$	$(\phi^\dagger i \overleftrightarrow{D}_\mu \phi) (\bar{u}_i \gamma^\mu u_j)$	cpt	-
$\dagger O_{\phi ud}^{(ij)}$	$(\phi^\dagger i D_\mu \phi) (\bar{u}_i \gamma^\mu d_j)$	cptb + i cptbI	Non-Hermitian, cptbI is set to 0
$\dagger O_{uW}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} \tau^L u_j) \tilde{\phi} W_{\mu\nu}^I$	ctW + i ctWI	Non-Hermitian, ctWI is set to 0
$\dagger O_{dW}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} \tau^L d_j) \phi W_{\mu\nu}^I$	cbW + i cbWI	Non-Hermitian, cbWI is set to 0
$\dagger O_{uB}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} u_j) \tilde{\phi} B_{\mu\nu}$	(cw*ctW - ctZ)/sw + i (cw*ctWI - ctZI)/sw	sw = $\sqrt{1 - (M_W/M_Z)^2}$, cw ² =1-sw ² , Non-Hermitian, ctWI and ctZI are set to 0
$\dagger O_{uG}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} T^A u_j) \tilde{\phi} G_{\mu\nu}^A$	ctG + i ctGI	Non-Hermitian, ctGI is set to 0
$O_{lq}^{1(ijkl)}$	$(\bar{l}_i \gamma^\mu l_j) (\bar{q}_k \gamma^\mu q_l)$	cQIM + cQI3	-
$O_{lq}^{3(ijkl)}$	$(\bar{l}_i \gamma^\mu \tau^L l_j) (\bar{q}_k \gamma^\mu \tau^L q_l)$	cQI3	-
$O_{lu}^{(ijkl)}$	$(\bar{l}_i \gamma^\mu l_j) (\bar{u}_k \gamma^\mu u_l)$	ctl	-
$O_{eq}^{(ijkl)}$	$(\bar{e}_i \gamma^\mu e_j) (\bar{q}_k \gamma^\mu q_l)$	cQe	-
$O_{eu}^{(ijkl)}$	$(\bar{e}_i \gamma^\mu e_j) (\bar{u}_k \gamma^\mu u_l)$	cte	-
$\dagger O_{lequ}^{1(ijkl)}$	$(\bar{l}_i e_j) \varepsilon (\bar{q}_k u_l)$	ctlS + i ctlSI	Non-Hermitian, ctlSI is set to 0
$\dagger O_{lequ}^{3(ijkl)}$	$(\bar{l}_i \sigma^{\mu\nu} e_j) \varepsilon (\bar{q}_k \sigma_{\mu\nu} u_l)$	ctlT + i ctlTI	Non-Hermitian, ctlTI is set to 0

Table 2: Operators that have effects on ttH, ttW, ttZ, and tqZ processes at order $1/\Lambda^2$. The field ϕ ($\tilde{\phi} = \epsilon\phi^*$) is the Higgs boson doublet. $D_\mu = \partial_\mu - ig_s \frac{1}{2} \lambda^A G_\mu^A - ig \frac{1}{2} \tau^I W_\mu^I - ig' Y B_\mu$ is the covariant derivative. $W_{\mu\nu}^I = \partial_\mu W_\nu^I - \partial_\nu W_\mu^I + g \epsilon_{IJK} W_\mu^J W_\nu^K$ is the W boson field strength, and $G_{\mu\nu}^A = \partial_\mu G_\nu^A - \partial_\nu G_\mu^A + g_s f^{ABC} G_\mu^B G_\nu^C$ is the gluon field strength. More details about the operators can be found in Ref [4].

dataset as the one with the larger integrated luminosity. Even though the 2018 dataset is now available to be analyzed, it is not possible for us to use it on the time scale of this analysis for two reasons. First, we would have to regenerate the MC. Secondly, we rely on the ttH multilepton group for the lepton MVA, including scale factors and fake rates, and those are not available for the 2018 dataset at this time.

One a more strategic level, we argue that it makes good sense to release this first result on only part of the full Run 2 dataset. This analysis represents a new approach to EFT analysis. We anticipate that this first result will generate important discussion in the HEP community and result in useful feedback for honing future iterations using this approach. In anticipation of this valuable feedback, we not only want to release a result as promptly as possible, not delaying to deal with the technical complications of combining 2016, 2017, and 2018 mentioned above, but we also think it's useful to have some of the Run 2 data held out of the analysis, so that as we adjust the technique based on feedback received, we can remain unbiased to most of the Run 2 data.

Because this is a new analysis approach, we think it is important to publish these results in a timely fashion so that we can collect feedback from the community, including theorists. Therefore, we push to publish the results with the 2017 data. Once that publication is in hand, we intend to pursue a second iteration of the analysis which will include more differential information, and possibly also target additional signatures. This will also give us to incorporate feedback received from the HEP community on the first version of the analysis. During the

Single lepton triggers	HLT_Ele32_WPTight_Gsf HLT_Ele35_WPTight_Gsf HLT_IsoMu24 HLT_IsoMu27
Double lepton triggers	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
Triple lepton triggers	HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL HLT_Mu8_DiEle12_CaloIdL_TrackIdL HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ HLT_TripleMu_12_10_5

Table 3: Triggers used to record the data for this analysis.

time it takes to develop the more advanced version of the analysis, we will also have time to generate the necessary signal samples for the 2016 and 2018 data-taking periods, which will allow us to publish a result on the full Run 2 dataset with not only more data, but improved techniques. This will certainly take some time to accomplish, however, so it is essential that we not delay this publication of this first result.

3.1 Data Samples and Triggers

The data used in this analysis has been collected with the CMS detector in 2017. We use data collected only during periods when the CMS magnet was on, corresponding to a total integrated luminosity of 41.29 fb^{-1} [5]. We use the 31Mar2018 version of the datasets. Details on the various datasets contributing to our analysis are given in table 4.

The HLT paths used to collect the data to be analyzed are listed in Table 3. A combination of single and double lepton triggers are used to record events in the $2\ell ss$ channels, while events in the 3ℓ and 4ℓ channels are recorded using a combination of single, double, and triple lepton triggers. The motivation for choosing a mix of triggers is to use the trigger of higher efficiency whenever it is available and to resort to triggers of lower efficiency whenever the trigger of higher efficiency is disabled or prescaled. Events recorded in any data-taking period as well as simulated events are selected in case they pass any of the HLT paths that are included in the mix. The effect of triggers that are disabled or prescaled in some part of the analyzed data is accounted for by applying suitably chosen weights to simulated events. These weights are detailed in Section 6.2.

3.2 Signal Samples

Samples of signal and of background events, produced by Monte Carlo (MC) simulation, are used to estimate event yields. The signal samples chosen for this analysis are those that include one or more top quarks, along with multiple leptons in the final state. These processes include $t\bar{t}H$, $t\bar{t}l\bar{l}$, $t\bar{t}l\nu$, $t\bar{t}lq$ and $t\bar{t}Hq$. An additional set of $t\bar{t}H$, $t\bar{t}Z$, $t\bar{t}Zq$, and $t\bar{t}Hq$ are simulated with aMC@NLO [7] and matched to PYTHIA [8] for the parton shower. These samples are only used for validation of the privately produced samples. Details on these centrally produced samples are in Table 5.

Dataset	Run Range	Int. Lumi (fb ⁻¹)
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.77 fb ⁻¹
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.58 fb ⁻¹
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.22 fb ⁻¹
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.26 fb ⁻¹
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.46 fb ⁻¹
Total DoubleMuon	297046–306462	41.37 fb⁻¹
/DoubleEG/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.77 fb ⁻¹
/DoubleEG/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.58 fb ⁻¹
/DoubleEG/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.22 fb ⁻¹
/DoubleEG/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.26 fb ⁻¹
/DoubleEG/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.46 fb ⁻¹
Total DoubleEG	297046–306462	41.37 fb⁻¹
/MuonEG/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.77 fb ⁻¹
/MuonEG/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.58 fb ⁻¹
/MuonEG/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.22 fb ⁻¹
/MuonEG/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.26 fb ⁻¹
/MuonEG/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.46 fb ⁻¹
Total MuonEG	297046–306462	41.37 fb⁻¹
/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.77 fb ⁻¹
/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.58 fb ⁻¹
/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.22 fb ⁻¹
/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.26 fb ⁻¹
/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.46 fb ⁻¹
Total SingleMuon	297046–306462	41.37 fb⁻¹
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.77 fb ⁻¹
/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.58 fb ⁻¹
/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.22 fb ⁻¹
/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.26 fb ⁻¹
/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.46 fb ⁻¹
Total SingleElectron	297046–306462	41.37 fb⁻¹

Table 4: The datasets used in this analysis, with corresponding run ranges and integrated luminosities. Luminosities listed here are derived using the brilcalc tool [6].

Sample	Xsec (pb)
/tHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.5085 ^(1-0.577)
/TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.2043
/TTZToLLNuNu_M-10_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.2529
/tZqJL4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.0758
/THQ_4f_Hincl_13TeV_madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.0771

Table 5: Central samples used only for validation.

For the signal processes, on top of the centrally produced samples, a set of privately produced MC is used. Each of these signal samples are produced using leading-order (LO) matrix elements implemented in the MADGRAPH5_aMC@NLO 2.6.0 program [7] with the dim6TopEFT physics model [4]. Each sample was generated setting the Wilson Coefficient (WC) parameter strengths to non-zero values. The effects of the EFT operators are not included in the decays of the matrix element final state particles (e.g. top and higgs decays), which are handled by PYTHIA. Details on the private samples are in Table 6.

The 4top sample is excluded from the analysis because the operators considered are expected to have a negligible impact on the 4top process relative to the other signal processes. The

expected impact on the 4top process is based on the results from the LHC Top WG EFT note [4]. Additionally, the higher jet multiplicity signal regions are not split up, thereby significantly reducing the sensitivity to the 4top process. Splitting up these jet bins requires producing the other signal MC with extra jets, which currently is not done.

In order to estimate the signal yields over a large range of Wilson coefficient strengths, MADGRAPH event re-weighting is used. A set of hard events are generated and for each event a set of rescaling factors are calculated and stored along with the normal event information. Since each individual event represents an infinitesimal section of the inclusive cross section, the rescaling factors also a quadratic parameterization based on the Wilson coefficient strengths. The re-weight points are then used to generate a n-D quadratic fit for each event in the sample that can be used to estimate all possible choices of WC values. The number of re-weight points needed to fully fit a generic n-D quadratic function is given by:

$$N = 1 + 2d + \frac{d}{2}(d - 1) \quad (6)$$

Where d is the number of independent WCs considered and N is the number of non-degenerate re-weight points. For this analysis there are 16 WCs, which requires a minimum of 153 re-weight points. In practice, due to the finite precision of the MADGRAPH calculation, we use an overdetermined least-squares fit with 184 points. The additional points are also used to validate the accuracy of the quadratic parameterization.

Multiple events can have their individual fits summed together into a single fit, that can rescale the entire set of events. This is implemented by using a custom `TH1EFT` histogram class, that stores and keeps track of each event fit on a bin by bin basis. To ensure the initial set of simulated events populate the full EFT phase space, each signal sample is generated with each Wilson coefficient set to a non-SM value.

Sample	Xsec (pb)
/store/user/awightma/FullProduction/Round4/Batch*/postLHE_step/v1/mAOD_step.tth.*	0.3858
/store/user/awightma/FullProduction/Round4/Batch*/postLHE_step/v1/mAOD_step.tll.*	0.0747
/store/user/awightma/FullProduction/Round4/Batch*/postLHE_step/v1/mAOD_step.tlnu.*	0.1121
/store/user/awightma/FullProduction/Round4/Batch*/postLHE_step/v1/mAOD_step.tllq.*	0.0677
/store/user/awightma/FullProduction/Round4/Batch*/postLHE_step/v1/mAOD_step.tHq.*	0.0730

Table 6: Privately produced signal samples used in this analysis, located on the Notre Dame T3.

The private samples give results equivalent to the SM when all WC are set to zero, via re-weighting. Validation is done by re-weighting the private samples to SM expectation and comparing to the relevant centrally produced SM samples.

The private MADGRAPH gridpacks are produced using the same `cmssw` `genproductions` framework that is used to produce the central samples, working on the `mg26x` branch. The samples produced by MADGRAPH are generated with the NNPDF3.1 [9] set of parton distribution functions. Parton shower and hadronization processes are modelled using the generator PYTHIA [8] with the CP5 tune [10].

Minimum bias events are overlaid on all simulated events, according to the luminosity profile of the analyzed data and for a pp inelastic cross section of 69.2 mb.

All generated events are passed through a detailed simulation of the CMS apparatus, based on GEANT4 [11], and are reconstructed using the same version of the CMS event reconstruction software as used for data.

Sample	Xsec (pb)
/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	4.42965
/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	10.481
/ZZTo4L_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	1.256
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.1651
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.05565
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	0.01398
/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	831.76
/ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	136.02
/ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	80.95
/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	3.68
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	35.6
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM	35.6
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	18610
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-ext1-v1/MINIAODSIM	6025.2
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-ext1-v2/MINIAODSIM	61526.7

Table 7: List of background samples used in this analysis (CMSSW 94X). In the first section of the table are listed the samples of the processes for which we use the simulation to extract the final yields and shapes, in the second section the samples of the processes we will estimate from data.

Each signal sample was validated by comparing gen-level and basic reconstructed objects to the corresponding centrally produced sample, details in Appendix B.

3.3 Background Samples

The simulated samples used to estimate the backgrounds include $t\bar{t}\gamma$ and $t\bar{t}\gamma^*$, diboson and tri-boson production, and single top (tW). Additional samples are also generated to simulate SM background processes that are estimated from data, and are used for validation purposes and is covered in more detail in Section 7. These samples include Z+jets, W+jets, $t\bar{t}$ +jets, and some single top processes (s,t-channel). The samples are generated using next-to-leading order (NLO) matrix elements implemented in the MADGRAPH5_aMC@NLO [7, 12, 13] and POWHEG [14–19] programs. No dim6 EFT effects are included in any of the background samples, that is, the EFT operators are all set to their SM expectation value of zero. The same PYTHIA chain that was outlined in section 3.2 is also used for every background sample. See table 7 for the full list of background samples used in this analysis.

4 Object reconstruction and identification

The information provided by all CMS subdetectors is employed by a particle-flow (PF) algorithm [20–24] to identify and reconstruct individual particles in the event, namely muons, electrons, photons, charged and neutral hadrons. These particles are then used to reconstruct jets and the missing transverse momentum vector, as well as to quantify the isolation of leptons.

Some of the criteria used for particle identification, in particular the isolation of electrons and muons, depend on the choice of a primary collision vertex (PV). We choose this vertex to be the one that has the highest $\sum p_T^2$ of associated tracks.

4.1 Lepton Identification

The identification of electrons and muons is performed in two stages: Basic electron (muon) identification criteria, developed by the EGamma (Muon) POG, and loose isolation criteria are applied in the first stage to separate genuine leptons from jet backgrounds. In the second stage, the leptons originating from decays of W and Z bosons and from τ lepton decays are separated from leptons produced in the decays of charm (c) and bottom (b) quarks. We will refer to the former as “prompt” or signal leptons and to the latter as “non-prompt” or background leptons. In order to maximize our acceptance of prompt leptons, while simultaneously maximizing our rejection of background leptons, we exploit a multivariate technique developed by another analysis, discussed in Section 4.1.4.

4.1.1 Basic electron identification

The identification of electrons is performed by a multivariate algorithm [25, 26], based on a BDT [27], which has been trained to separate electrons from jets. The training, performed by the EGamma POG, is done in two bins of p_T and three bins in η , using simulated samples of electrons and jets. Separate discriminants have been trained on electrons passing the requirements of the single electron trigger and for an inclusive sample of electron candidates. We use the discriminants trained on the inclusive electron selection. Loose and tight working-points (WP) are defined by corresponding cuts on the BDT output. The cuts, applied in different p_T and η bins, are given in Ref. [26].

In order to remove electron candidates that are due to photon conversions we require that the electron track is associated to a hit in each layer of the pixel detector that is crossed by the track, except for at most one layer. We further reject electron candidates in case there exist a track of opposite charge near the electron track that, if paired with the electron track, can be fitted to a common vertex within the volume of the tracking detector.

4.1.2 Basic muon identification

The identification of muons is based on linking track segments reconstructed in the silicon tracking detector and in the muon system [28]. The matching between track segments is done outside-in, starting from a track in the muon system, and inside-out, starting from a track reconstructed in the inner detector. In case a link can be established, the track parameters are refitted using the combination of hits in the inner and outer detectors and the track is referred to as global muon track. Quality cuts are applied on the multiplicity of hits, on the number of matched segments and on the quality of the global muon track fit, quantified by χ^2 . The muon candidates used in the analysis are required to pass the medium PF muon identification criteria [29].

4.1.3 Electron and muon isolation

Electrons and muons in signal events are expected to be isolated, while leptons from c and b quark decays, as well as from in-flight decays of pions and kaons, are often reconstructed within jets. Isolated leptons are distinguished from leptons in jets by means of the sum of scalar p_T values of charged particles, neutral hadrons, and photons, that are reconstructed within a narrow cone centered on the lepton direction. The size R of the cone shrinks inversely proportional with the p_T of the lepton in order to increase the efficiency for leptons reconstructed in events with “boosted” topologies and/or high hadronic activity to pass the isolation criteria. The narrow cone size, referred to as “mini isolation”, has the further advantage that it reduces the effect of pileup (PU). Efficiency loss due to PU is further reduced by considering only charged particles originating from the lepton production vertex in the isolation sum. Residual contributions of PU to the neutral component of the isolation of the lepton is taken into account by means of so-called effective area corrections:

$$I_\ell = \sum_{\text{charged}} p_T + \max \left(0, \sum_{\text{neutrals}} p_T - \rho \mathcal{A} \left(\frac{R}{0.3} \right)^2 \right), \quad (7)$$

where ρ represents the energy density of neutral particles reconstructed within the geometric acceptance of the tracking detectors (`fixedGridRhoFastJetAll`), computed as described in Refs. [30, 31]. The size of the cone is given by:

$$R = \begin{cases} 0.05 & \text{if } p_T > 200 \text{ GeV} \\ 10 \text{ GeV}/p_T & \text{if } 50 < p_T < 200 \text{ GeV} \\ 0.20 & \text{if } p_T < 50 \text{ GeV} \end{cases}. \quad (8)$$

The effective area \mathcal{A} is obtained from the simulation, by studying the correlation between I_ℓ and ρ , and is determined in bins of η , separately for electrons and muons. Numerical values are given in Table 8.

Electrons	
Pseudorapidity range	\mathcal{A}
$0.0 < \eta < 1.0$	0.1566
$1.0 < \eta < 1.479$	0.1626
$1.479 < \eta < 2.0$	0.1073
$2.0 < \eta < 2.2$	0.0854
$2.2 < \eta < 2.3$	0.1051
$2.3 < \eta < 2.4$	0.1204
$2.4 < \eta < 2.5$	0.1524

Muons	
Pseudorapidity range	\mathcal{A}
$0.0 < \eta < 0.8$	0.0566
$0.8 < \eta < 1.3$	0.0562
$1.3 < \eta < 2.0$	0.0363
$2.0 < \eta < 2.2$	0.0119
$2.2 < \eta < 2.5$	0.0064

Table 8: Effective areas \mathcal{A} for electrons (top) and muons (bottom).

4.1.4 Separation of prompt from non-prompt electrons and muons

The final separation of prompt leptons from non-prompt and fake leptons is performed by a BDT-based algorithm [32, 33]. The algorithm has been retrained on MC samples with 2017 detector conditions, separately for electrons and muons. For convenience (and for reasons described in Section 7.1 and Appendix A), we use the exact same BDT training as is used by the 2017 ttH multilepton analysis. The method is described briefly below; for more information please see here [34, 35].

The following observables are inputs to the BDT:

- `pt` and `eta`, the p_T and η of the lepton.
- `miniRelIsoCharged`, the isolation of the lepton with respect to charged particles, $I_\ell^{\text{charged}} = \sum_{\text{charged}} p_T$.
- `miniRelIsoNeutral`, the isolation of the lepton with respect to neutral particles, corrected for PU effects, $I_\ell^{\text{neutrals}} = \max\left(0, \sum_{\text{neutrals}} p_T - \rho \mathcal{A} \left(\frac{R}{0.3}\right)^2\right)$.
- `jetPtRatio`, the ratio of the transverse momentum of the lepton to the transverse momentum of the nearest jet, p_T^ℓ / p_T^j .
- `jetBTagCSV`, the discriminant value of the CSVv2 b-tagging algorithm of the jet (see Section 4.2).
- `jetNDauChargedMVASel`: the number N_{charged} of charged particles within the jet.
- `jetPtRelv2`, the component of the lepton momentum in direction transverse to the jet, $p_T^{\text{rel}} = p_\ell \sin \theta$. where θ denotes the angle between the lepton and jet momentum vectors.
- `dxy` and `dz`, the transverse and longitudinal impact parameters of the lepton track with respect to the PV.
- `sip3d`, the signed impact parameter, in three dimensions, of the lepton track with respect to the PV, divided by its uncertainty, which corresponds to its significance d/σ_d .
- `mvaIdFall17noIso`, the output of the BDT that separates electrons from jets, trained by the EGamma POG.
- `segmentCompatibility`, the compatibility of track segments in the muon system with the pattern expected for a minimum ionizing particle.

The observable `mvaIdFall17noIso` (`segmentCompatibility`) is used only for electrons (muons). A lower jet p_T threshold of 15 GeV is used for the matching of jets to leptons. In case no jet of $p_T > 15$ GeV is within a distance $\Delta R < 0.4$ to the lepton, the value of the observable `jetPtRatio` is set to $p_T^\ell / (p_T^\ell + I_\ell)$ and the value of the observables `jetBTagCSV`, `jetNDauChargedMVASel`, and `jetPtRelv2` is set to zero. The p_T of the jet is computed by applying the “lepton-aware” jet energy corrections (JEC) [36], which apply the JEC to the difference of uncalibrated jet minus lepton and subsequently add the lepton momentum to the calibrated jet. The inputs are complemented by the p_T and η of the lepton. We refer to the output of the BDT trained on electrons (muons) as prompt-e (prompt- μ) MVA.

The BDTs are trained on simulated samples of prompt leptons in ttH events (signal) and non-prompt leptons in tt+jets events (background). The leptons used for the training are required to pass loose preselection criteria (referred to as “loose” lepton selection), given by Tables 9 and 10. Further “tight” and “fakeable” lepton selection criteria are used for the purpose of selecting events in the signal region and for estimating the fake lepton background from control

regions in data, respectively. In order to reduce potential biases of the background estimation procedure, the p_T of leptons that pass the fakeable, but fail the tight lepton selection criteria is set to 0.90 times the p_T of the nearest jet in case the distance between lepton and nearest jet satisfies $\Delta R < 0.4$. In case $\Delta R > 0.4$ the p_T of fakeable leptons is set to $p_T^\ell / (p_T^\ell + I_\ell)$. We refer to the p_T of fakeable leptons, computed in this way, as “cone- p_T ”. The cone- p_T in general exceeds the transverse momentum of the lepton that is determined by the electron and muon reconstruction algorithms. We refer to the latter as “reco- p_T ” to distinguish it from the cone- p_T when there is danger of confusion. When there is no danger of confusion, we use the terms reco- p_T and p_T synonymously. Electrons passing the fakeable and tight lepton selection criteria are required to satisfy a set of conditions on the width of the electron cluster in η -direction ($\sigma_{i\eta i\eta}$), the ratio of energy in the HCAL to the energy in the ECAL that is associated to electron (H/E), and the difference between the reciprocal of the electron cluster energy and the reciprocal of its track momentum ($1/E - 1/p$). They mimic the electron identification criteria applied at trigger level. Finally, any electron overlapping with a loose muon such that $\Delta R = \sqrt{(\eta^e - \eta^\mu)^2 + (\phi^e - \phi^\mu)^2} < 0.05$ is removed from consideration.

Observable	Loose	Fakeable	Tight
p_T	$> 7 \text{ GeV}$	$> 10 \text{ GeV}$	$> 10 \text{ GeV}$
$ \eta $	< 2.5	< 2.5	< 2.5
$ d_{xy} $	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$ d_z $	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
EGamma POG MVA	$> \{-0.86 / -0.81 / -0.72\}^1$	$> \{-0.86 / -0.81 / -0.72\}^2$	$> \{-0.86 / -0.81 / -0.72\}$
$\sigma_{i\eta i\eta}$	—	$< \{0.011 / 0.011 / 0.030\}$	$< \{0.011 / 0.011 / 0.030\}$
H/E	—	< 0.10	< 0.10
$1/E - 1/p$	—	> -0.04	> -0.04
Conversion rejection	—	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
p_T^e / p_T^j	—	$> 0.6 \dagger (-)$	—
DeepCSV(b) of nearby jet	—	$< 0.07 \dagger (< 0.4941)$	< 0.4941
Prompt-e MVA	—	$< 0.90 \dagger (> 0.90)$	> 0.90

¹ $> \{-0.13 / -0.32 / -0.08\}$ if $p_T < 10 \text{ GeV}$

² > 0.50 if prompt-e MVA < 0.90

Table 9: Loose, fakeable, and tight selection criteria for electrons. The requirement on the output of the BDT trained by the EGamma POG (see Section 4.1.1) and on the observables $\sigma_{i\eta i\eta}$, H/E, and $1/E - 1/p$ are varied as function of η of the electron candidate. The numbers are separated by slashes and refer to $|\eta| < 0.8$, $0.8 < |\eta| < 1.479$, and $|\eta| > 1.479$, respectively. The conditions on p_T^e / p_T^j and on the CSV discriminant of the jet nearest to the electron are tightened (relaxed) for fakeable electrons that fail (pass) the requirement prompt-e MVA > 0.90 , in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavour composition (see Section 7.4.1 in Ref. [37]). The tightened (relaxed) conditions are indicated by the symbol \dagger (put in parentheses). A hyphen (—) indicates selection criteria that are not applied.

4.2 Jets and B-tagging

Jets are reconstructed by clustering PF candidates using the anti- k_T algorithm with distance parameter $\Delta R = 0.4$ as implemented in the FASTJET package [38, 39]. The charged hadrons

Observable	Loose	Fakeable	Tight
p_T	$> 5 \text{ GeV}$	$> 10 \text{ GeV}$	$> 10 \text{ GeV}$
$ \eta $	< 2.4	< 2.4	< 2.4
$ d_{xy} $	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$ d_z $	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
Loose PF muon	✓	✓	✓
Medium PF muon	—	—	✓
Segment compatibility	—	$> 0.3 \dagger (-)$	—
p_T^μ/p_T^j	—	$> 0.6 \dagger (-)$	—
DeepCSV(b) of nearby jet	—	$< 0.07 \dagger (< 0.4941)$	< 0.4941
Prompt- μ MVA	—	$< 0.90 \dagger (> 0.90)$	> 0.90

Table 10: Loose, fakeable, and tight selection criteria for muons. The conditions on p_T^μ/p_T^j and on the CSV discriminant of the jet nearest to the muon are tightened (relaxed) for fakeable muons that fail (pass) the requirement prompt- μ MVA > 0.90 , in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavour composition (see Section 7.4.1 in Ref. [37]). The tightened (relaxed) conditions are indicated by the symbol \dagger (put in brackets). A hyphen (—) indicates selection criteria that are not applied.

not coming from the primary vertices are subtracted from the PF candidates considered in the clustering. Fake jets, mainly arising from calorimeter noise, are rejected by requiring reconstructed jets to pass a set of tight jet identification criteria [40]. The prescribed jet energy corrections in Table 11 are applied as a function of the jet E_T and η [41]. Within the tracker acceptance the jet tracks are also required to be compatible with the primary vertex. Jets are only considered if they have a transverse energy above 30 GeV and $|\eta| < 2.4$. In addition, they have to be separated from any lepton candidates passing the Loose selection, by requiring $\Delta R = \sqrt{(\eta^\ell - \eta^{jet})^2 + (\phi^\ell - \phi^{jet})^2} > 0.4$.

Table 11: Global tags used to apply jet energy corrections to data and simulated events.

Event type	global tag
Simulation	94X_mc2017_realistic_v13
Data	94X_dataRun2_v6

The DeepCSV b-tagging algorithm [42] is used to identify jets that are likely to originate from the hadronization of bottom quarks. This algorithm uses a deep neural network. The input is the same set of observables used by the existing CSVv2 b-tagger, with the extension that it uses information of more tracks. The discriminant output value ranges from zero to one. It distinguishes between b-jets and jets originating from light quarks, gluons and charm quarks. The efficiency to tag b-jets and the rate of misidentification of non-b jets depend on the operating point chosen. Both the efficiency and the fake rate are parameterised as a function of the transverse momentum and pseudorapidity of the jets. These performance measurements are obtained directly from data in samples that can be enriched in b jets, such as $t\bar{t}$ and multijet events where a muon can be found inside the one of jets. Two working points for the DeepCSV(b) output discriminant are used in the analysis. The *loose* working point (DeepCSV(b) > 0.1522) has approximately 85% efficiency to tag jets with b quarks and a 10% chance to tag jets with only light quarks or gluons. The *medium* working point (DeepCSV(b) > 0.4941) has approximately

425 70% efficiency for tagging jets with b quarks and 1.0% chance to tag jets with only light quarks
426 or gluons [42].
427

DRAFT

5 Event selection and categorization

The goal of the event selection is to select $t\bar{t}H$, $t\bar{t}l\nu$, $t\bar{t}l\bar{l}$ and $t\bar{t}lq$ events while excluding as many contributions from background processes as possible. To accomplish this, we target events containing multiple leptons and jets, as well as b-tagged jets, with additional requirements that depend on the lepton flavor and multiplicity.

We begin by requiring that at least two leptons passing the tight selection are present in the event. Events where a pair of loose leptons with an invariant mass smaller than 12 GeV is found are rejected, as they are not modeled by the simulation. For all events passing the selection, we require at least two jets with transverse momentum greater than 30 GeV be reconstructed in the $|\eta| < 2.4$ region. We also require the presence of one or more jets that pass the medium working point of the DeepCSV b-tag algorithm, as a top quark pair decaying into b-jets is present in all signal events. We place no explicit identification requirements on τ leptons, but allow them to enter our event selection via their decay products.

5.1 $2\ell ss$ category

The $2\ell ss$ category primarily targets $t\bar{t}H$ and $t\bar{t}l\nu$ signal events in which the $t\bar{t}$ system decays semi-leptonically, with an additional, identically-charged lepton decaying from the W boson produced in association with the top quark pair (in the case of $t\bar{t}l\nu$), or decaying from a W or τ from the decay of the H boson (in the case of $t\bar{t}H$). In addition, $t\bar{t}l\bar{l}$ events may contribute events in which there is at least one lepton that is not reconstructed or is out of acceptance. Selected events are required to contain two leptons of the same charge and passing the tight object selection criteria. The lepton of higher (lower) p_T is required to have $p_T > 25$ GeV (> 15 GeV). Requiring both leptons to be of the same charge removes almost all of the large $t\bar{t}$ +jets background. Residual $t\bar{t}$ +jets background contributions are further suppressed by requiring that the charge of all electrons and muons in the event, which pass the fakeable object selection criteria, is well measured. Electrons are required to pass the two conditions `isGsfCtfScPixChargeConsistent` and `isGsfScPixChargeConsistent`, which test the consistency between the independent measurements of the electron charge obtained from the position of the ECAL cluster and from its track, while muons must satisfy the condition that the estimated uncertainty on the p_T of the muon track is below 0.2 times its p_T . The latter is equivalent to requiring that the sign of the curvature of the muon track, which determines its charge, is measured with a significance of 5σ . After applying these conditions, the charge misidentification rate is on the per-mille level for electrons and negligible for muons [43]. The events are required to contain at least four jets of $p_T > 30$ GeV and $|\eta| < 2.4$, at least one(two) of which must satisfy the medium(loose) working point of the DeepCSV b-tag algorithm. Events containing more than two tight leptons are vetoed to avoid overlap with the 3ℓ and 4ℓ categories.

5.2 3ℓ category

The 3ℓ category selects $t\bar{t}l\nu$ events in which all three W bosons decay leptonically, $t\bar{t}l\bar{l}$ events in which the $t\bar{t}$ system decays semi-leptonically and the Z decays to two leptons, $t\bar{t}lq$ events in which the top quark decays leptonically and the Z decays to two leptons, and $t\bar{t}H$ events in which the H boson decays to W or τ particles, at least one of which then decays leptonically (with one or more leptons also coming from the decay of the $t\bar{t}$ system). Selected events are required to contain exactly three leptons passing the tight object selection criteria. The lepton of highest, second, and third highest p_T is required to have $p_T > 25$ GeV, > 15 GeV, and > 10 GeV, respectively. In the case that the trailing lepton is an electron, the requirement on it

is instead $p_T > 15$ GeV. Two or more jets of $p_T > 30$ GeV and $|\eta| < 2.4$ are required, at least one of which must satisfy the medium working point of the DeepCSV b-tag algorithm.

5.3 4ℓ category

The 4ℓ category targets $t\bar{t}l\bar{l}$ events in which all the W and Z bosons decay leptonically, and $t\bar{t}H$ events in which the H boson decays into a pair of W bosons or τ leptons, and all W bosons and τ leptons in the event decay leptonically. Events selected in this category are required to contain four or more leptons passing the tight object selection criteria and passing p_T thresholds of $p_T > 25$ GeV, > 15 GeV, > 10 GeV, and > 10 GeV, for the lepton of highest, second, third, and fourth highest p_T , respectively. In the case of electrons, the requirement on the third and fourth leptons is instead $p_T > 15$ GeV. Two or more jets of $p_T > 30$ GeV and $|\eta| < 2.4$ are required. At least one(two) jets must satisfy the medium(loose) working point of the DeepCSV b-tag algorithm.

5.4 Separation into sub-categories

Events in the $2\ell ss$, 3ℓ and 4ℓ categories are further separated based on a number of criteria. In the $2\ell ss$ category, events are distinguished based on flavor and charge multiplicity. The latter allows us to take advantage of the fact that the $t\bar{t}W^+$ cross section is roughly a factor of two larger than that of the $t\bar{t}W^-$ cross section, so it is enhanced in $2\ell ss$ events where both leptons are positively charged, and the former increases sensitivity to those variations in Wilson coefficients which produce flavor-dependent effects on event yields. In the 3ℓ category, we consider separately events which contain a same-flavor, oppositely-charged pair of leptons that fall within 10 GeV of the mass of the Z boson, which primarily serves to create a region where the contribution from (on-shell) $t\bar{t}Z$ is enhanced. For 3ℓ events that do not fall within this region, we further separate based on the sum of the charge multiplicity of the leptons, considering events where the sum of the charge is positive separately from those where the sum of the charge is negative (this again exploits the difference in cross section between $t\bar{t}W^+$ and $t\bar{t}W^-$). In all 3ℓ events, we further separate based on whether the event contains exactly one jet passing the medium DeepCSV(b) working point, or two or more jets passing the medium DeepCSV(b) working point. The contribution from $t\bar{t}lq$ is enhanced in the former case.

Events in all categories are then separated into jet-multiplicity bins, which are then used to fit to data and place limits on EFT parameters, as described in section 8. Table 12 summarizes the expected number of events in each category, and figures 1 and 2 show the expected number of events per category as a function of jet multiplicity.

	$2\ell ss (e^+e^+)$	$2\ell ss (e^+\mu^+)$	$2\ell ss (\mu^+\mu^+)$	$2\ell ss (e^-e^-)$	$2\ell ss (e^-\mu^-)$	$2\ell ss (\mu^-\mu^-)$	$3\ell (1b "p")$	$3\ell (1b "m")$	$3\ell (\geq 2b "p")$	$3\ell (\geq 2b "m")$	$3\ell (SFOS Z, 1b)$	$3\ell (SFOS Z, \geq 2b)$	$\geq 4\ell$
Diboson	0.30	0.72	0.50	0.46	0.45	0.37	5.93	3.49	0.35	0.16	52.34	4.16	0.64
Triboson	0.01	0.30	0.18	0.15	0.29	0.09	0.17	0.23	0.05	0.06	3.36	0.58	0.09
Charge Flips	4.23	4.26	0.00	4.33	4.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fakes	5.27	12.54	7.75	4.71	13.03	9.08	11.28	12.99	3.27	2.46	16.90	3.83	0.00
Convs	3.95	5.71	1.21	2.86	4.54	1.91	2.35	2.55	1.73	1.95	0.85	0.36	0.00
Sum Background	13.76	23.54	9.64	12.52	22.44	11.44	19.73	19.25	5.39	4.63	73.44	8.92	0.73
$t\bar{t}l\bar{l}$	13.42	35.42	23.56	7.28	19.12	12.73	16.51	9.20	13.21	7.33	3.64	2.88	0.01
$t\bar{t}l$	3.80	8.48	4.10	3.67	8.31	4.51	10.54	9.88	7.02	7.33	85.17	55.16	8.87
$t\bar{t}H$	4.72	11.85	7.89	4.63	12.18	7.64	8.24	8.15	5.57	5.67	3.46	2.36	1.12
$t\bar{t}lq$	0.41	0.68	0.34	0.15	0.36	0.21	3.69	2.18	1.03	0.49	38.74	9.99	0.01
$t\bar{t}Hq$	0.09	0.43	0.11	0.11	0.08	0.04	0.32	0.05	0.11	0.03	0.07	0.01	0.01

Table 12: SM expected yields after the selection in all categories and final states.

6 Data-to-MC corrections

In order to improve the modeling of the data, we apply corrections to simulated events in terms of:

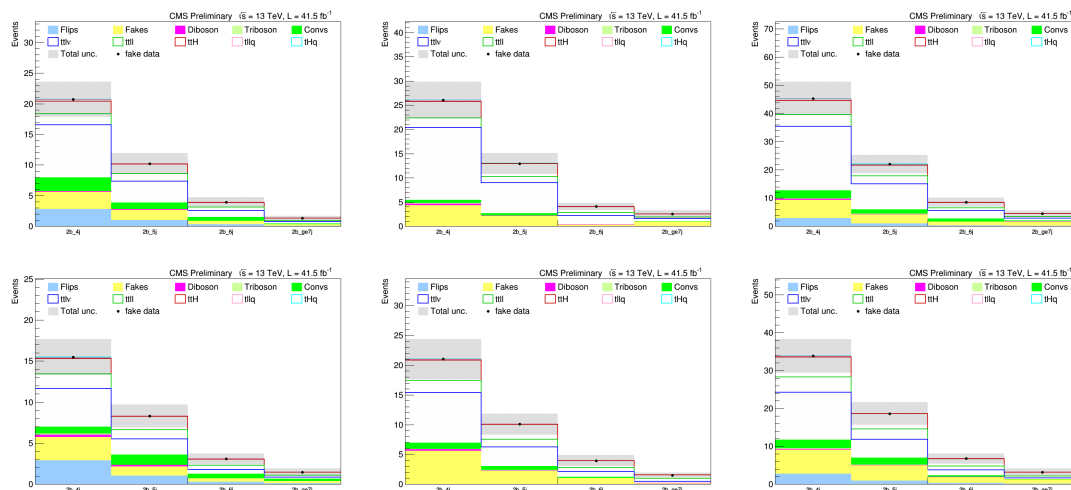


Figure 1: SM Expected number of signal and background events in the 2lss categories. From left to right in the top row, the plots show events in the 2lss (e^+e^+), 2lss ($\mu^+\mu^+$) and 2lss ($e^+\mu^+$) categories. From left to right on the bottom row, the plots show the 2lss (e^-e^-), 2lss ($\mu^-\mu^-$) and 2lss ($e^-\mu^-$) categories.

- Pileup reweighting
- Trigger efficiency
- e and μ identification and isolation efficiency
- b-tag efficiency and mistag rate

6.1 Pileup reweighting

The PU present in the MC samples does not exactly match the PU present in the data. The difference is corrected by reweighting simulated events to match the PU distribution in data [44]. We use the PU reweighting based on the mean of the Poisson distribution. The corresponding PU distribution in data is computed as described in Ref. [45], using a value of 69.2 mb for the inelastic pp scattering cross section. A problem has been reported with the simulation of PU in the RunIIFall17 MC production [46]. We work around the problem using the distribution in `PileupSummaryInfo::getTrueNumInteractions`, obtained individually for each MC sample before any cuts are applied, as input for the PU reweighting.

6.2 Trigger efficiency

Differences in efficiency for events in data and MC simulation to pass the triggers given in Table 3 are corrected by applying the ratio of the efficiency in data to the efficiency in the MC simulation as weight to simulated events. We refer to these weights as “scale factors” (SF). The trigger SFs are measured as function of lepton multiplicity and type of the leptons as well as of the leading lepton p_T . An additional uncertainty of 2% is attributed to 2ℓ -triggered events, associated with the parametrization of the trigger efficiency as a function of the leading lepton p_T alone.

The SF are determined by comparing the combined efficiencies of single, double, and triple lepton triggers between data and MC simulation using events recorded by E_T^{miss} triggers. The sample predominantly consists of $t\bar{t}$ +jets events. The scale factors were derived in a similar analysis [34] using this same method are re-used here. The SFs are given in Table 13.

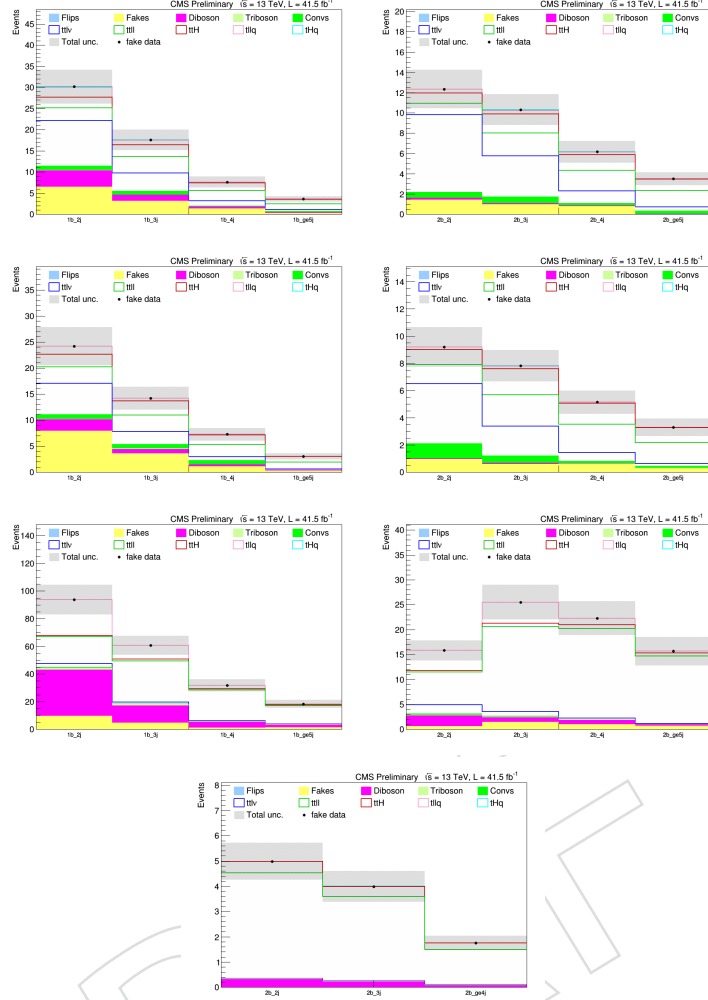


Figure 2: SM Expected number of signal and background events in the 3l and 4l categories. Going from left to right and top to bottom on a per-row basis, the plots show events in the 3l (1b “p”), 3l ($\geq 2b$ “p”), 3l (1b “m”), 3l ($\geq 2b$ “m”), 3l (SFOS Z, 1b), 3l (SFOS Z, $\geq 2b$) and $\geq 4l$ categories, respectively.

6.3 Identification and isolation efficiency for e and μ

The efficiency for electrons and muons to pass the loose and the tight selection criteria defined in Tables 9 and 10 have been measured via the Tag-and-Probe technique using samples of $Z/\gamma^* \rightarrow ee$ and $Z/\gamma^* \rightarrow \mu\mu$ events. The measurement described in Ref. [47] has been repeated for data recorded in 2017 (see Ref. [35] for details on the 2017 measurement). Briefly, the efficiencies are measured separately for electrons and muons and are parametrized as function of lepton p_T and η . The measurement is performed in two stages: the efficiency for leptons to pass the loose selection criteria is measured first, and then the conditional probability for leptons that pass the loose selection criteria to also pass the tight selection criteria is measured. The ratio of the efficiency measured in data to the efficiency in MC simulation yields a SF per lepton. The product of per-lepton SF is applied as weight to simulated events.

6.4 b-tag efficiency and mistag rate

Small differences between data and MC simulation in the efficiency for b-jets and c-jets to pass the loose and medium WPs of the DeepCSV algorithm, and in the mistag rate for light flavor (u,

Lepton multiplicity and flavour	Leading lepton p_T	SF
2μ	$p_T < 35 \text{ GeV}$	0.972 ± 0.006
	$p_T \geq 35 \text{ GeV}$	0.994 ± 0.001
$e + \mu$	$p_T < 35 \text{ GeV}$	0.952 ± 0.008
	$35 \leq p_T < 50 \text{ GeV}$	0.983 ± 0.003
	$p_T \geq 50 \text{ GeV}$	1.000 ± 0.001
$2e$	$p_T < 30 \text{ GeV}$	0.937 ± 0.027
	$p_T \geq 30 \text{ GeV}$	0.991 ± 0.002
$\geq 3\ell$	—	1.000 ± 0.050

Table 13: Trigger efficiency SF applied to simulated events selected in different categories.

d, s) quark and gluon jets have been observed by the BTV POG. We correct for the differences by applying suitably chosen weights to simulated events, using SF and tools (“method 1d”) provided by the BTV POG [48]. The per-jet SF are parametrized as function of jet p_T and η , of the b-tagging discriminant of the jet, and of the flavor of the quark or gluon that is matched to the jet on generator level. The per-event weight is then taken as the product of per-jet weight of all selected jets in the event.

6.5 ECAL prefiring correction

In 2016 and 2017, the gradual timing shift of ECAL was not properly propagated to Level-1 trigger primitives (TP) resulting in a significant fraction of high eta TP being mistakenly associated to the previous bunch crossing. Due to Level-1 trigger rules, events can self-veto if a significant amount of ECAL energy is found in the region of $2 < |\eta| < 3$. This effect is not described by the simulations, which must be corrected to match the data. We apply a correction to each of our MC samples following the procedure described in [49]. We found this to have a small effect on the analysis. The uncertainty on the correction is taken from the maximum between 20 percent of the object prefiring probability and the statistical uncertainty associated to the considered bin in the prefiring map. We found that the up/down variation of the correction by its uncertainty had a negligible effect on the normalization and shape of the njets distributions in our signal regions, and so we do not include it as a source of systematic uncertainty.

7 Background estimation

In this section we describe how we assess the contributions of the various backgrounds to events in our signal regions. Here, we distinguish between “reducible” and “irreducible” background contributions.

A background is considered as “reducible” in case one or more of the reconstructed electrons or muons passing the tight object selection criteria detailed in Section 4.1.4 are not due to genuine prompt leptons. In the $2\ell ss$ channels, a further source of reducible background arises from events containing lepton pairs of opposite charge, in which the charge of either lepton is mismeasured. Both of these reducible backgrounds are determined from data. We refer to the former as “fake” and to the latter as charge “flip” background. The procedures used to estimate the fake and charge flip backgrounds is described in Sections 7.1 and 7.2, respectively.

While the production of top quark pairs in association with virtual photons is included as part of the $t\bar{t}ll$ signal, the production of top quark pairs in association with real photons makes a small contribution to the reducible background. This background has been studied in Refs. [37, 47, 50, 51]. The contribution of $t\bar{t}$ events with real photons is typically due to asymmetric conversions of the type $\gamma \rightarrow e^+e^-$, in which either the electron or the positron carries most of the energy of the photon, while the other electron or positron is of low energy and fails to get reconstructed. Such events are suppressed by the photon conversion rejection criteria and by requiring that the tracks of electron candidates have hits in each layer of the pixel detector that is crossed by the track (see Table 9). The contribution of this background is modeled using MC simulation.

Irreducible background contributions are modeled using the MC simulation. The dominant contributions are due to the production of W or Z boson pairs in association with jets. Minor contributions arise from triboson production. The modeling of the data by the simulation is validated in a $3l$ control region enriched in the contribution of the dominant irreducible background process: WZ+jets. Figure 4 shows plots of event-level quantities in this control region.

7.1 “Fake” background

The estimation of the fake background is based on the fake-factor (FF) method. The method is applied to each of the $2\ell ss$, 3ℓ , and 4ℓ channels separately. Because we use an identical set of lepton identification techniques as is used in the $t\bar{t}H$ multilepton analysis [34] (see Section 4), and explicitly check that we remain synchronized with that analysis on an object-ID level (see Appendix A), we are able to use the exact same FF weights derived in that analysis here. We explicitly check the modelling of the non-prompt background using this method in a dedicated $2\ell ss$ control region, as shown in Fig 5.

The FF method selects events which pass all selection criteria for the respective category, except that the leptons in these events are required to pass the fakeable instead of the tight object selection criteria. The selected event sample is referred to as the “application region” (AR) of the FF method. Events in which all e and μ pass the tight object selection criteria are vetoed in order to ensure orthogonality with the signal region (SR). An estimate of the fake background in the SR is obtained by applying appropriately chosen FF weights as a function of lepton p_T and η to the events selected in the AR. These weights are shown in Figure 3. Again, the FF weights are derived in another analysis [34] and reused here. The method for deriving the FF weights is described in detail in [34, 37, 47] and is summarized below.

The FFs are derived in a “measurement region” (MR) composed of an event sample enriched in multijet production, which is selected by requiring the presence of exactly one electron or

muon that passes the fakeable lepton selection criteria given by Tables 9 and 10 and at least one jet, which is separated from the lepton by $\Delta R > 0.7$. The events are recorded using single electron and muon triggers. Events selected in the multijet sample are required to contain at least one “offline” reconstructed jet with a p_T above the p_T threshold of the trigger.

The offline reconstructed leptons are required to satisfy conditions on the reco- p_T and on the cone- p_T . The requirement on the reco- p_T is applied to ensure that the trigger is fully efficient, while the condition on the cone- p_T is applied to reduce biases in the measurement of the lepton misidentification rates. The p_T thresholds that are applied on trigger level impose an implicit condition on the isolation of the lepton, because the lepton misidentification rates are parametrized by the cone- p_T , and not the reco- p_T , of the lepton. Such bias on the isolation of the lepton would cause a bias on the output of the BDT that is used to separate prompt leptons from non-prompt and fake leptons, increasing the probability for fakeable leptons to pass the tight lepton selection criteria. This in turn would cause the estimate of the fake background in the SR to be overestimated, as all leptons selected in the MR are required to pass the trigger requirements, whereas events selected in the AR contain multiple leptons and only one of these leptons must pass the trigger requirements. A bias of this kind is avoided by requiring that the cone- p_T of the offline reconstructed lepton exceeds the p_T threshold applied on trigger level by about a factor two. The condition $I_\ell < 0.4 \times \text{reco-}p_T$ that is applied in the selection of fakeable electrons and muons then ensures that leptons passing the threshold on the cone- p_T pass the p_T threshold of the trigger regardless of their isolation, avoiding the aforementioned bias. A small residual bias, which is caused by the electron identification criteria applied on trigger level, is reduced by requiring fakeable electrons to satisfy conditions on $\sigma_{i\eta i\eta}$, H/E, and $1/E - 1/p$ (cf. Table 9).

The FFs are measured in bins of η and cone- p_T of the lepton. In each bin, the the FF is measured by determining the number of multijet events with fakeable leptons that pass the tight lepton selection criteria and those that fail. The number of multijet events in the pass and fail regions, denoted by N_{pass} and N_{fail} , is determined by a maximum-likelihood (ML) fit of the distribution in the observable:

$$m_T^{\text{fix}} = \sqrt{2 p_T^{\text{fix}} E_T^{\text{miss}} (1 - \cos \Delta\phi)} \quad (9)$$

with shape templates for W+jets, Z+jets, diboson, top quark pair, and multijet production. The pass and fail regions for each p_T and η bin are fitted simultaneously, while individual p_T and η bins are fitted independently. Given the event yields N_{pass} and N_{fail} of multijet production in the pass and fail regions, the FF f_i for a given bin i in η and cone- p_T of the lepton are computed by:

$$f_i = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}} \quad (10)$$

The symbol $\Delta\phi$ in Eq. (9) refers to the angle in the transverse plane between the lepton momentum and the E_T^{miss} vector, while the symbol p_T^{fix} denotes a constant of value 35 GeV. The observable m_T^{fix} exploits the fact that the E_T^{miss} reconstructed in multijet events is mainly due to resolution effects and is typically small, resulting in a falling distribution in m_T^{fix} , while the distribution in W+jets events, which constitute the main background for the FF measurement, exhibits a broad maximum around $m_W \approx 80$ GeV. Compared to the usual transverse mass, the observable m_T^{fix} has the advantage of not depending on the p_T of the lepton, which is better suited for the purpose of measuring the FF in bins of lepton p_T [47].

A correction needs to be applied to the measured FF for electrons to account for the contribution of photon conversions to the pass and fail regions. The event yields N_{pass} and N_{fail} obtained by the fit are scaled by the relative fraction of multijet events in which the reconstructed fakeable

electron is matched, within $\Delta R < 0.3$, to a generator level prompt photon (i.e. not a photon arising from the decays of hadrons) The fraction is determined separately for the pass and fail region and for each bin in cone- p_T and η of the lepton. The effect of this correction is to reduce the FF for electrons by 10-20%.

Small differences between the FF in multijet and top quark pair events arise from a higher fraction of fake leptons from light quark and gluon jets in multijet events, and top quark pair events containing a higher fraction of non-prompt leptons from decays of heavy (c and b) quarks. The differences between the FF have been reduced by tuning the fakeable lepton selection criteria, in particular by adjusting the cut on the CSVv2 b-tagging discriminant that is applied in the selection of fakeable electrons and muons. The small residual differences are not corrected for, and are covered by the uncertainties on the FF weights.

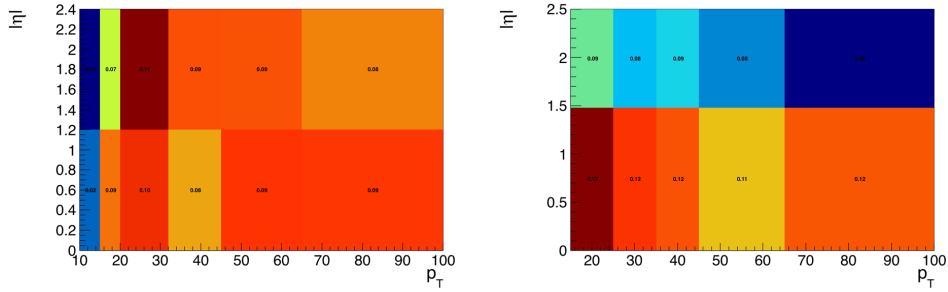


Figure 3: Maps of the p_T - and η -dependant fake-factor (FF) weights for muons (left) and electrons (right). The weights are used to estimate the background due to non-prompt leptons.

7.2 Charge “flip” background

The charge “flip” background in the $2\ell ss$ categories is dominated by $t\bar{t}$ +jets events with two prompt leptons, produced in the decay $t\bar{t} \rightarrow bW^+bW^- \rightarrow b\ell^+\nu b\bar{\ell}^-\bar{\nu}$, in which the charge of either prompt lepton is mismeasured. The background is estimated from data, following a strategy similar to the one used for the estimation of the fake background. Again, as with the FF weights described in Section 7.1, we make use of the exact same set of charge flip weights derived by the $t\bar{t}H$ multilepton analysis [34] for this background.

The charge misidentification rates for electrons are extracted from data, in events with two same-sign electrons with invariant mass close to the mass of the Z boson. Electron pairs in the peak are overwhelmingly likely to be from $Z/\gamma^* \rightarrow ee$ events; i.e., real opposite sign electron pairs with a wrongly assigned charge on one leg. Charge misassignment probabilities are calculated for different bins of electron p_T and η by extracting same-sign and opposite-sign event yields categorized in the kinematics of the two lepton legs. In each category, the event yield of electron pairs from Z decays is determined from a fit to the invariant mass shape, and depends on the charge misassignment probabilities of each leg. The invariant mass shape is modeled with a crystal ball and Breit-Wigner function for the signal and an exponentially falling function for the backgrounds. Electron kinematics are separated in three p_T (15–25 GeV, 25–50 GeV, and ≥ 50 GeV) and two η bins (0–1.479 and 1.479–2.5), resulting in a total of 21 distinct categories of electron pairs. The six charge misassignment probabilities are then determined in a simultaneous fit to the 21 same-sign and opposite-sign event yields. The resulting misassignment probabilities range between about 0.02% in the barrel and about 0.2% in the end caps and are shown in Table 14.

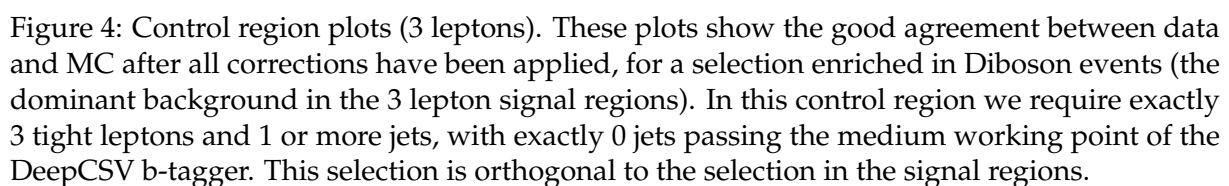
Alternative measurements using template histograms obtained from the simulation are also performed. These templates are used to model the $Z/\gamma^* \rightarrow ee$ signal and background, in both

Table 14: Electron charge misassignment probabilities (in percent) as determined in data. These probabilities are used to estimate the background contribution to $2\ell_{ss}(2e)$ and $2\ell_{ss}(e\mu)$ categories where the charge of one electron has been misidentified.

	$15 \leq p_T < 25 \text{ GeV}$	$25 \leq p_T < 50 \text{ GeV}$	$50 \text{ GeV} \leq p_T$
$0 \leq \eta < 1.479$	0.0137 ± 0.0041	0.0243 ± 0.0042	0.0269 ± 0.0055
$1.479 \leq \eta < 2.5$	0.0178 ± 0.0071	0.0543 ± 0.0042	0.1742 ± 0.0163

the same-sign and opposite-sign regions. The advantage of using template histograms is that tails in the invariant mass distribution are better-modelled in the high-stats OS region. The electron charge misidentification rates measured with these different approaches differ by up to 30% [34]. This variation is treated as a source of systematic uncertainty when applying the charge misidentification rates to estimate the charge “flip” background in the $2\ell_{ss}$ categories.

The application region used to estimate the charge flip background contains events that pass all selection criteria of the SR, except that the two leptons are required to be of opposite charge. The sum of the probabilities to mismeasure the charge of either one of the two leptons is then applied as an event weight. The charge misidentification rates for muons are negligible [43]. Thus, we assess the contribution of the charge flip background to the $2\ell_{ss}(2e)$ and $2\ell_{ss}(e\mu)$ categories only, and neglect it otherwise. We explicitly check the modelling of the charge flip background in a dedicated $2\ell_{ss}$ control region, as shown in Fig 5.



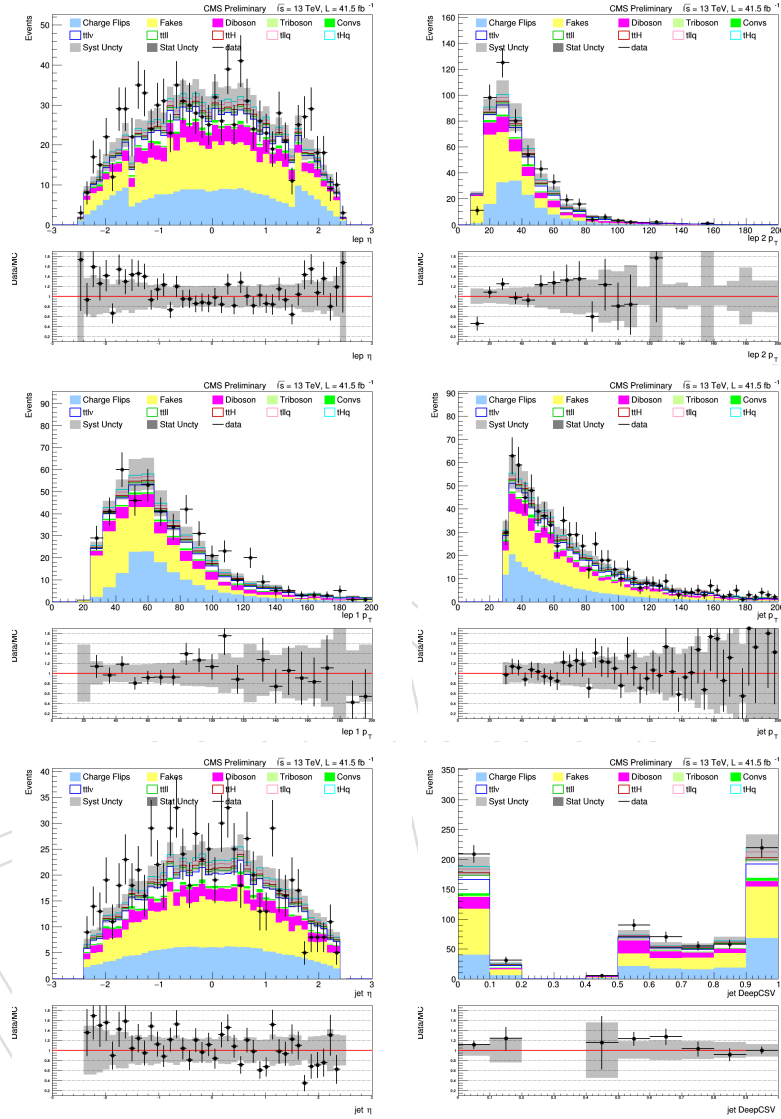


Figure 5: Control region plots (2 leptons). These plots show the good agreement between data and the data-driven charge flip and non-prompt backgrounds. The latter is the dominant background in the 2 lepton signal regions. In this control region we require exactly 2 tight, same-sign leptons, and 1 or 2 jets, with exactly 1 jet passing the medium working point of the DeepCSV b-tagger. This selection is orthogonal to the selection in the signal regions.

8 Signal extraction and EFT model parameters fitting

To properly assess a fit for EFT effects, we must heavily consider the information contained in our generated signal samples. Our privately produced samples are generated with weights on each event that scale with the values of the Wilson coefficients. Because of these weights, we can account for the changing kinematic profile of the samples for EFT scenarios both similar to the standard model and very significantly different from it. These weights also include interference effects between the operators themselves and with the Standard Model. This template morphing can be parameterized and used as the tool to fit the yields to determine the most likely EFT scenario. The details of this approach are described below.

The parameterization of the event yields is determined by reweighting the event yields to many different points in WC space. The sum of the resulting weights of the events in an analysis bin gives the yield for that bin. This result is identical to rescaling the inclusive cross-section to that point in WC space. The new yield for each bin follows a second degree polynomial as a function of the coefficients. By sampling multiple points in WC space, we can parameterize the quadratic function as function all 16 Wilson coefficients. The final quadratic function contains all of the cross-terms as well that account for interference between different Wilson Coefficients. With the parameterizations we can predict the yields for other points in WC space without needing to perform the reweighting again nor generating a new MC sample.

The parameterizations along with the yields for the SM point are stored in a custom ROOT histogram class called TH1EFT. This histogram class contains the n_{jet} distribution for a given process+category combination. The histogram bin values default to the standard model point in WC space. However, the histogram class can be reweighted to a different point in WC space through the "Scale" function. Within the TH1EFT class is the WCFit object. This is where the parameterizations are stored. The coefficients of terms in the parameterization are stored as items in an ordered vector of doubles. Another ordered vector of string pairs acts as a key to the coefficient vector. These pairs of strings reference the Wilson Coefficients involved in those terms. For example, the parameterization term with $\text{ct}W^2$ would have a string pair $\langle \text{"ctW"}, \text{"ctW"} \rangle$. Linear terms are of the form $\langle \text{"sm"}, \text{"ctW"} \rangle$ or $\langle \text{"ctW"}, \text{"sm"} \rangle$. The constant term, or in other words the standard model yield, is given by the pair $\langle \text{"sm"}, \text{"sm"} \rangle$. To obtain the coefficient for a specific term, one must find the index of the string pair of the desired term and obtain the coefficient at the corresponding index. This process is expediated through the "getCoefficient" function.

This analysis uses two repositories for the process of obtaining best fit results and limits. Both can be found in the "cms-govner" group's github page ¹. The first is a fork of the CombineHarvester package [52]. The scripts here build the single datacard needed for use in the Higgs PAG "Combine" tool [53]. The other "cms-govner" repository named "EFTFit" provides the scripts and modules necessary to run combine with our specific physics model.

The CombineHarvester package streamlines the process of building the datacards. The top-level script is "DatacardMaker.py". This immediately calls "HistogramProcessor.py" to retrieve all of the information from the input ROOT tree, which is filled with TH1EFT histograms. Each histogram's name specifies which category, process, and relevant systematics that histogram contains data for. At the top of HistogramProcessor.py, the signals, backgrounds, and data processes are specified. Only those named here will be extracted from the ROOT tree. The script then loops through all of the histograms to extract the yield information. The observation part of the datacard is set to be the asimov dataset by default (sum of signal+background

¹<https://github.com/cms-govner>

predictions at the SM point). If the “fake.data” switch is turned on, the signal processes’s contribution to the observation is scaled to a specified point in WC space. This is used to test the robustness of the fit and sensitivity to different values of the Wilson Coefficients. For the systematics, the yield from the histograms is normalized to the nominal yield so that the ratio can be correctly used in the datacard. At the end of “HistogramProcessor” all of the categories are scanned. Any category with signal or background yield ≤ 0.01 does not get used. The objects returned are lists of the accepted categories, data process names, signal process names, background process names, and systematics extracted along with python ‘dict’ objects for the data yields, fake data yields, signal yields, and background yields. The signal and background dicts are encoded with two keys, the process and category. The data and fake data dicts are only encoded by the category. The DatacardMaker script takes all of these objects and feeds them into an instance of CombineHarvester. The datacard is automatically generated at the end of this script and named “EFT_MultiDim_Datacard.txt”.

The EFTFit repository supports running the physics model for this analysis in the Combine tool. The main script is “EFTFitter.py”. Within this python class are functions to do various types of fits and scans over Wilson Coefficients. Due to the complexity of the fit, first a workspace must be generated using the ‘makeWorkspace’ function. This workspace named “16DWorkspace.root” uses a custom physics model file named EFT16DModel.py. It also takes a numpy file containing the EFT parameterizations as an argument. The EFT16DModel uses the parameterizations from the numpy file to scale the yields. These are created in the workspace during the “setup” phase. The functions are quite large, so they are initially compiled in smaller chunks before the full function is compiled in the RooWorkspace. When Combine asks for the yield to be scaled for some point in WC space, the model returns those functions evaluated at that point. The EFT parameterization file is generated by a separate script, FitConversion16D.py. This script takes the ROOT trees as input and saves the parameterizations extracted from the WCfit objects to the numpy file as a dict of dicts. The keys for the outer dict are the process and category. The keys of the inner dict are the two Wilson Coefficient strings needed to extract the fits, as described above. The values stored are that term’s contribution to the yield normalized to the standard model yield. (For example, one term might be stored as $[('ttH', '2lss_m_emu_2b_5j)][('ctW', 'sm')] = 0.112$.)

After the workspace has been generated, Combine can be invoked normally. Various functions in EFTFit.py provide a preferred set of arguments for combine. This analysis simultaneously fits for multiple Wilson Coefficients, so these functions all use the MultiDimFit method. To compute the best fits for each Wilson Coefficient, we scan over each of them one at a time. In this 1D scans, the other 15 Wilson Coefficients are either left floating as unconstrained nuisance parameters or frozen to their Standard Model values. The scan point with the lowest value in negative log-likelihood is the best fit. A central 95% uncertainty interval is also calculated.

9 Systematic uncertainties

There are three types of systematic effects considered in this analysis: those that affect only the rates of signal or background processes, those that affect only the shapes of the n_{jet} distribution for signal or background processes, and those that affect both the rate and the shape (since we are doing a counting experiment, here “shape” simply refers to migrations between n_{jet} bins). In the last case, the rate and shape effects are treated simultaneously so that they are considered completely correlated. Unless otherwise noted, all of the uncertainties listed here apply equally to signal and background and are treated as 100% correlated between the two. Below is a list of systematic effects considered for this analysis.

Luminosity: Rate systematic. The uncertainty on the luminosity estimate is 2.3%. [5].

Muon and electron ID, isolation: Rate + shape systematic. Following the recommendations of the MUO POG, a total rate+shape uncertainty is assigned to the tracking, muon ID and isolation [54] scale factors used to correct the efficiency in MC to reproduce the efficiency in data, which are derived with a tag-and-probe method. The total uncertainty of the resulting muon scale factors are the quadratic sum of the statistical uncertainties of the scale factors (as returned by `TH1 : : GetBinError()` of the SF histograms), a 0.5% tracking systematic uncertainty, a 1% ID systematic uncertainty, and a 0.5% Iso systematic uncertainty. Similarly, following the recommendations of the EGM POG, the p_T and η dependent uncertainties of the electron ID and Gsf tracking efficiency scale factors are propagated to the final fitting variable distributions [55]. Uncertainties are typically of the order of 1-2% per lepton.

Trigger efficiency: Rate + shape systematic. The impact due to the trigger efficiency is estimated as well by varying the scale factors within its uncertainties, as a function of p_T of the leading lepton as shown in Table 13.

Pileup: Rate + shape systematic. Effects due to the uncertainty in the distribution of the number of pileup interactions are evaluated by varying the cross section used to predict the number of pileup interactions in the MC by 4.6% from its nominal value. The changes in the weight factor are propagated to the final fitting variable distributions.

Jet energy scale (JES): Rate + shape systematic. The JES systematic is evaluated by shifting the jet energy scale applied to the reconstructed jets up and down by one standard deviation using the standard JETMET procedure. [56] The events are then reanalyzed, including reapplying the jet-based selection and recalculating all kinematic quantities.

b -tag scale factors: Rate + shape systematics. We are using differential b -tagging scale factors described in [42]. There are three sources of systematic uncertainty on both the heavy-flavor and light-flavor scale factors: JES, purity of heavy-flavour or light-flavour in the control samples used to derive the SF, and the statistics of the samples used to derive the SFs. Each of these systematic variations is considered separately.

JES: The b -tag uncertainty associated with JES is evaluated at the same time the JES uncertainty is considered. In other words, when the JES is shifted for the jet kinematics up or down by 1σ , the b -tag scale factor values are also shifted accordingly. Thus, the b -tag uncertainty from JES is considered completely correlated with the JES uncertainty.

purity: The purity uncertainty is controlled by a separate nuisance parameter for light and for heavy flavor. Shifting this parameter to $\pm 1\sigma$ corresponds to changing the

heavy-flavour (light-flavour) SF while having a higher or lower contamination of light-flavour (heavy-flavour) jets in the control sample, according to the errors described in [57].

sample statistics: Finally, the impact of statistical uncertainties associated with the size of the samples used to derive the SF are controlled by means of four total nuisance parameters, two for heavy-flavor and two for light-flavor. In each class of jets, the first nuisance parameter controls distortions in the CSV distribution corresponding to an overall tilt that would be consistent with the statistical uncertainties on the SF, while the second nuisance parameter controls distortions of a more complicated nature, where the upper and lower ends of the distribution change relative to the center.

For charm jets scale factors, we retain the overall relative uncertainty from the heavy flavor scale factors, double it in size, and use it to construct two separate nuisance parameters to control the uncertainties. These two uncertainties associated with charm jets scale factors will be treated independently with respect to all the uncertainties for the heavy flavor and light flavor scale factors. More details on these uncertainties are given in Ref. [57].

Uncertainty on the Lepton Fake Rate estimate: Rate + shape systematic. We estimate the uncertainty on the lepton fake rate following the procedure described in [47]. Several sources of systematic uncertainty are considered. The measurement of the FF weights is affected by the statistics in the MR, subtraction of prompt lepton contamination in the MR, as well as differences in the background composition between MR (dominated by multijet background) and AR (dominated by $t\bar{t}$ +jets background). The effect on the fake rate due to the overall uncertainty of the FF measurement is taken into account by varying the entire map of fake factor weights up or down by one standard deviation. This is the largest source of uncertainty on the fake rate, and amounts to ≈ 30 -60%, depending on the njet bin (see Figs 6 and 7). In addition, limited statistics in the AR of the FF method have a significant effect on the estimate of the fake rate and must be considered a separate source of uncertainty. This again varies with njet bin, and amounts to ≈ 10 -50%. We treat this latter uncertainty as completely uncorrelated between bins.

Uncertainty on the Charge Flip estimate: Rate systematic. We assign an uncertainty of 30% on the yield of the Charge Flip background, due to differences in the charge flip probabilities obtained by varying the measurement method (see section 7.2). The uncertainty due to finite statistics in the charge flip AR is negligible and is not considered.

Cross sections: Rate systematics. The expectation for signal and background yields are derived from theoretical predictions of at least NLO accuracy. Uncertainties affecting these normalizations are summarized in Table 15.[58]Where appropriate, factors contributing to these uncertainties that are common to multiple processes are treated as 100% correlated. For signal processes, this uncertainty is considered on the whole process (SM+EFT). Therefore, it will also cover theory uncertainties on the EFT part as far as the theory uncertainties on the EFT part are not bigger than the SM part. Since we are looking at the same process for the SM and EFT, we expect to find similar uncertainties of the SM and EFT parts.

PDF shape variations Shape systematic. The shape variation of the final fitting variable distributions due to the uncertainty on the PDF set is estimated by reweighting the spectra according to the 100 replica sets given by the NNPDF30_nlo_as_0118 PDF parameterization.

Process	pdf				QCD Scale						
	$gg_{t\bar{t}H}$	gg	$q\bar{q}$	$qg tHq$	$t\bar{t}V$	$t\bar{t}\gamma$	$t\bar{t}H$	tHq	V	VV	VVV
$t\bar{t}H$	3.6%						-9.2%/+5.8%				
$tllq$			4%						1%		
$t\bar{t}l\nu$			2%		-12%/+13%						
$t\bar{t}ll$		3%			-12%/+10%						
tHq				3.7%				-8%/+6%			
DiBoson			2%							2%	
TriBoson			4.2%								2.6%
$t\bar{t}\gamma$ +jets		5%				10%					

Table 15: Cross section (rate) uncertainties used for the fit. Each column in the table is an independent source of uncertainty. Uncertainties in the same column for different processes (different rows) are completely correlated.

The total uncertainty is estimated as the envelope of all variations using the algorithm implemented in the `uncertainty` method provided by the `LHAPDF::PDFUncertainty` interface [59, 60].

Renormalization and factorization scale: Shape systematics. Uncertainties due to the renormalization scale μ_R and the factorization scale μ_F in the `t\bar{t}matrix`-element generator are modelled by varying the scales independently by a factor of 0.5 or 2, and propagating the changes to the final fitting variable distribution in the fit. This is accomodated via weights obtained directly from the generator information. Since the normalization uncertainties of the matrix-element generator are covered by the cross setion uncertainties (listed in Tab. 15), only the shape variation of discriminant distributions is considered, i.e. they are scaled to retain the overall normalization.

Parton shower: Shape systematic. The uncertainty in the parton-shower simulation is estimated by varying the renormalization scales for initial- and final-state radiation up and down by a factor of 2 and $\sqrt{2}$. Only an incomplete subset of processes we consider had MC available that included this variation, so we evaluated the effect using the $t\bar{t}Z$, $t\bar{t}W$ and tZq samples listed in Table 5. We found this to have a negligible effect on the normalization, and a small effect on the shape of the fitting variable distributions, and is dominated by the ISR component of the systematic. So, we account for this uncertainty by introducing a single $\pm 5\%$ linear distortion in shape of the fitting variable distribution for all the categories, and for all MC samples.

Table 16 summarizes the systematic uncertainties assessed on the signal and backgrounds for this analysis. It describes how each systematic is treated in our fit to extract the amount of signal.

Table 16: Summary for the systematic uncertainties considered on the inputs to the limit calculation. Here “shape” means that the systematic accounts for the migration of events from one njet bin to another but not the overall normalization (unless it is also a rate systematic). Except where noted, each row in this table will be treated as a single, independent nuisance parameter.

Source	Type	Notes
Luminosity	rate	Signal and all MC backgrounds
Lepton ID/Iso	rate+shape	Signal and all MC backgrounds
Trigger efficiency	rate+shape	Signal and all MC backgrounds
Pileup	rate+shape	Signal and all MC backgrounds
JES	rate+shape	Signal and all MC backgrounds
b-Tag HF fraction	rate+shape	Signal and all MC backgrounds
b-Tag HF stats (linear)	rate+shape	Signal and all MC backgrounds
b-Tag HF stats (quadratic)	rate+shape	Signal and all MC backgrounds
b-Tag LF fraction	rate+shape	Signal and all MC backgrounds
b-Tag LF stats (linear)	rate+shape	Signal and all MC backgrounds
b-Tag LF stats (quadratic)	rate+shape	Signal and all MC backgrounds
b-Tag Charm (linear)	rate+shape	Signal and all MC backgrounds
b-Tag Charm (quadratic)	rate+shape	Signal and all MC backgrounds
Lepton Fake Rate	rate+shape	uncertainty on data-driven non-prompt background
Charge Flip	rate	uncertainty on data-driven charge flip background
QCD Scale ($t\bar{t}H$)	rate	Scale uncertainty for NLO $t\bar{t}H$ prediction
QCD Scale ($t\bar{t}\gamma$)	rate	Scale uncertainty for NLO $t\bar{t}\gamma$ prediction
QCD Scale ($t\bar{t}V$)	rate	Scale uncertainty for NLO $t\bar{t}l\nu$ and $t\bar{t}ll$ prediction
QCD Scale (tHq)	rate	Scale uncertainty for NLO tHq prediction
QCD Scale (V)	rate	Scale uncertainty for NNLO W and Z prediction
QCD Scale (VV)	rate	Scale uncertainty for NLO diboson prediction
QCD Scale (VVV)	rate	Scale uncertainty for NLO triboson prediction
pdf (gg)	rate	Pdf uncertainty for gg initiated processes except $t\bar{t}H$ ($t\bar{t}Z$)
pdf ($gg t\bar{t}H$)	rate	Pdf uncertainty for $t\bar{t}H$
pdf ($q\bar{q}$)	rate	Pdf uncertainty for $q\bar{q}$ initiated processes ($t\bar{t}W$, W , Z).
pdf ($qg tHq$)	rate	Pdf uncertainty for tHq
PDF	shape	Pdf shape variations for all MC processes
Q2 Scale Renorm.	shape	Renormalization scale uncertainties.
Q2 Scale Fact.	shape	Factorization scale uncertainties.
Parton Shower	shape	Parton-shower scale uncertainties.

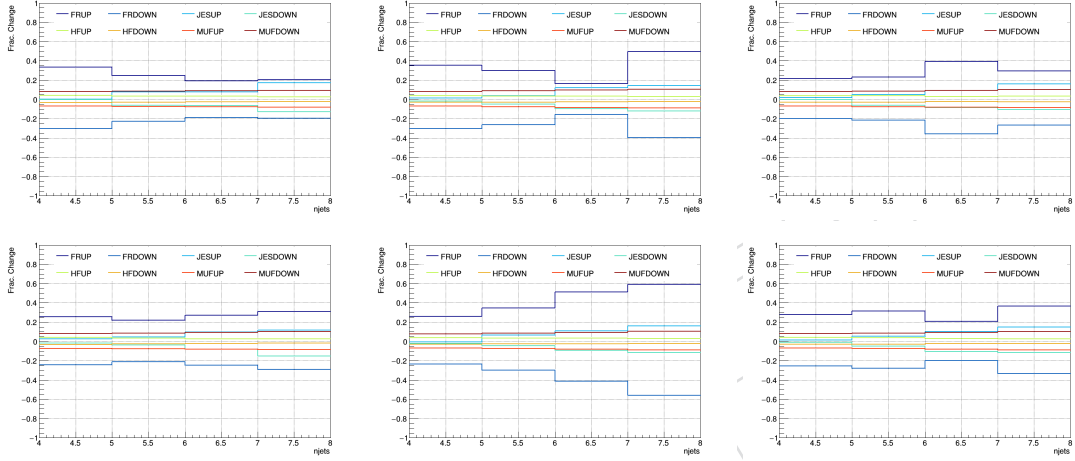


Figure 6: Fractional change in the expected yield w.r.t. the nominal yields for signal and background events, as a function of n_{jets} , as different systematic variations are considered. The systematics are a subset chosen to highlight those that have a relatively large effect on the shape of the n_{jets} distribution. The effect of the up/down variation of the FF measurement systematic (FRUP/DOWN) on the fake lepton background is shown. The rest of the curves show the effect of the up/down variation of different systematics on the SM $t\bar{t}H$ yield; they are the jet energy scale (JESUP/DOWN), the heavy-flavor btagging purity systematic (HFUP/DOWN), and the factorization scale systematic (MUFUP/DOWN). The 2lss categories are shown. From left to right in the top row, the plots show events in the 2lss (e^+e^+), 2lss ($\mu^+\mu^+$) and 2lss ($e^+\mu^+$) categories. From left to right on the bottom row, the plots show the 2lss (e^-e^-), 2lss ($\mu^-\mu^-$) and 2lss ($e^-\mu^-$) categories.

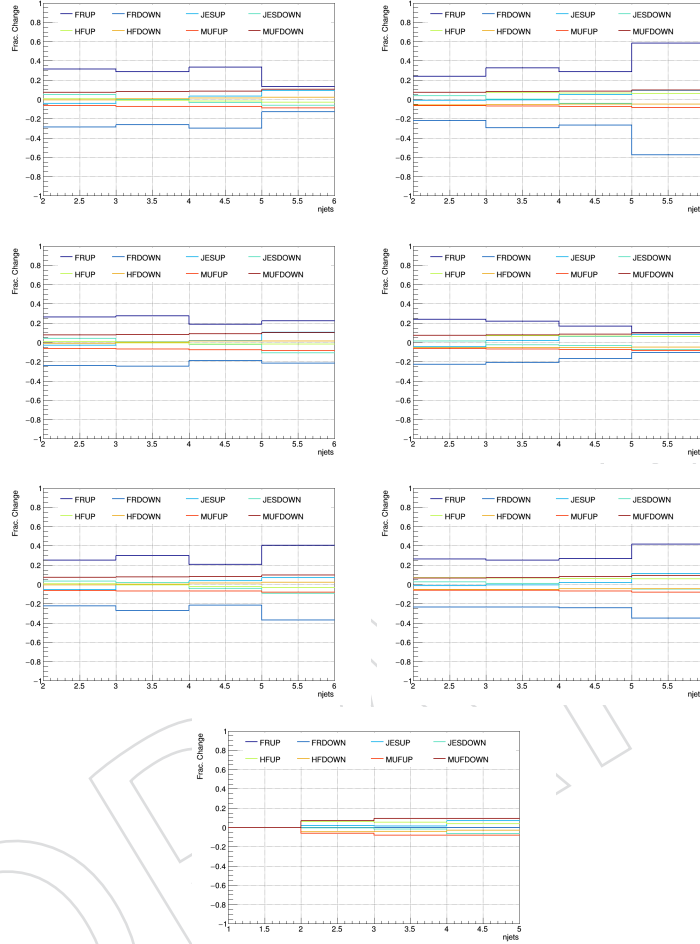


Figure 7: Fractional change in the expected yield w.r.t. the nominal yields for signal and background events, as a function of n_{jets} , as different systematic variations are considered. The systematics are a subset chosen to highlight those that have a relatively large effect on the shape of the n_{jets} distribution. The effect of the up/down variation of the FF measurement systematic (FRUP/DOWN) on the fake lepton background is shown. The rest of the curves show the effect of the up/down variation of different systematics on the SM $t\bar{t}H$ yield; they are the jet energy scale (JESUP/DOWN), the heavy-flavor btagging purity systematic (HFUP/DOWN), and the factorization scale systematic (MUFUP/DOWN). The 3l and 4l categories are shown. Going from left to right and top to bottom on a per-row basis, the plots show events in the 3l (1b “p”), 3l ($\geq 2b$ “p”), 3l (1b “m”), 3l ($\geq 2b$ “m”), 3l (SFOS Z, 1b), 3l (SFOS Z, $\geq 2b$) and $\geq 4l$ categories, respectively.

10 Results

10.1 Standard Model Signals

As a consistency check, a fit was performed on the signal strength of each of our five signal processes, after scaling them to SM expectation via their WCFits (see Sec. 8). The fit was carried out simultaneously in each of the five signal strengths, as well as individually with other processes frozen to $\mu = 1$. The results of both checks are shown in Table 18, along with comparisons of the best-fit signal strengths obtained in dedicated analyses.

SM Signal Strength Results (Asimov Dataset)			
Signal Process	Best Fit μ (Simultaneous)	Best Fit μ (Individual)	Dedicated Analysis
ttll	1.00 ± 0.17	1.00 ± 0.18	$1.00^{+0.15}_{-0.13}$ (TOP-18-009)
ttlnu	1.00 ± 0.29	1.00 ± 0.22	1.00 ± 0.22 (TOP-17-005)†
ttH	1.00 ± 0.60	1.00 ± 0.53	$1.00^{+0.39}_{-0.35}$ (HIG-18-019)
ttlq	1.00 ± 0.51	1.00 ± 0.44	$1.00^{+0.23}_{-0.21}$ (TOP-18-008)
tHq	1.05 ± 14.9	1.03 ± 13.1	$1.00^{+0.70}_{-0.65}$ (HIG-18-009)†

Table 17: Best fit values of the different SM signal strengths for each of this analysis' signal processes. The second column contains the results from a simultaneous fit of all signal strengths. The third column contains the results from a fit on each signal strength individually (others frozen to their SM value of 1). The fourth column shows the expected signal strengths in analyses dedicated to measuring individual processes using a similarly-sized dataset. The dedicated analyses each use the full 2017 dataset, except those marked with a †, which use the 2016 dataset.

In general, we found the results of the fits to be consistent with SM expectation, with the fits to individual processes having somewhat lower uncertainty than the simultaneous fit to all processes. This is expected due to correlations among the μ parameters (see Fig. 8). Further, we find the uncertainties on the expected μ parameters to be generally larger than those obtained in dedicated analyses, which is also expected as each dedicated analysis is finely tuned to be as sensitive as possible to a given process. For example, the tZq analysis uses a lepton ID tuned for that specific process, as well as a BDT to discriminate between signal and background events. We use an identical object selection to the ttH multilepton analysis, but they consider taus, and consequently have a different event selection with more categories, which boosts sensitivity. They also do a fairly complex multivariate analysis and fit to the shape of the output distribution, whereas our result is a counting experiment. Our analysis is in no way optimized for sensitivity to SM tHq (in contrast to the dedicated analysis), but is instead sensitive to the effect of WC variations that enhance its cross section. When comparing to the dedicated ttW and ttZ results, our results are more comparable than those for the other signals, mainly due to the greater number of events expected from these processes. In the case of ttW, the additional factors of the difference in integrated luminosity between 2016 and 2017, as well as the fact that we consider off-Z-shell 3 lepton channels that the dedicated analysis does not, serve to bring our result into line with the dedicated one. Both ttW and ttZ make up a large fraction of our yield, with ttZ being the dominant process (among both signal and background processes) in the 3l (SFOSZ) and 4l categories. Optimizations to object ID, event selection or signal extraction

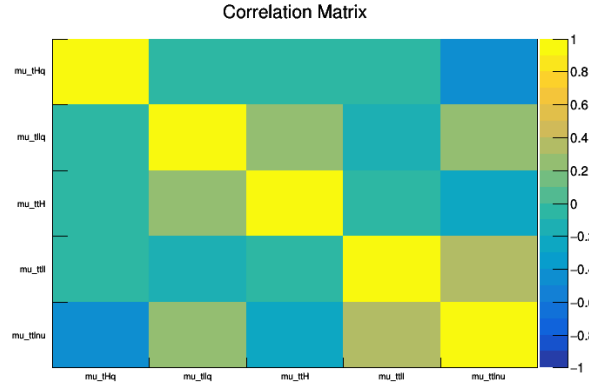


Figure 8: Correlation Matrix of Standard Model signal strengths. This is a result of a simultaneous fit of all five signal strengths on the SM-expected asimov dataset. The processes ttlnu and ttll are moderately correlated, as are ttH and ttllq and ttllq and ttlnu. There is also a significant anticorrelation between tHq and ttlnu. The percent variance of the anticorrelation (40%) is larger than the correlations (10-15%). Overall, we thus expect to see a larger uncertainty on the simultaneous fit to multiple signals than from a dedicated analysis that fits signals individually.

in the dedicated analyses therefore do not have as large of an effect in this case.

Figures 9-13 show the pull and impacts for each signal strength fit to the SM asimov dataset. For each fit, the other signal strength parameters are allowed to float freely.

10.2 EFT Signals

Due to the complex nature of the wilson coefficients' interactions and overlapping effects, there are many different ways to consider the results of this analysis. First, there is the consideration of the best fit values. Second, the uncertainty contours around this point exists in a 16-dimensional hyperprism, so we also present a series of 2D projections of this hyperprism along with 1D projection. All results are obtained using the the HiggsAnalysis-CombinedLimit tool as described in Section 8.

Tables 18 and 19 show the best fit values for each coefficient and their 2σ uncertainty intervals. Error bars are shown for two scenarios. In the black, the uncertainties intervals are calculated while the other Wilson Coefficients are treated as unconstrained nuisances. In the red, the uncertainties are determined while the other Wilson Coefficients are frozen to their Standard Model values of 0. Wilson Coefficients with multiple, distinct uncertainty intervals have multiple intervals listed. These results are displayed graphically in Figure 14.

Figure 15 shows the 1D likelihood scans of each of the 16 Wilson coefficients for the Standard Model expectation. These fits were made by either treating the other coefficients as floating nuisance parameters or by freezing the other coefficients to their Standard Model value of 0. These two results are overlaid with lines at the 1σ and 2σ uncertainty value for both.

The images in Figure 16 are the 2D likelihood scans of pairs of coefficients for the Standard Model expectation. The plots show the 1σ , 2σ , and 3σ contours in this cross section of the phase space. Also shown are markers for the SM expected result. These are split into two sets based on their treatment of the other 14 coefficients; either they are treated as unconstrained nuisance parameters (left column) or fixed to their Standard Model values of 0 (right column). In the case that the coefficients were left to float, the contours represent the profiled contours of the 16D hyper-prism on this 2D plane.

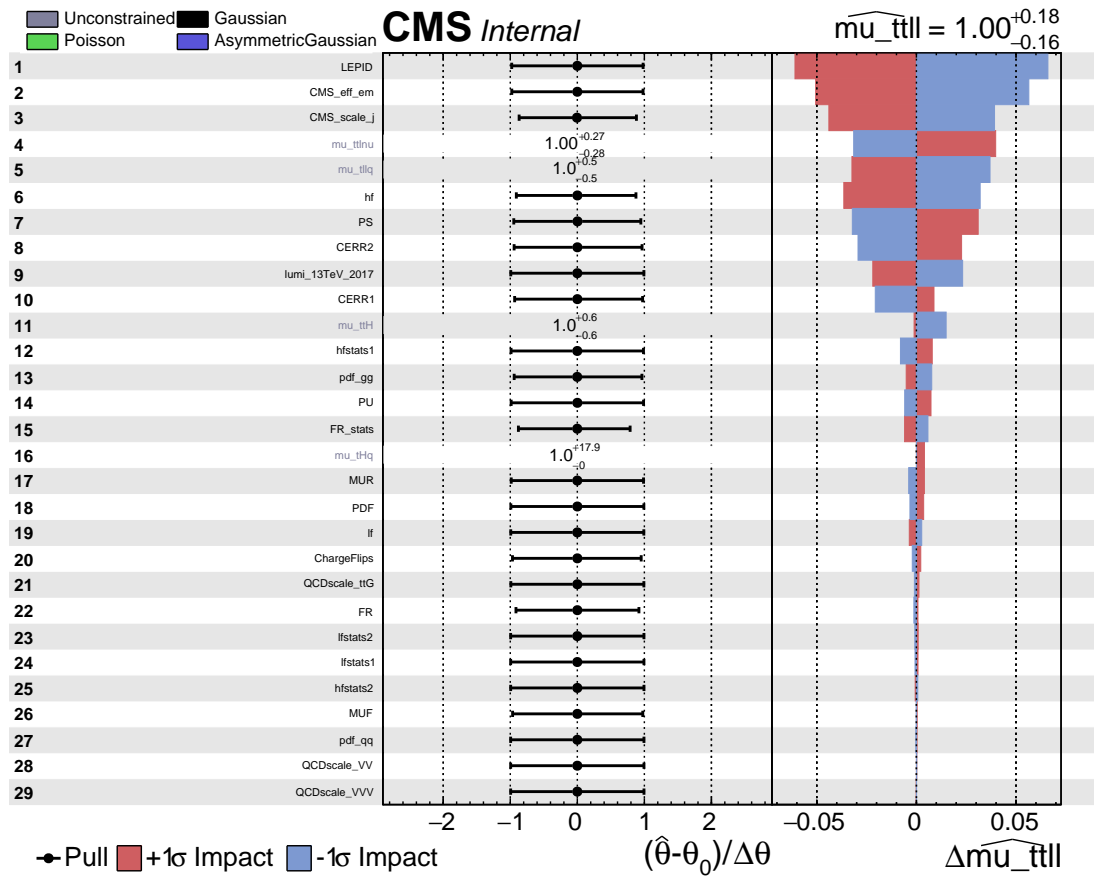


Figure 9: Impact plot for $\widehat{\mu}_{\text{ttll}}$. The other process signal strengths are allowed to float freely as unconstrained nuisances.

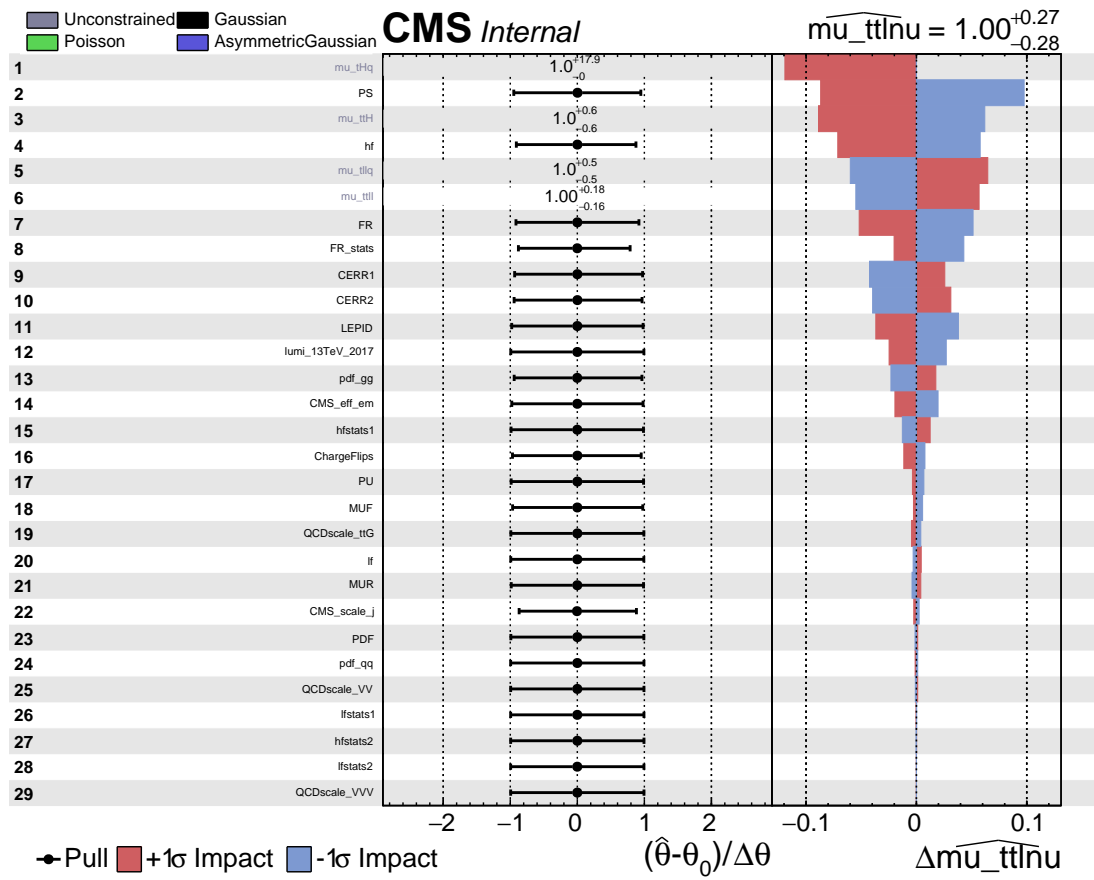


Figure 10: Impact plot for $\text{t}\bar{\text{t}}\text{l}\nu$. The other process signal strengths are allowed to float freely as unconstrained nuisances.

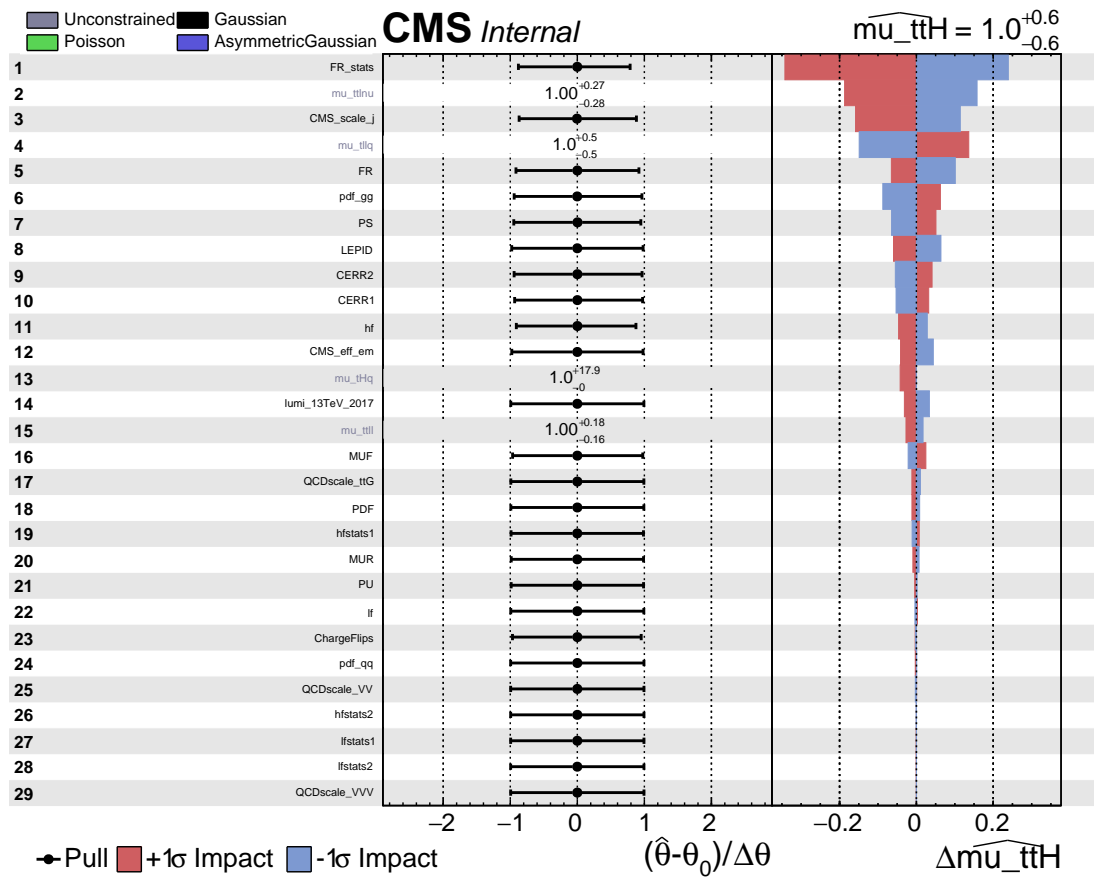


Figure 11: Impact plot ttH . The other process signal strengths are allowed to float freely as unconstrained nuisances.

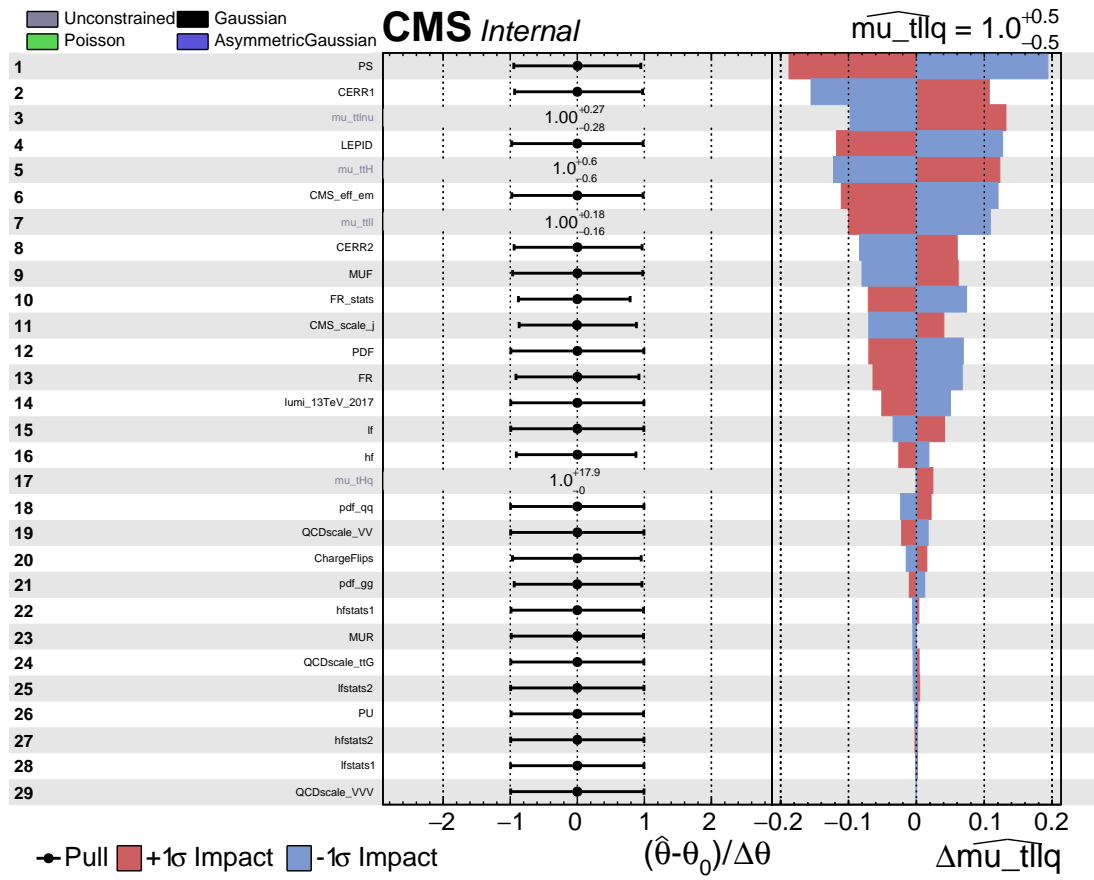


Figure 12: Impact plot for $t\bar{t}lq$. The other process signal strengths are allowed to float freely as unconstrained nuisances.

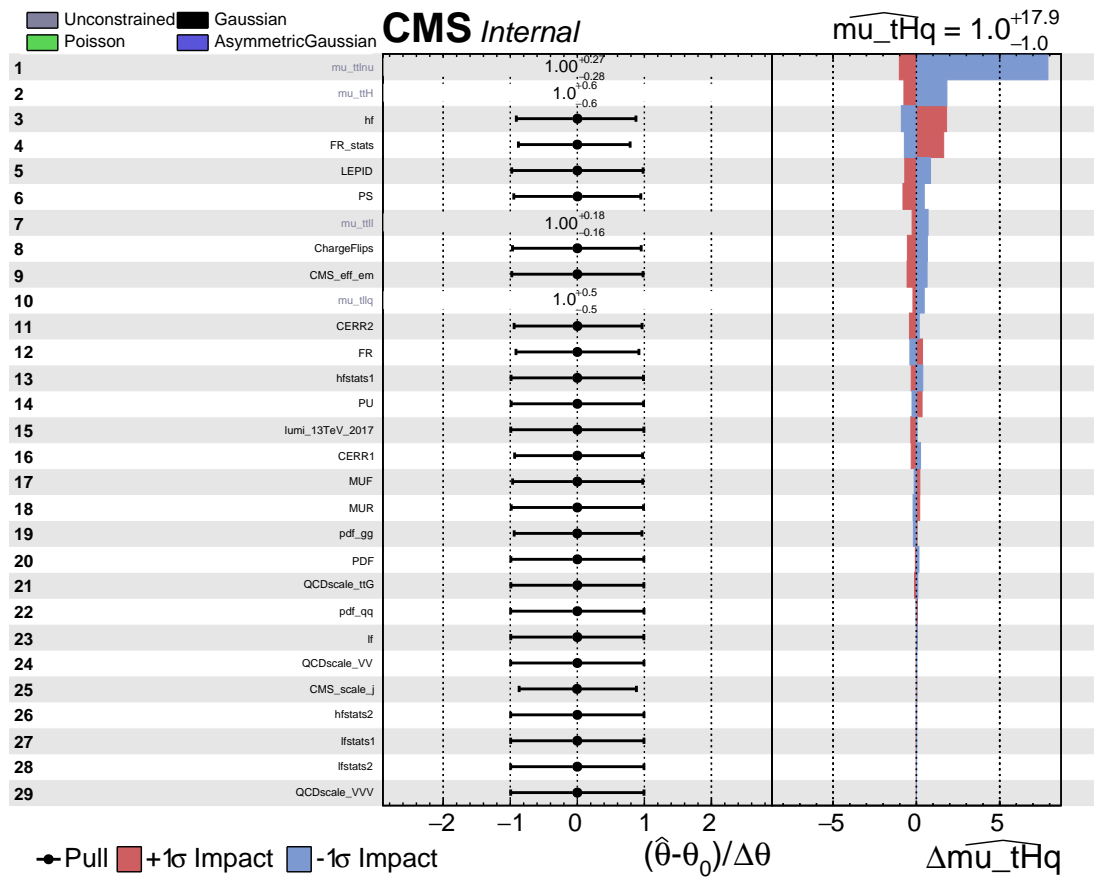


Figure 13: Impact plot for tHq. The other process signal strengths are allowed to float freely as unconstrained nuisances.

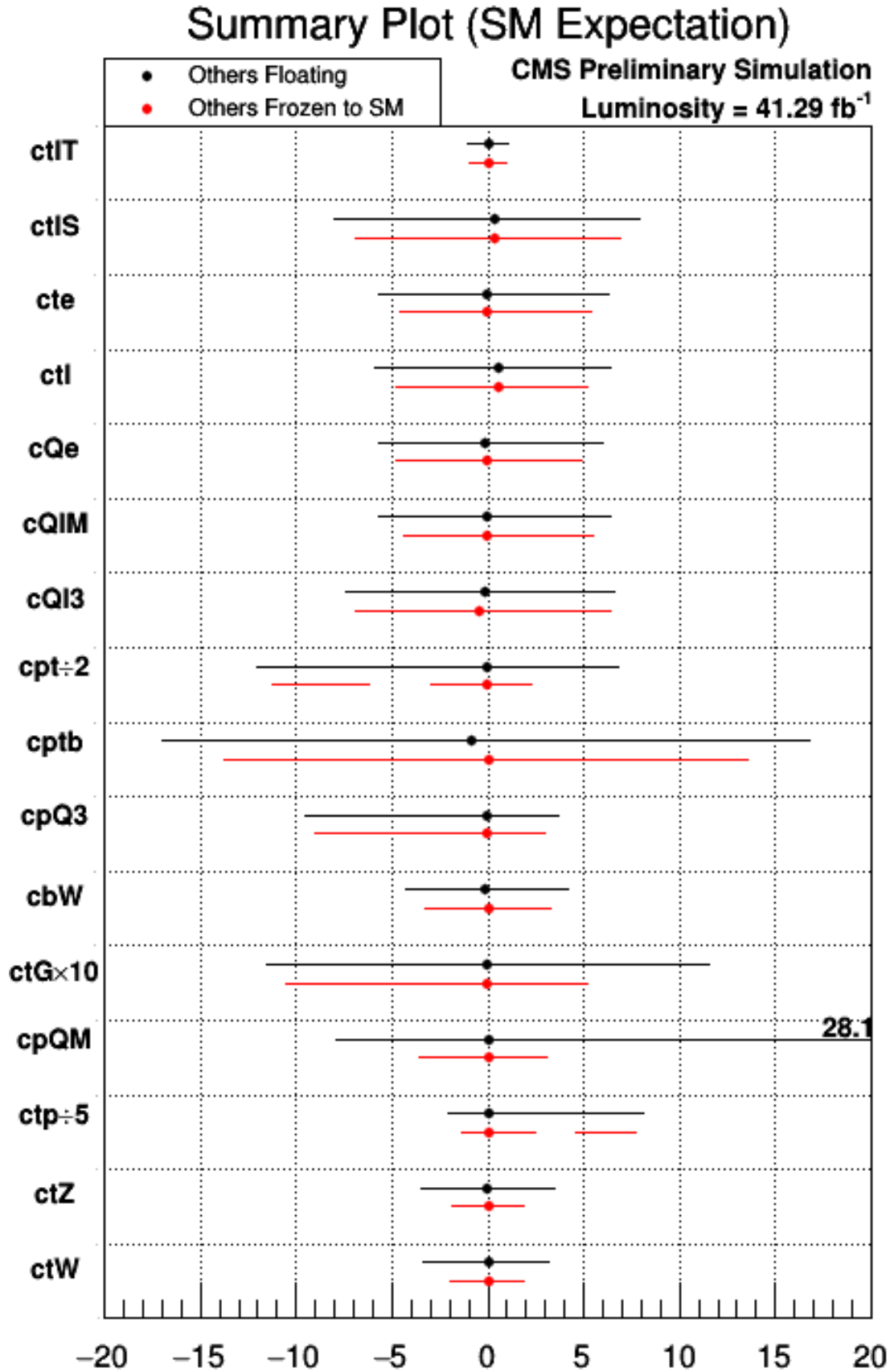
Summary Table (other WC floating)		
Wilson Coefficient	Best Fit	2σ Interval
ctW	0.06	[-3.34, 3.19]
ctZ	-0.04	[-3.53, 3.50]
ctp	0.00	[-10.40, 40.40]
cpQM	0.00	[-7.92, 28.08]
ctG	0.01	[-1.15, 1.15]
cbW	-0.18	[-4.24, 4.18]
cpQ3	0.07	[-9.50, 3.68]
cptb	-0.88	[-16.96, 16.80]
cpt	-0.03	[-24.04, 13.61]
cQl3i	0.20	[-7.44, 6.64]
cQlMi	-0.03	[-5.71, 6.36]
cQei	-0.16	[-5.70, 5.98]
ctli	0.60	[-5.92, 6.40]
ctei	-0.04	[-5.67, 6.32]
ctlSi	0.31	[-8.01, 7.92]
ctlTi	0.04	[-1.09, 1.07]

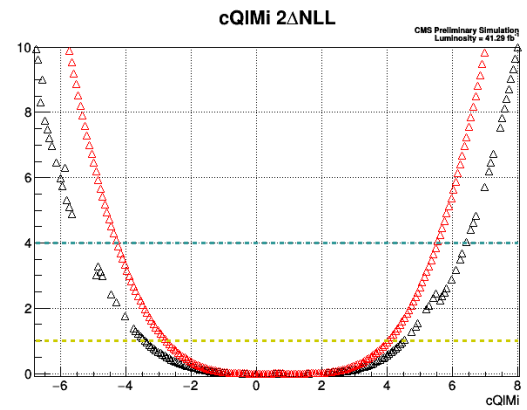
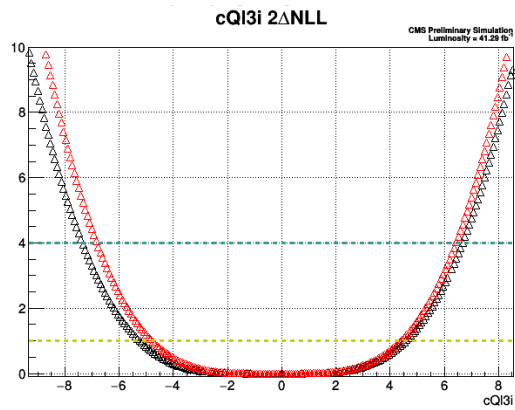
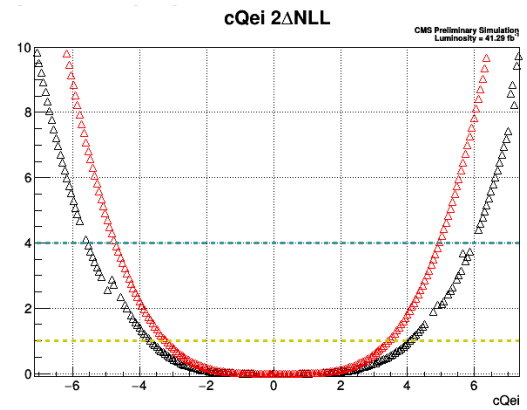
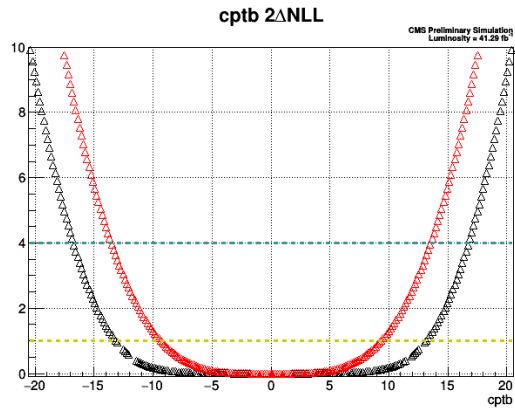
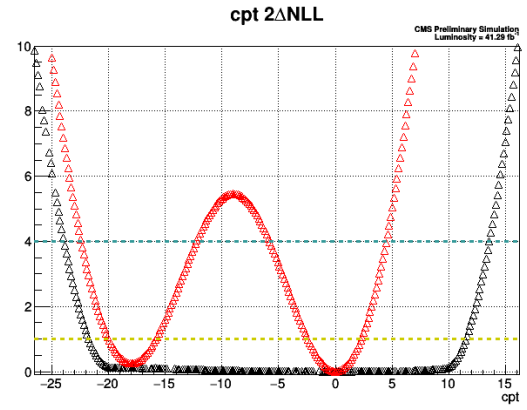
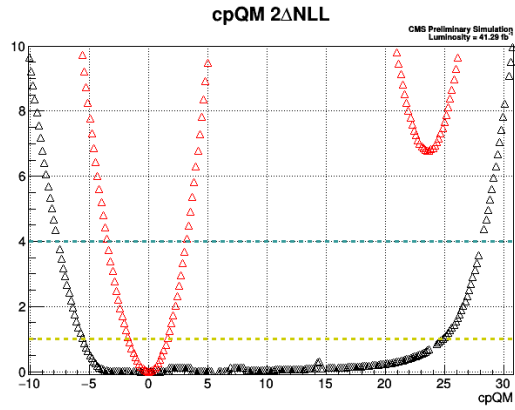
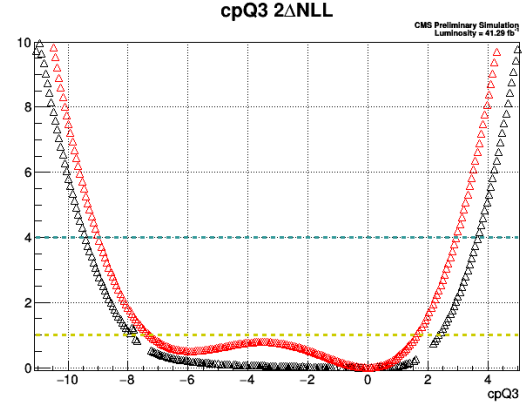
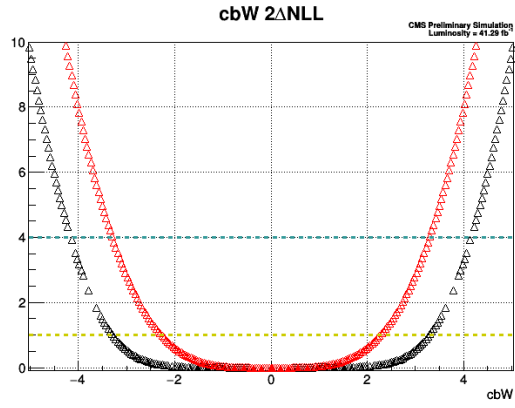
Table 18: Standard Model-expected best fit values and 2σ uncertainty intervals. Taken from 1D scans with other Wilson Coefficients profiled.

Summary Table (other WC frozen)		
Wilson Coefficient	Best Fit	2σ Interval
ctW	0.00	[-1.94, 1.82]
ctZ	0.01	[-1.88, 1.88]
ctp	0.01	[-6.60, 12.20] and [22.80, 38.60]
cpQM	0.00	[-3.60, 3.12]
ctG	0.00	[-1.06, 0.52]
cbW	0.00	[-3.32, 3.28]
cpQ3	0.00	[-9.06, 2.98]
cptb	0.08	[-13.76, 13.60]
cpt	0.00	[-22.5, -12.42] and [-5.98, 4.52]
cQl3i	-0.44	[-6.88, 6.40]
cQlMi	-0.03	[-4.35, 5.51]
cQei	-0.10	[-4.80, 4.93]
ctli	0.52	[-4.80, 5.20]
ctei	-0.05	[-4.56, 5.44]
ctlSi	0.31	[-6.95, 6.86]
ctlTi	0.04	[-0.94, 0.93]

Table 19: Standard Model-expected best fit values and 2σ uncertainty intervals. Taken from 1D scans with other Wilson Coefficients frozen to their Standard Model value of 0.

940 The number of events selected in different categories is compared to the postfit expected contri-
941 butions of the signal processes and of the different background processes in Figures 17 and 18.
942 Figures 19 and 20 show an example of the expected yields after setting a single WC (ctlTi) to its
943 $+1\sigma$ (red) and $+2\sigma$ (green) values. The process yields themselves are set to their SM expectation
944 for reference.

Figure 14: Standard Model-expected best fit values and 2σ uncertainty intervals.



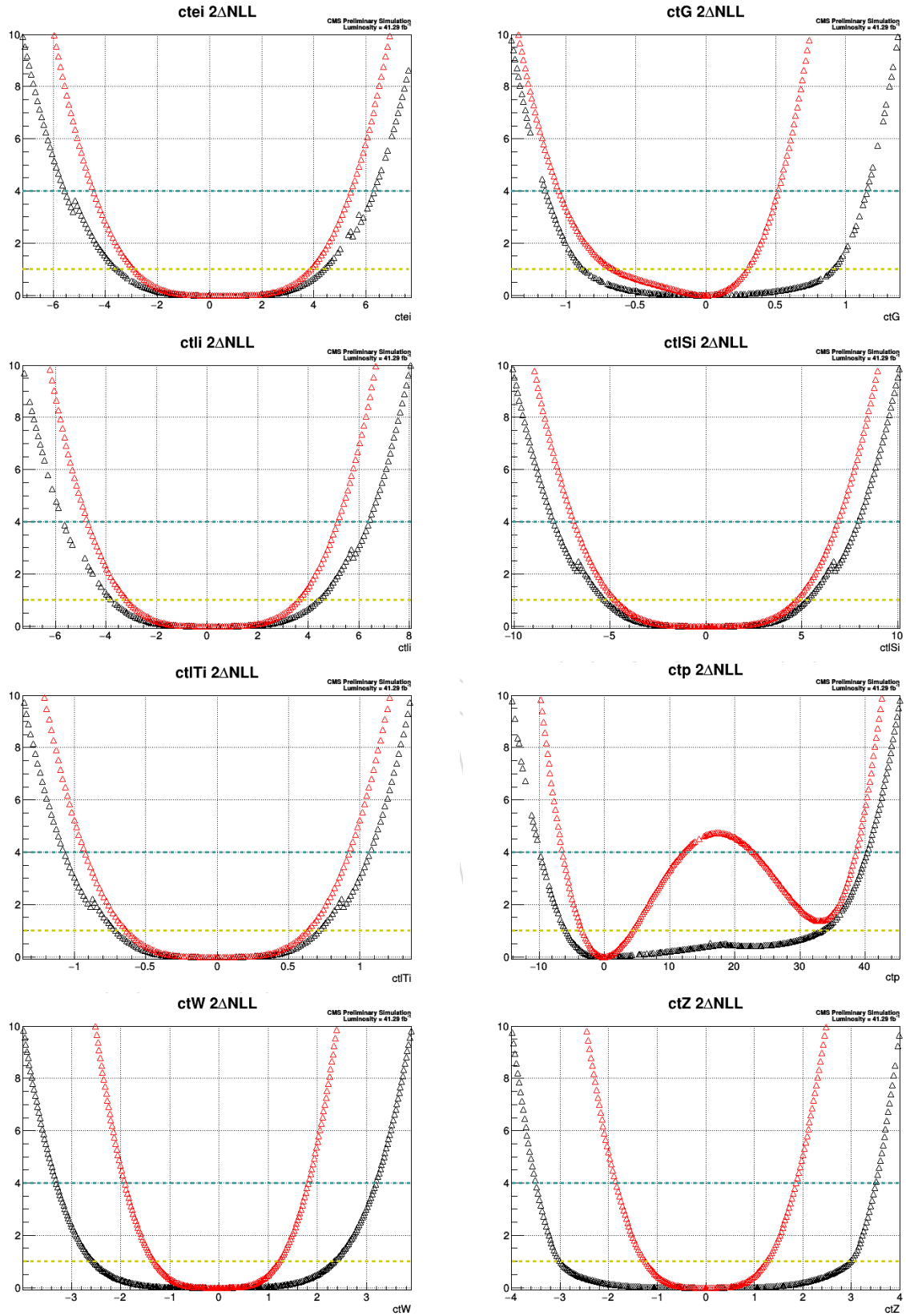
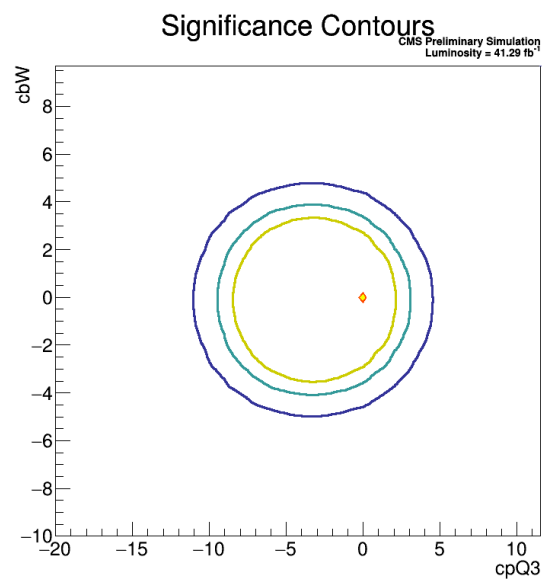
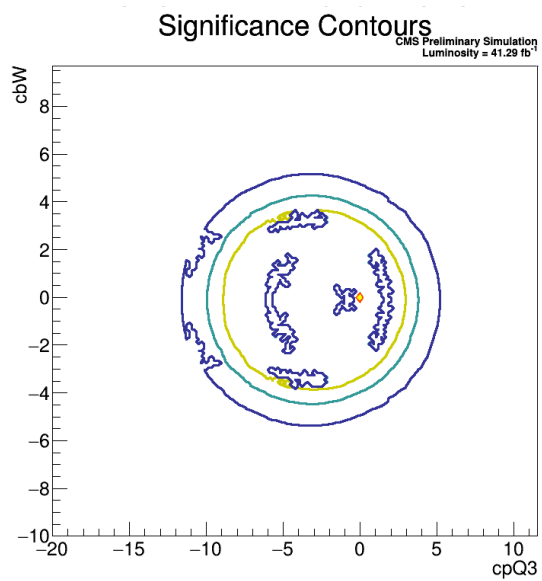
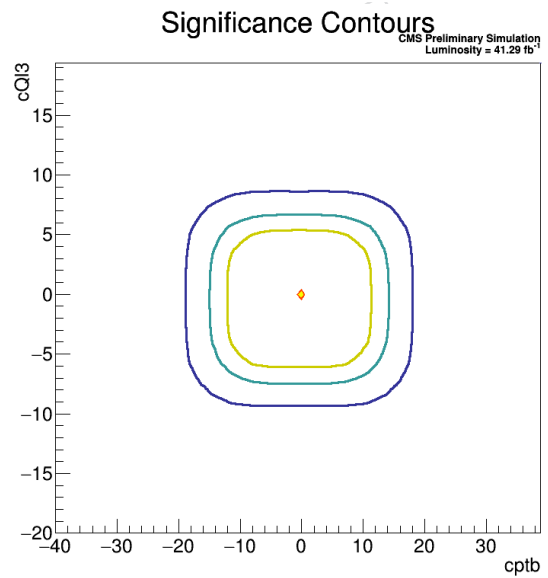
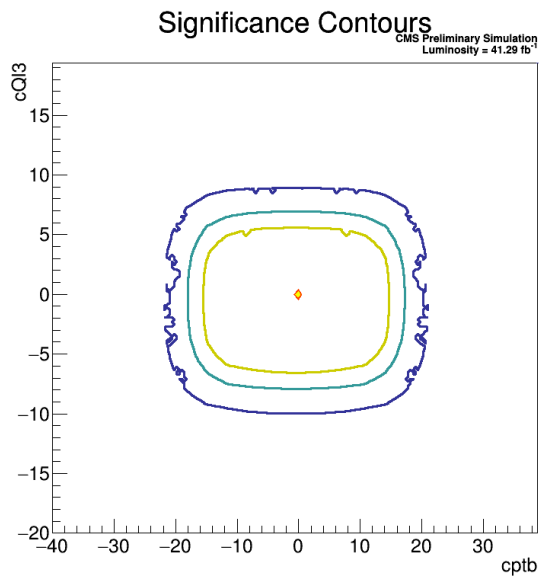
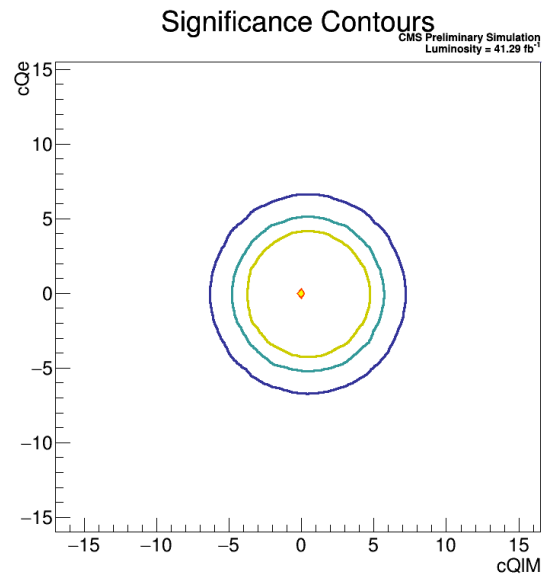
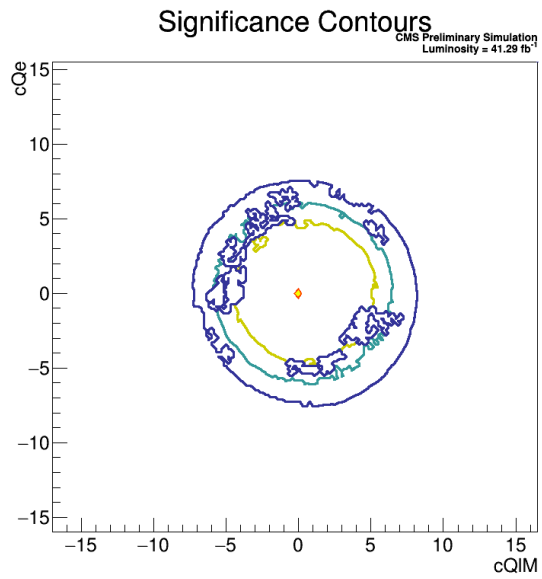
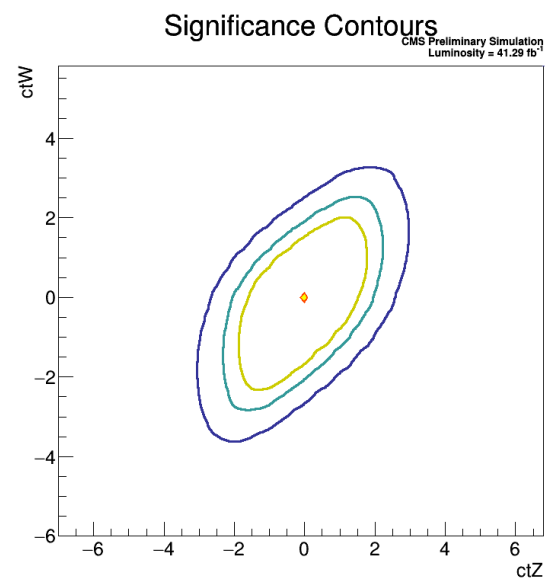
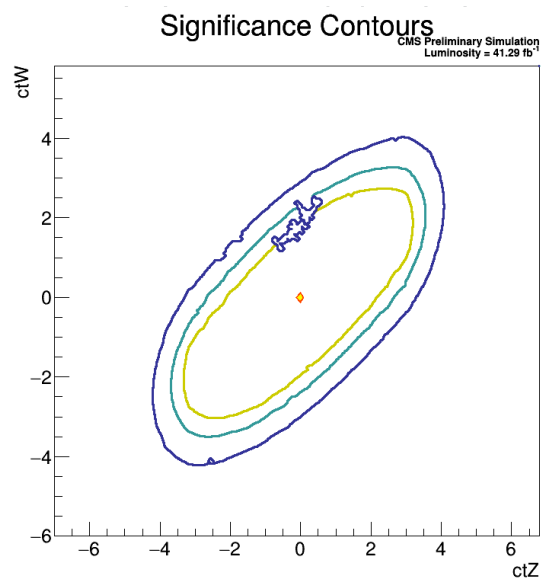
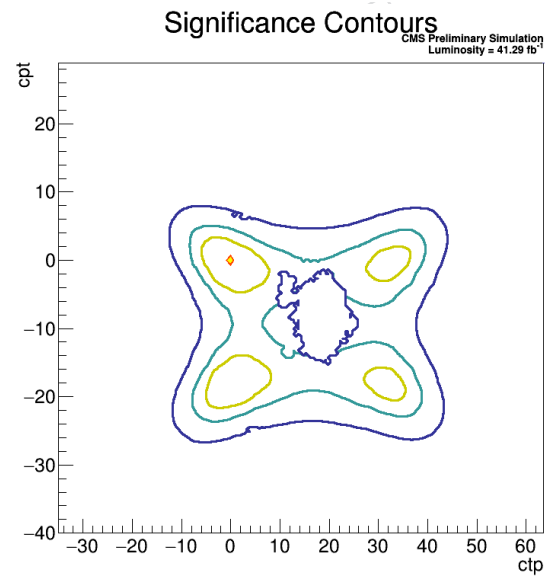
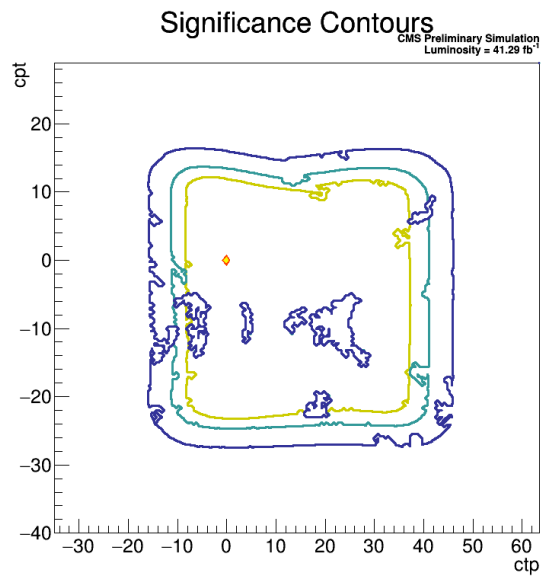
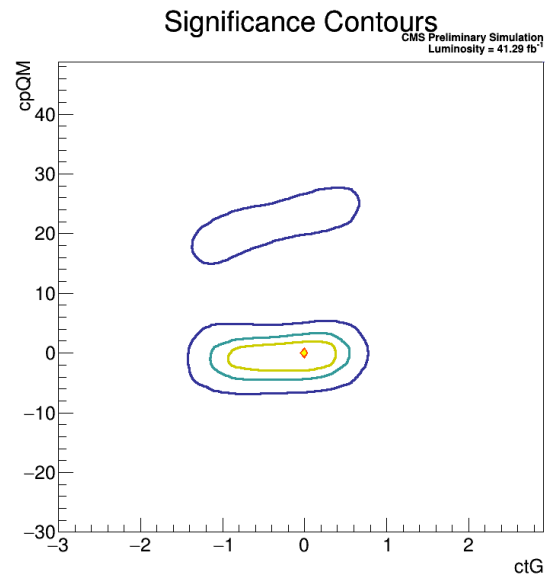
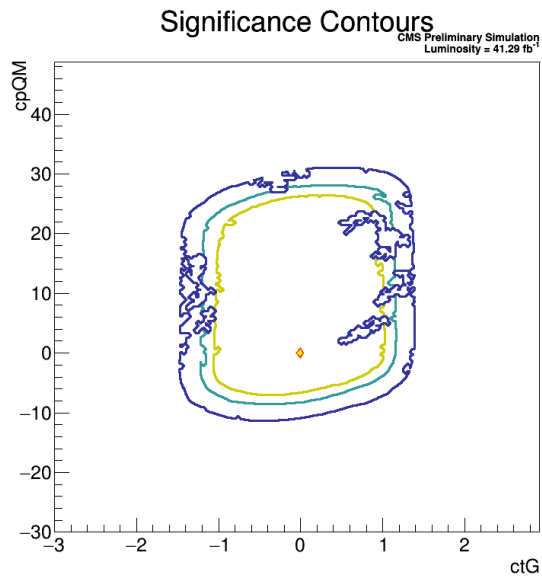


Figure 15: Scans over values of single Wilson Coefficients. Standard Model-expected results. Black points are profiled scans while the other 15 Wilson Coefficients are allowed to float. Red points are scans while the other 15 Wilson Coefficients are frozen to their Standard Model values.





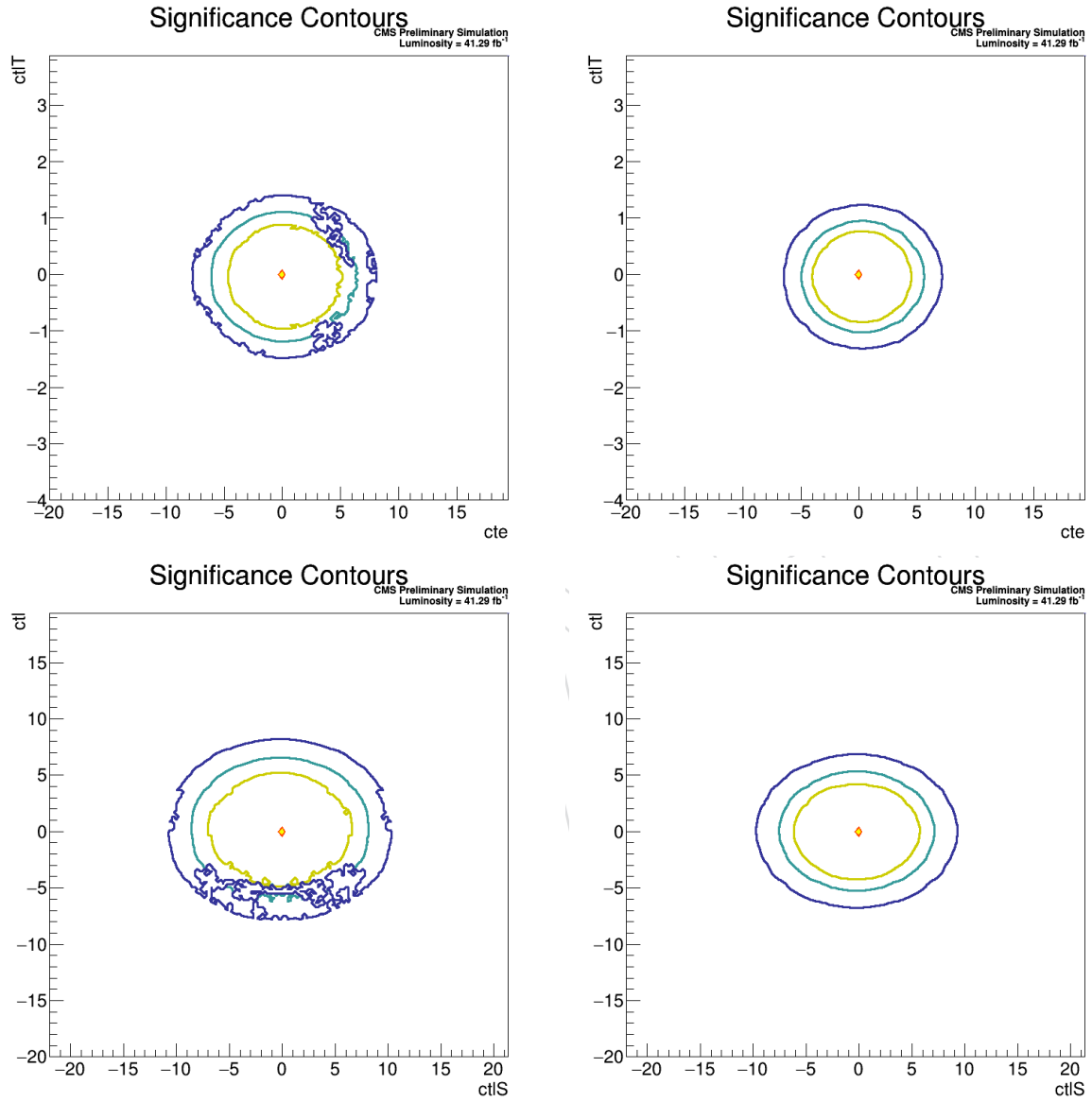
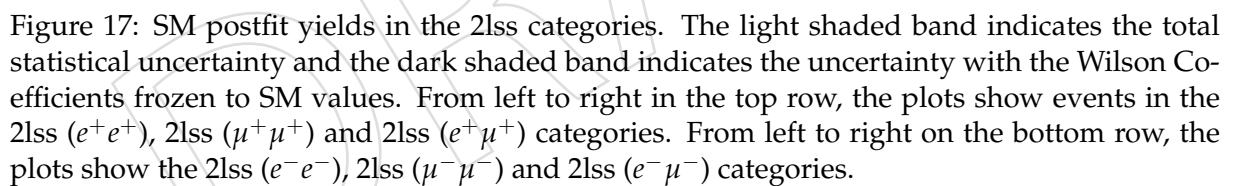


Figure 16: Left column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients profiled. Right column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients frozen to their Standard Model value. Diamond markers are shown for the Standard Model prediction.



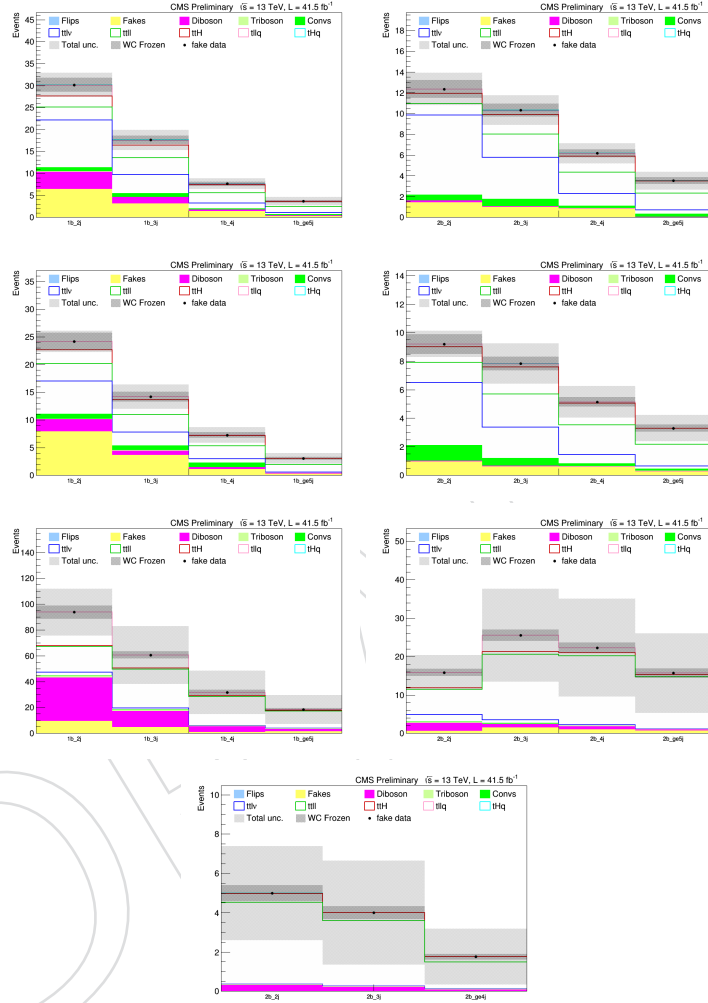


Figure 18: SM postfit yields in the 3l and 4l categories. The light shaded band indicates the total statistical uncertainty and the dark shaded band indicates the uncertainty with the Wilson Coefficients frozen to SM values. Going from left to right and top to bottom on a per-row basis, the plots show events in the 3l (1b “p”), 3l ($\geq 2b$ “p”), 3l (1b “m”), 3l ($\geq 2b$ “m”), 3l (SFOS Z, 1b), 3l (SFOS Z, $\geq 2b$) and $\geq 4l$ categories, respectively.

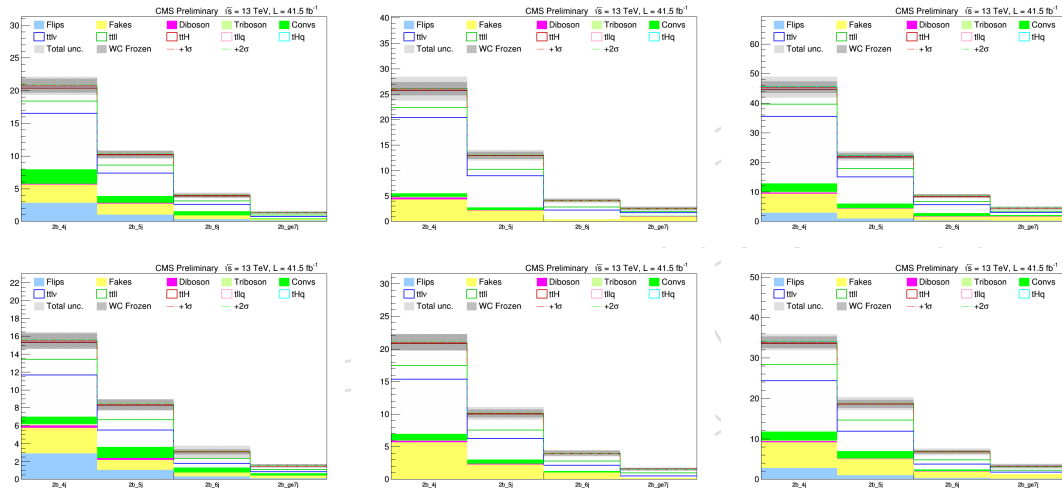


Figure 19: From left to right in the top row, the plots show events in the 2lss (e^+e^+), 2lss ($\mu^+\mu^+$) and 2lss ($e^+\mu^+$) categories. From left to right on the bottom row, the plots show the 2lss (e^-e^-), 2lss ($\mu^-\mu^-$) and 2lss ($e^-\mu^-$) categories.

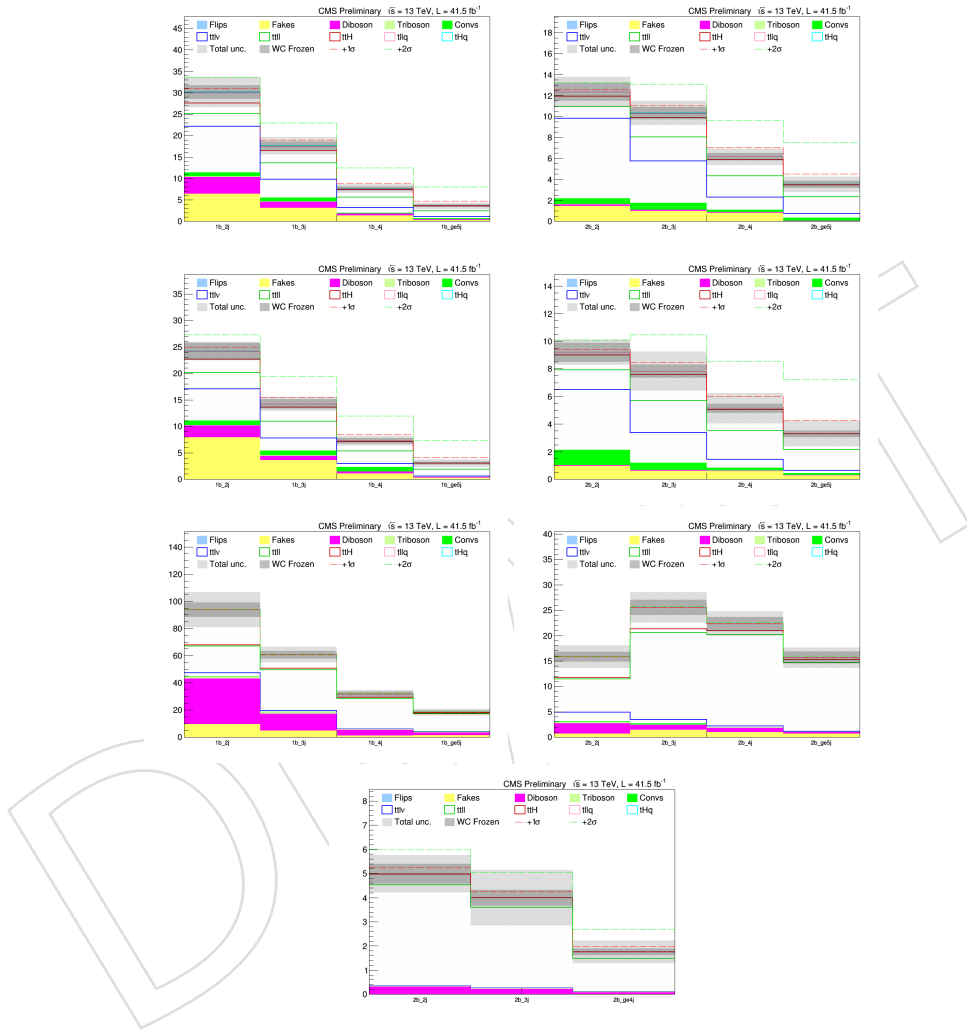


Figure 20: Going from left to right and top to bottom on a per-row basis, the plots show events in the 3l (1b “p”), 3l ($\geq 2b$ “p”), 3l (1b “m”), 3l ($\geq 2b$ “m”), 3l (SFOS Z, 1b), 3l (SFOS Z, $\geq 2b$) and $\geq 4l$ categories, respectively.

A Synchronization Plots

In order to maximize the acceptance of prompt leptons, while simultaneously maximizing the rejection of background leptons in our signal regions, we capitalize off of a multivariate lepton identification technique developed for the ttH multilepton analysis. Not only does this help us optimize sensitivity to our signal processes, it allows us to use the same technique as is used in the ttH analysis for estimating the contamination from non-prompt leptons, together with the method for estimating the uncertainties (both on the identification of prompt leptons, as well as on the fake rate estimation).

To ensure that the estimation of the fake background as performed in the ttH analysis is also applicable to our analysis, we employ the exact same set of lepton identification criteria, and remain synchronized with the ttH analysis on the object-ID level. We compared the yields of muons, electrons and jets selected by running our analysis framework over 56465 ttH MC events with those obtained by the ttH analysis when running their framework over the same events, and found that the number of objects in each collection agreed to within less than 1%. We also compared kinematic and other distributions of the objects passing the loose selection criteria; those comparisons are shown in Figures 21 and 22.

DRAFT

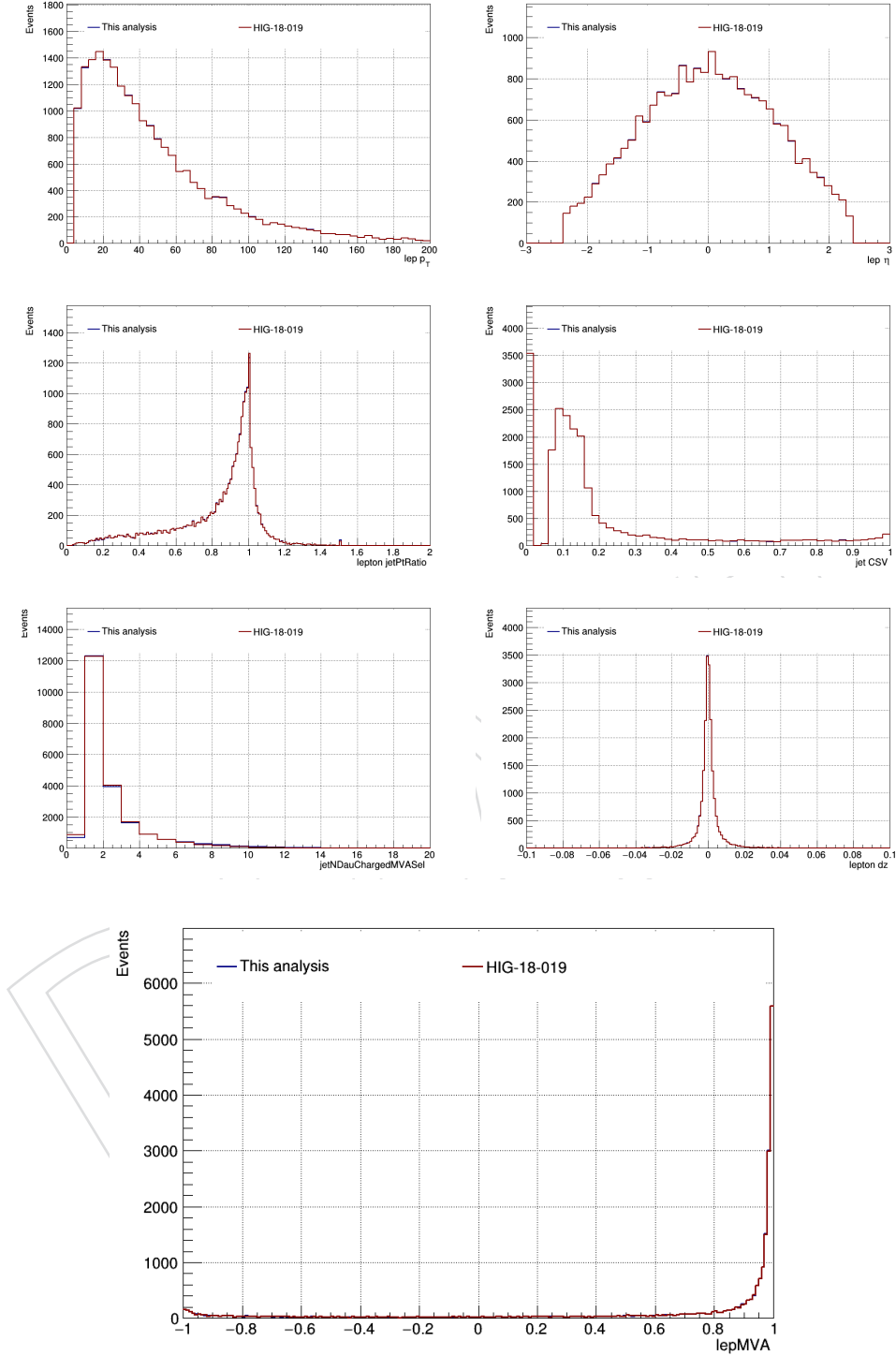


Figure 21: Comparison between distributions used to identify muons obtained from running two different analysis frameworks over the same set of $t\bar{t}H$ MC events. The plots are filled after applying the loose selection cuts listed in table 10, as well as the cross-cleaning between objects described in Section 4. Very good agreement is seen for the BDT inputs (top three rows), as well as the lepton BDT output distribution (bottom row).

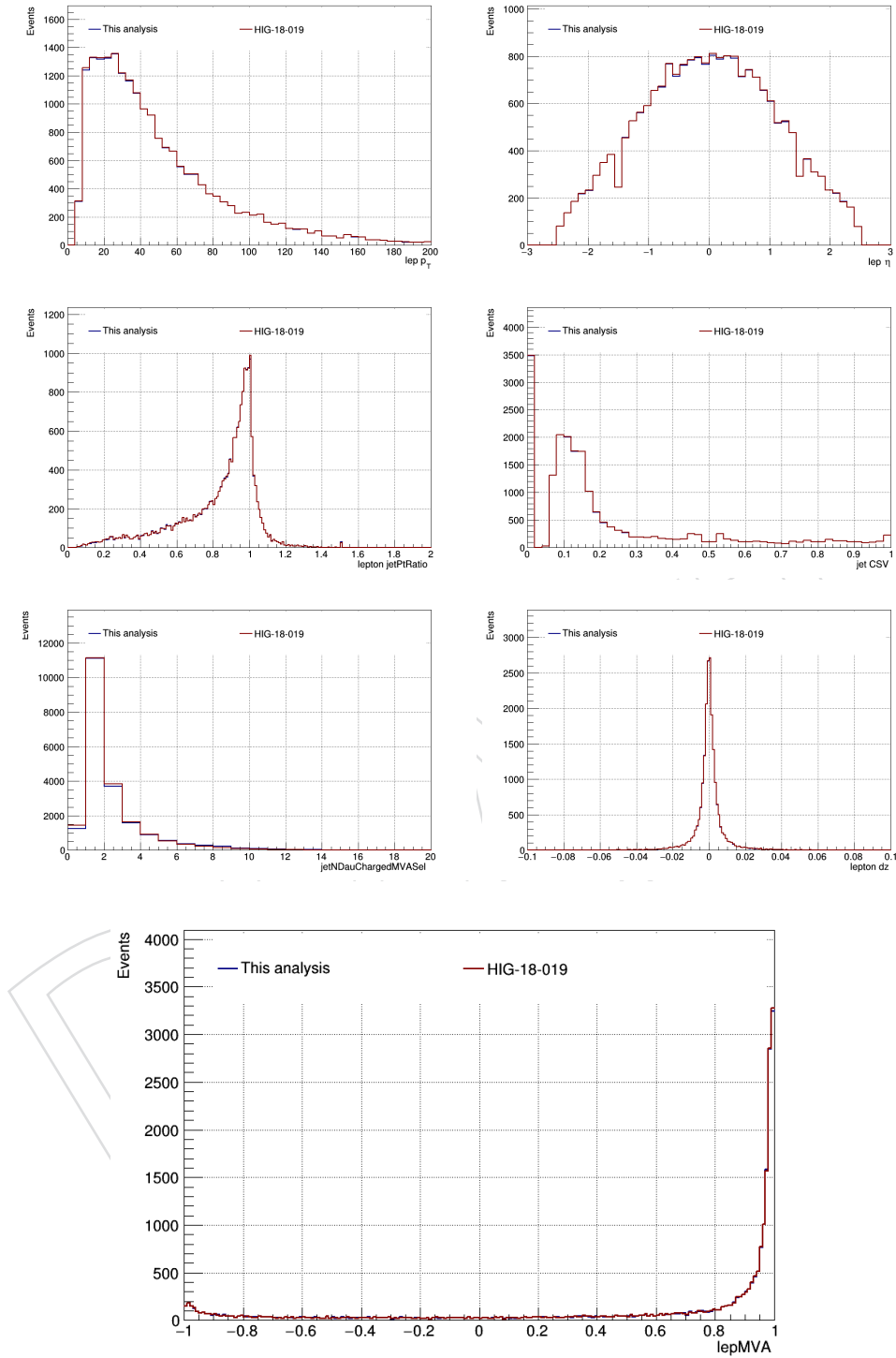


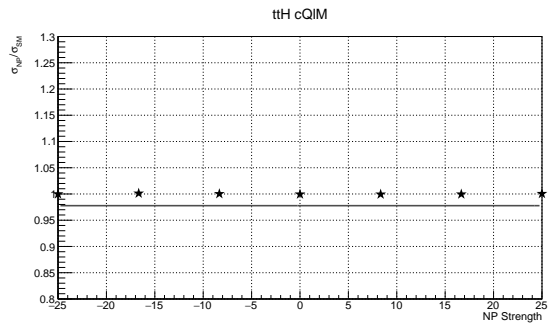
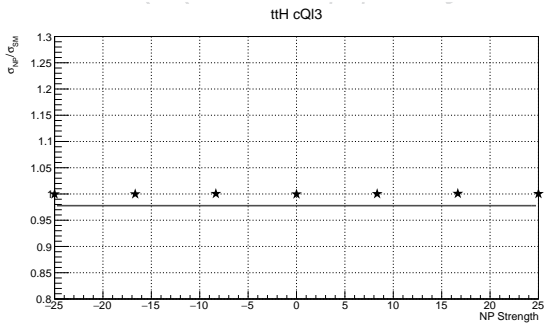
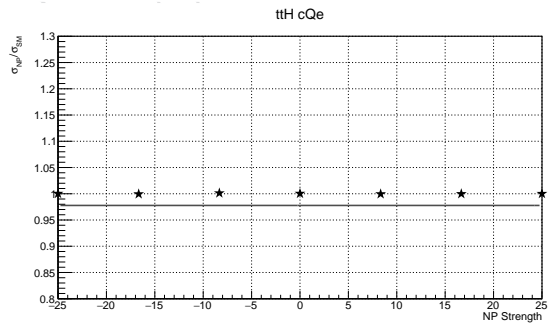
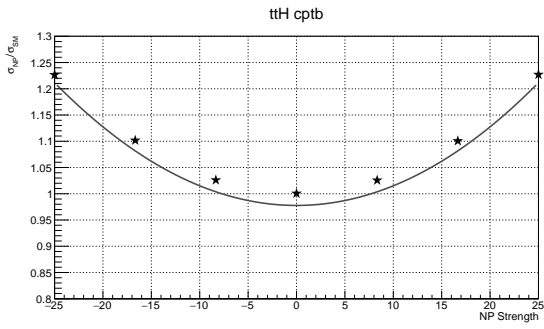
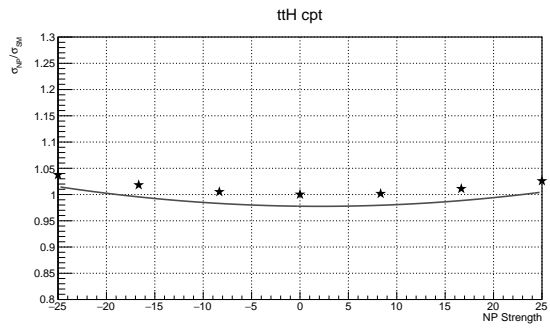
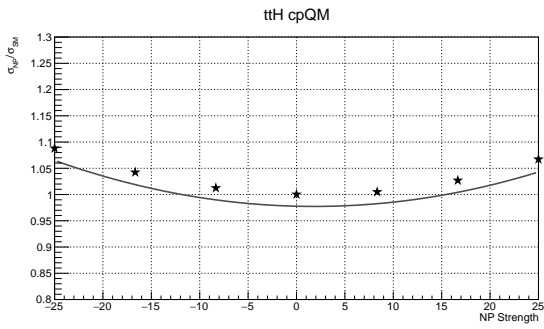
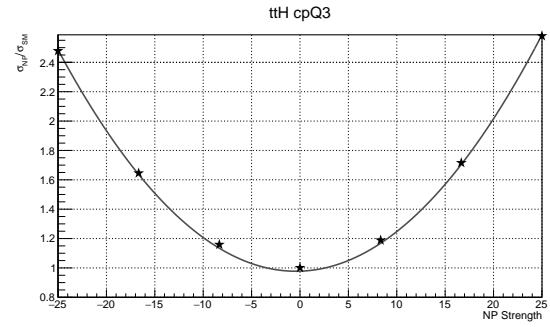
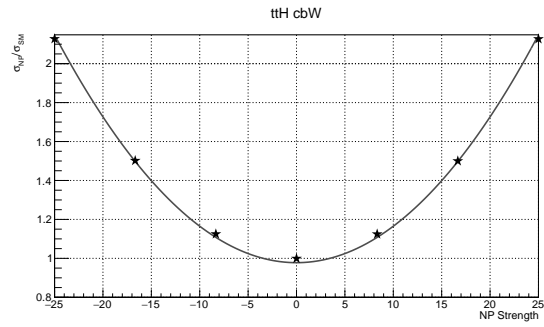
Figure 22: Comparison between distributions used to identify electrons obtained from running two different analysis frameworks over the same set of $t\bar{t}H$ MC events. The plots are filled after applying the loose selection cuts listed in table 9, as well as the cross-cleaning between objects described in Section 4. Very good agreement is seen for the BDT inputs (top three rows), as well as the lepton BDT output distribution (bottom row).

B MC Validation

The MC signal samples used in this analysis make use of the MADGRAPH event-by-event reweighting feature. This involves starting with a set of hard events, which are generated for some specific choice of WC strengths. The sample is then rescaled on an event-by-event basis to represent a sample that was generated with some different choice of WC strengths. A key aspect of the event reweighting, is the fact that if the set of initial hard events are generated in a region of phase space that does not overlap with the region to which the events are reweighted to, the sample could contain events with very large weights. This results in a sample that is dominated almost entirely by the behavior of only a few events.

A set of reference samples are used to check the accuracy of the reweighting used in the nominal sample. These reference samples are produced for a particular choice of WC strengths without any reweighting done. The reweighted events from the nominal sample are then compared to each reference sample in order to validate the nominal sample reweighting. Figures 23,24,25, and 26 show the comparison between the nominal sample reweighting and the dedicated reference samples.

As a final step of the validation, we check the event yields obtained using the nominal reweighted samples by comparing them to event yields obtained from centrally-produced, NLO samples that do not use reweighting. The comparison is done in the n_{jets} bins of the signal regions, after the full object ID and event selection and categorization of the analysis. Plots of these distributions are shown in Figure 27. Reasonably good agreement is seen between the privately-produced reweighted samples and the centrally-produced, non-reweighted samples.



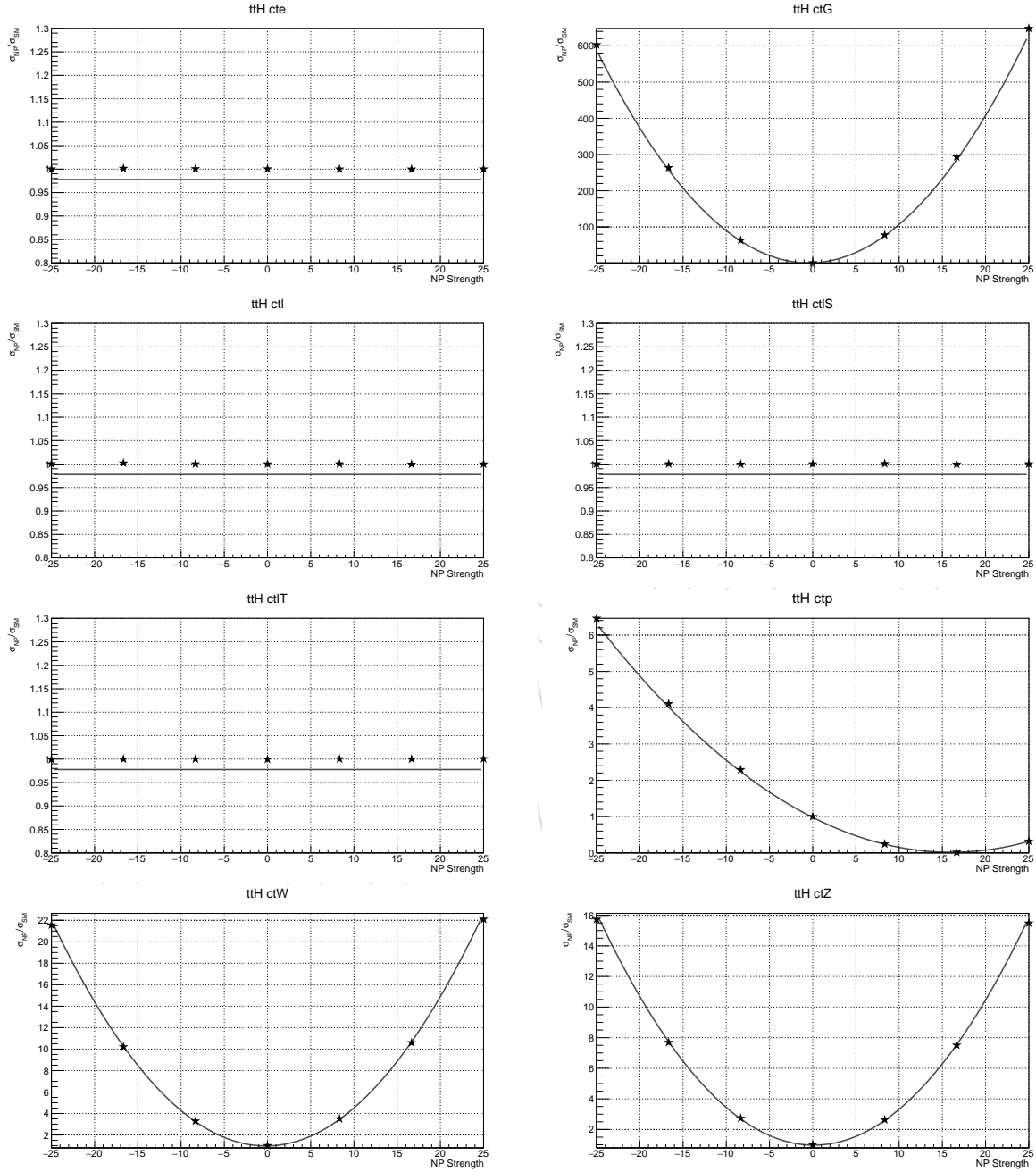
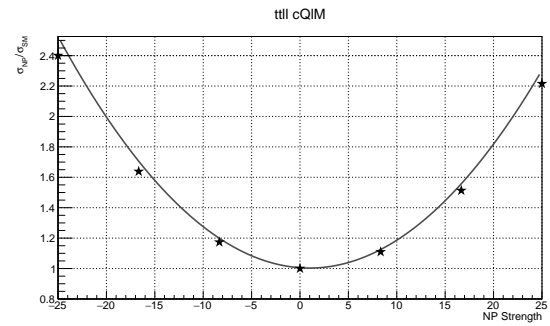
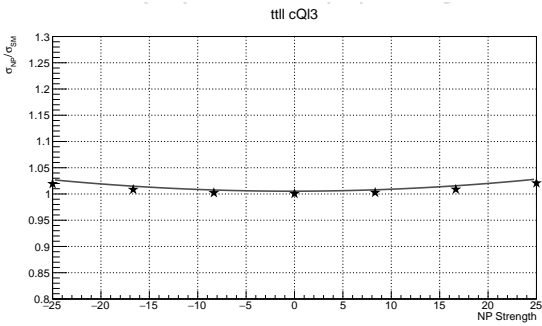
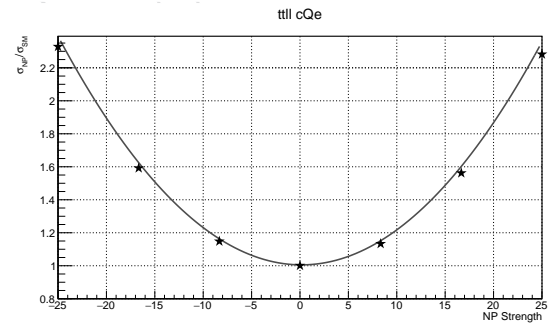
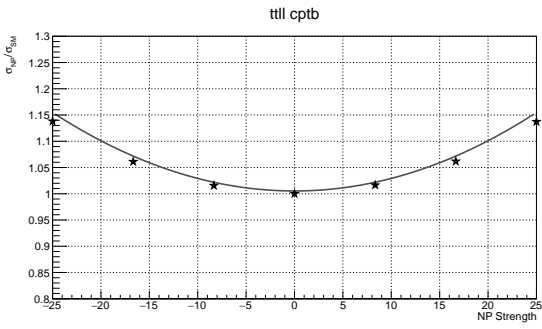
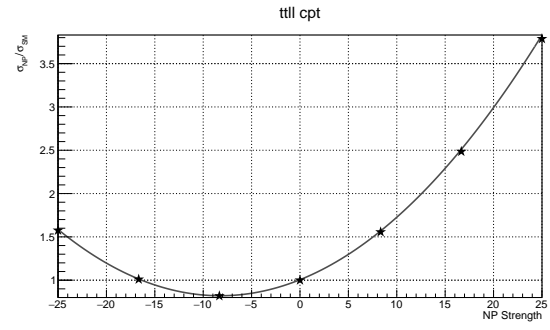
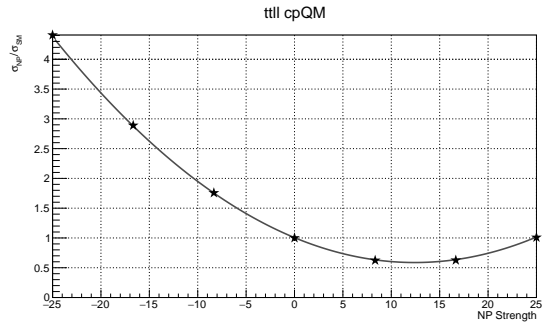
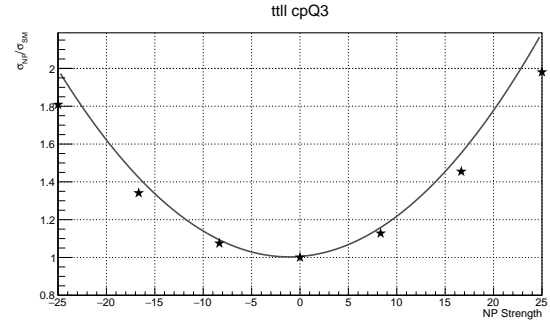
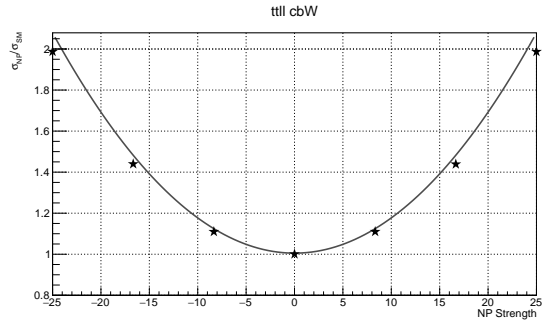


Figure 23: Comparison of the WC parameterization fit for ttH. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



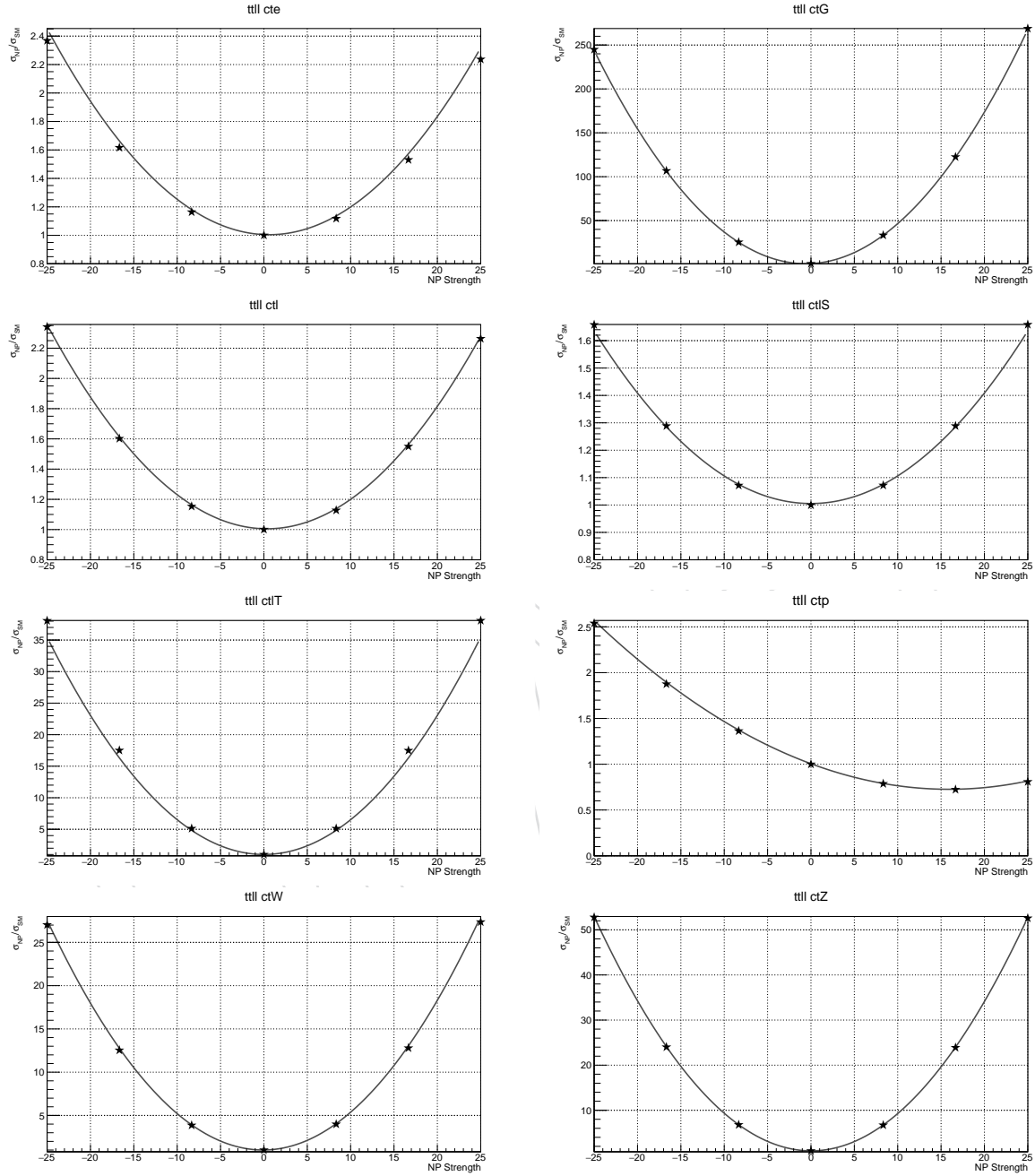
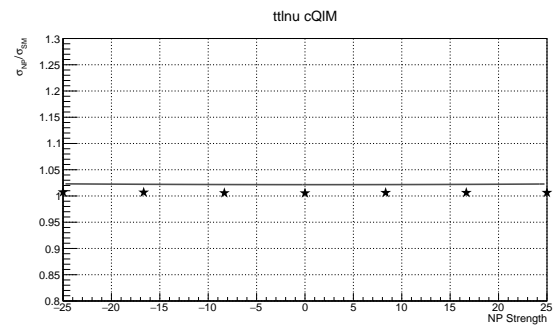
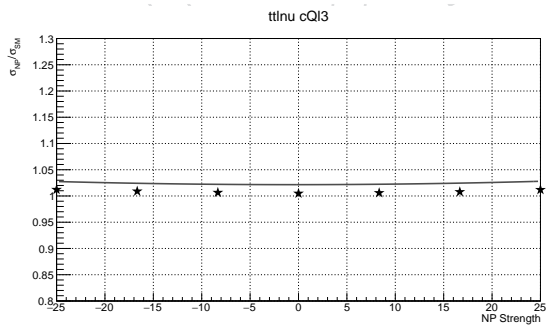
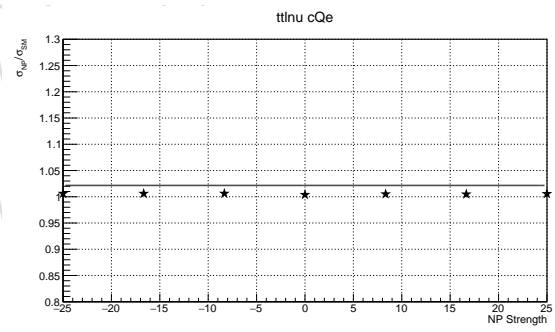
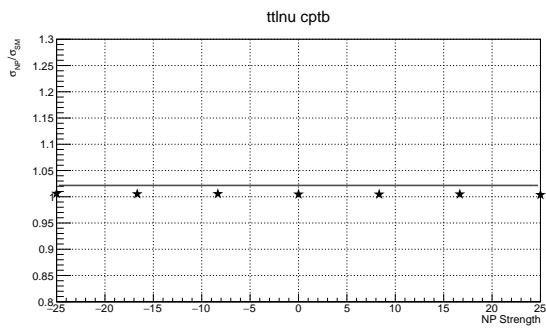
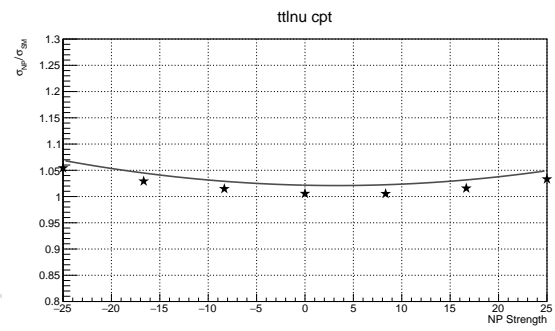
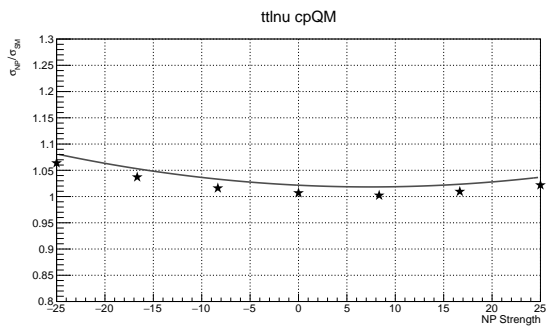
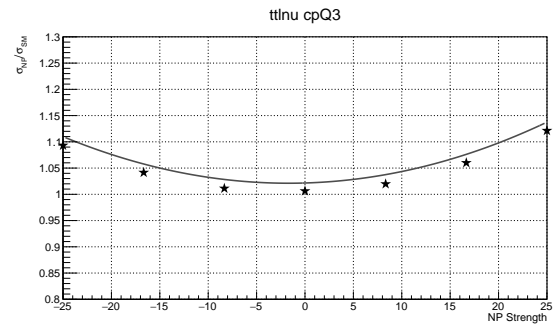
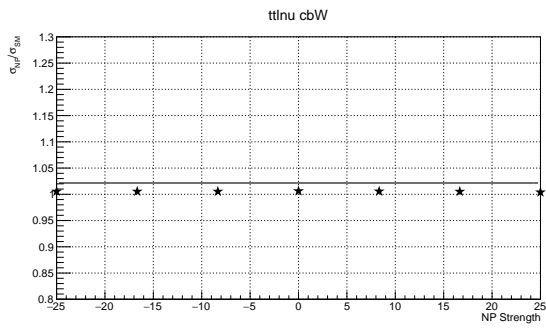


Figure 24: Comparison of the WC parameterization fit for ttll. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



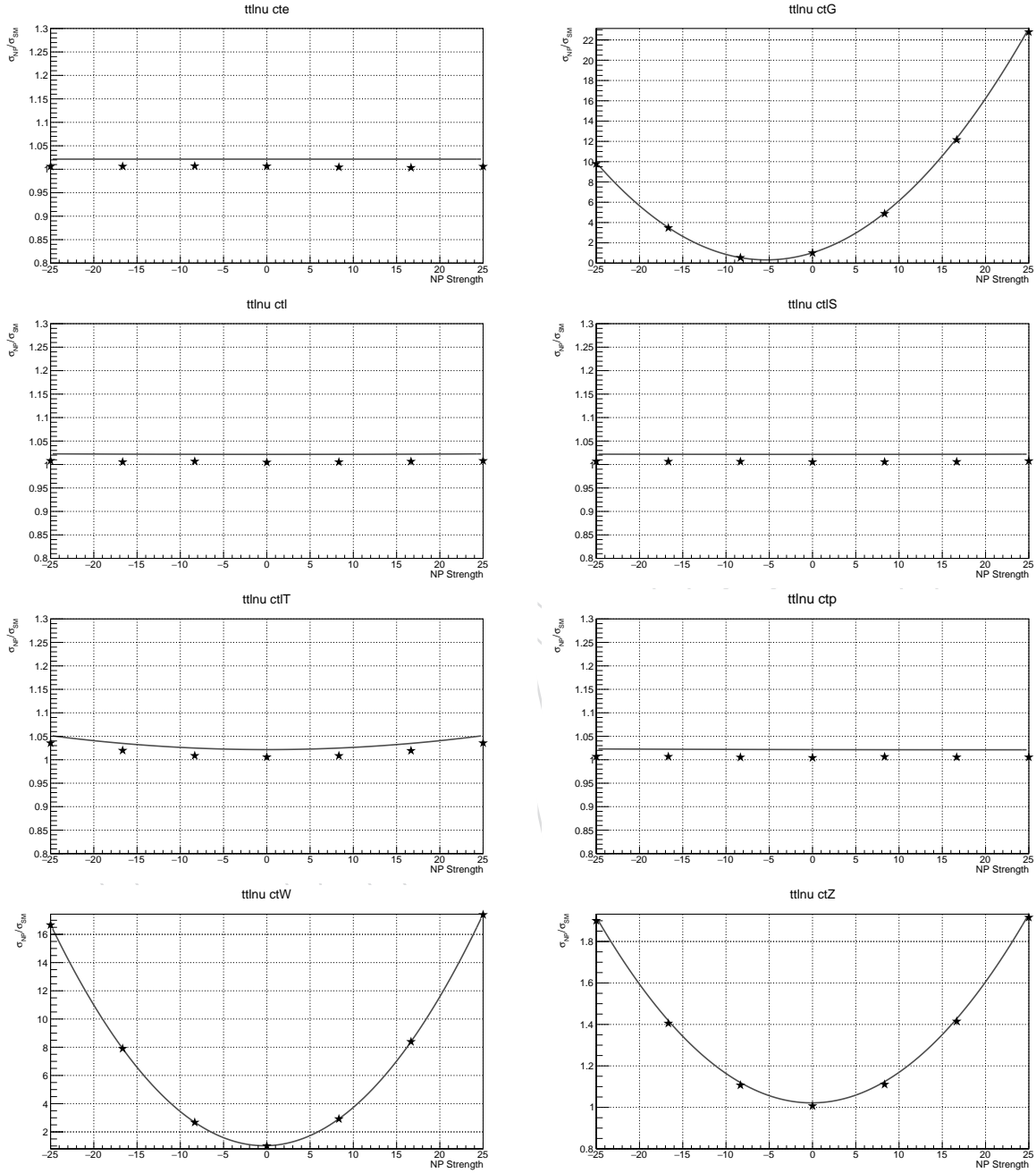
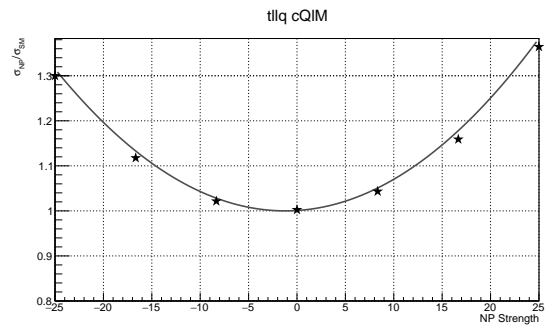
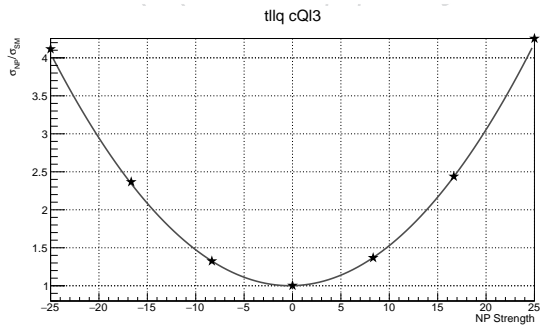
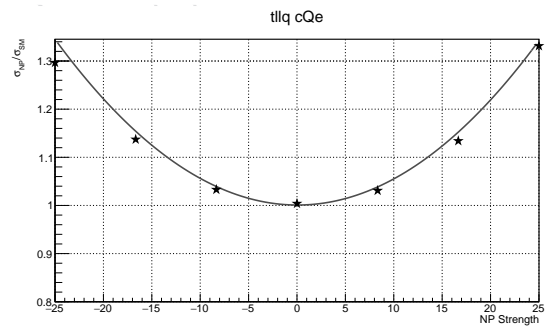
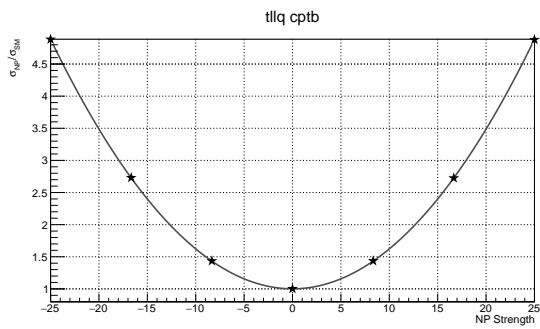
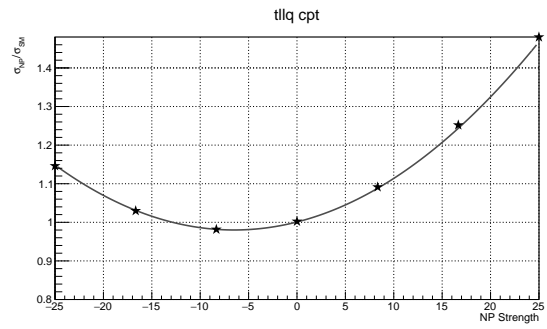
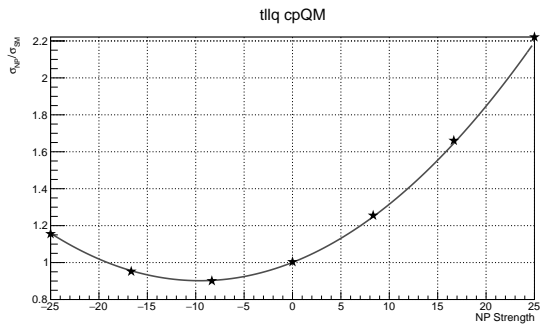
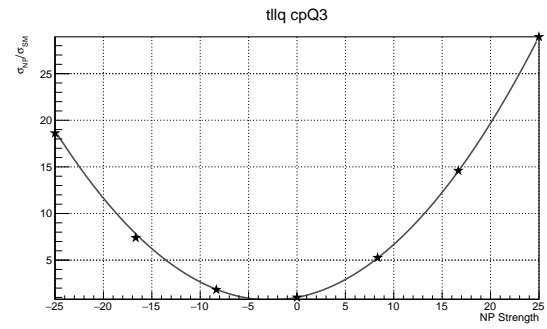
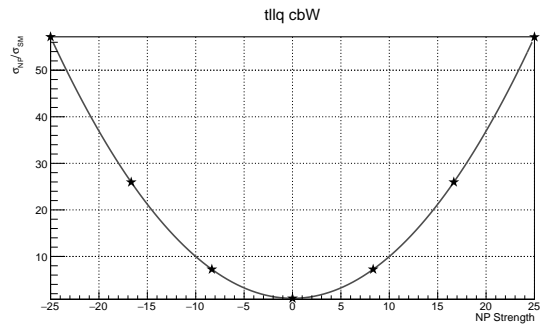


Figure 25: Comparison of the WC parameterization fit for $tt\bar{\nu}$. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



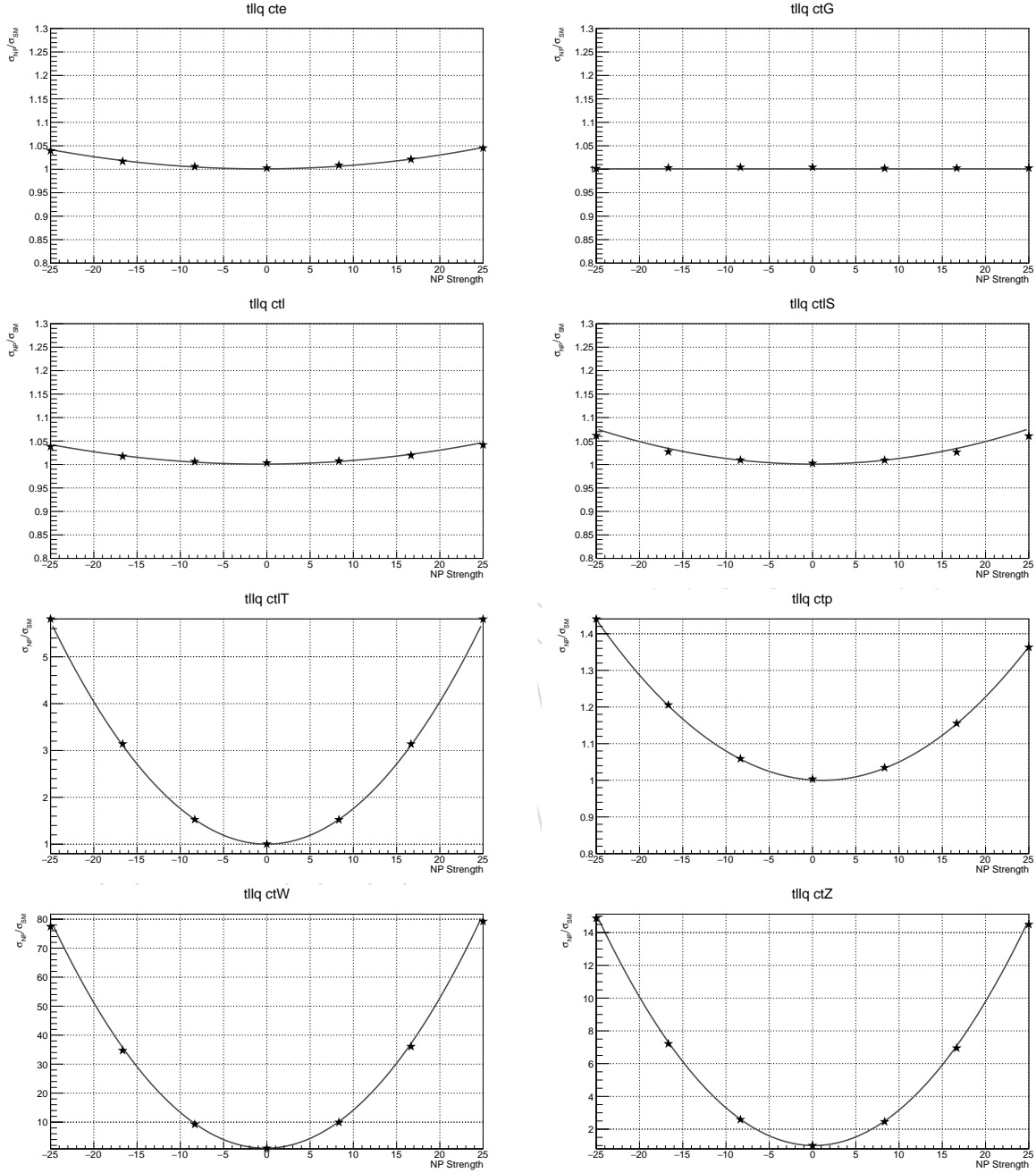


Figure 26: Comparison of the WC parameterization fit for $tllq$. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.

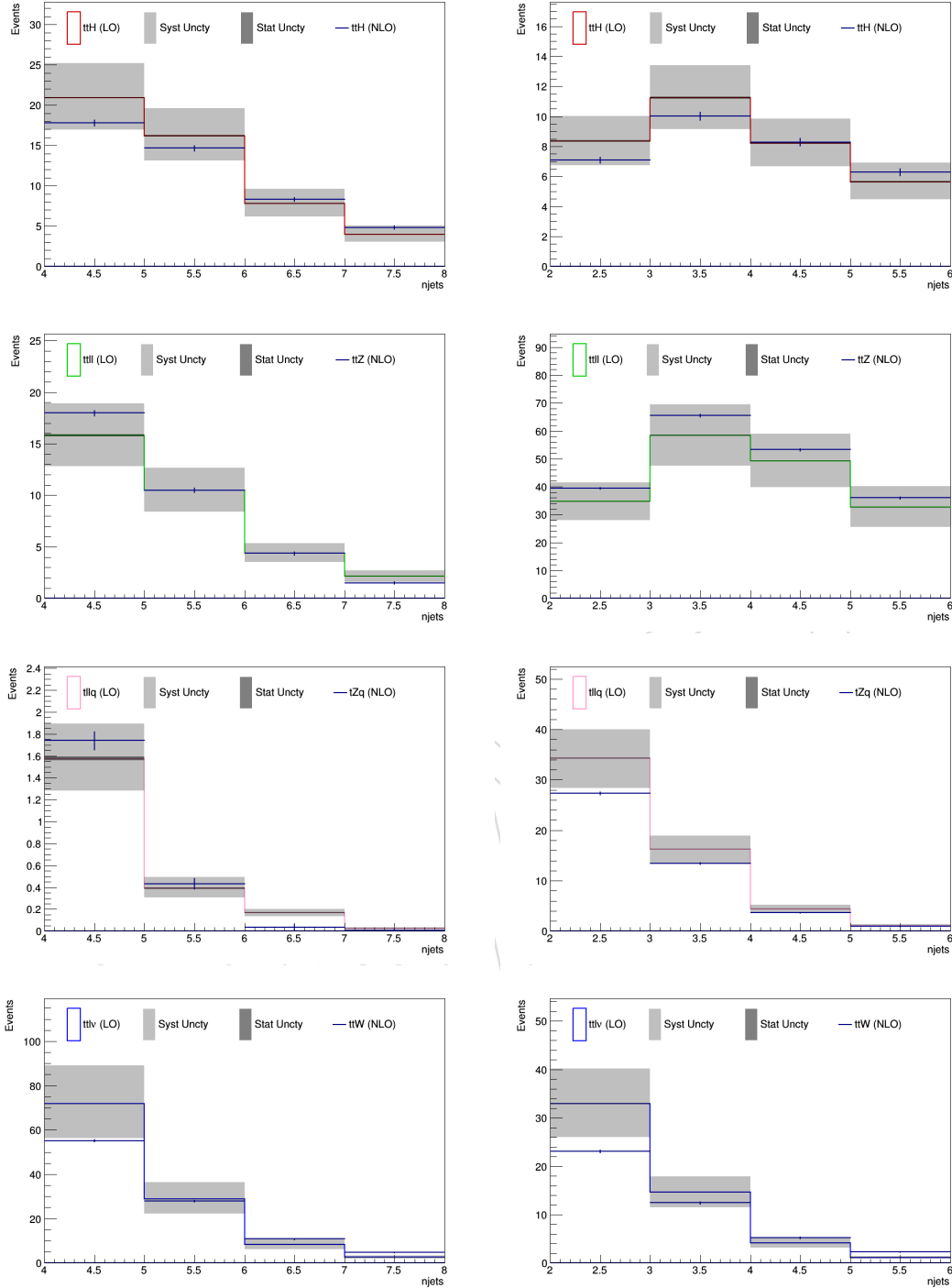


Figure 27: Comparison between centrally-produced NLO MC samples and LO privately-produced MC samples with EFT weights, showing n_{jets} distributions in the signal regions. Events are scaled to the SM expectation; in the case of the LO samples, a k -factor is used to adjust (uniformly and simultaneously across all categories) for the difference between the NLO and LO inclusive SM cross sections. Left column: SM expected yields in combined $2\ell_{\text{ss}}$ categories. Right column: SM expected yields in combined 3ℓ categories. The light(dark) error band centered on the LO sample prediction shows the combined systematic(statistical) uncertainty on the yield, while the error bars on the NLO samples are stats-only. Top to bottom, the rows show the LO(NLO) SM expectation for $t\bar{t}H$ ($t\bar{t}H$), $t\bar{t}l\bar{l}$ ($t\bar{t}Z$), $t\bar{t}l\bar{l}q$ ($t\bar{t}Zq$) and $t\bar{t}l\bar{l}\nu$ ($t\bar{t}W$), respectively. Differences between LO and NLO are expected, but generally fall within the uncertainty of the yield prediction.

C A comparison of the Limits

The wealth of precision measurements presented by the LHC collaborations in recent years, together with the significant progress in the corresponding theoretical calculations and modeling of collider processes, has motivated many groups to pursue (partial) SMEFT analyzes of the LHC data. We have chosen the most recent results of a global EFT fit for doing a comparison of the obtained limits [61]. In Ref. [61], more than 30 independent measurements from 10 different processes at $\sqrt{s} = 8$ and 13 TeV such as inclusive $t\bar{t}$ and single-top production and the associated production of top quarks with weak vector bosons and the Higgs boson are used in a global EFT fit. In figure 28, we have compare the limits obtained in this analysis to the limits obtained in [61]. There are no direct bounds on two leptons - two quarks operators.

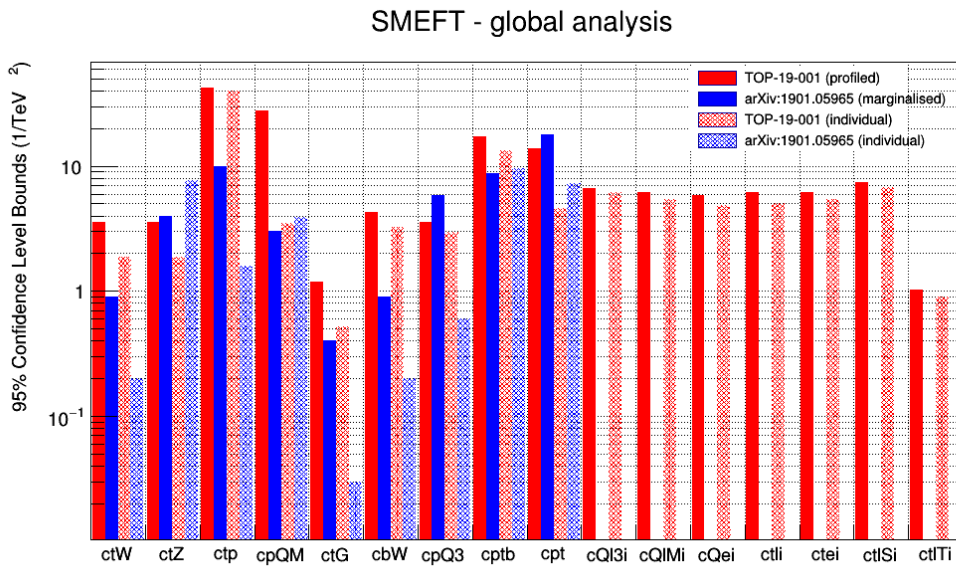


Figure 28: Limits obtained in this analysis (red) are compared to the limits obtained in Ref. [61] (blue). Filled histograms show the limits when other couplings are float and hatched histograms show the limits with other couplings are set to zero.

One should taken into account the following points before comparing our results to the results in Ref. [61].

- In Ref. [61], results of other sensitive processes, like $t\bar{t}$ and single top, are included in the global fit. So some of the effective couplings (e.g. C_{tG}) are more constrained.
- In Ref. [61], results are observed while our results are expected (blinded yet).
- In Ref. [61], effects of EFT operators on backgrounds are ignored.
- In Ref. [61], correlation between various measurements (e.g. $t\bar{t}Z$, tqZ , ...) are ignored.

In addition, some of the EFT operators considered in this analysis are probed in other channels by the CMS collaboration [62, 63]. In figure 29, expected limits obtained in this analysis for each operator at a time (individual) are compared to other CMS analyzes.

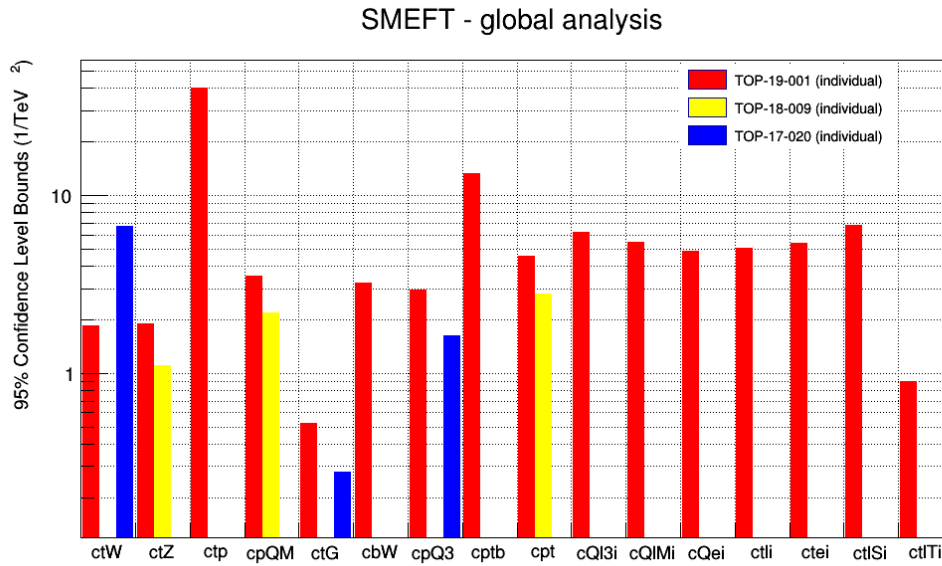


Figure 29: Limits obtained in this analysis (red) are compared to the limits obtained in other CMS analyzes [62, 63].

D Additional plots requested during review

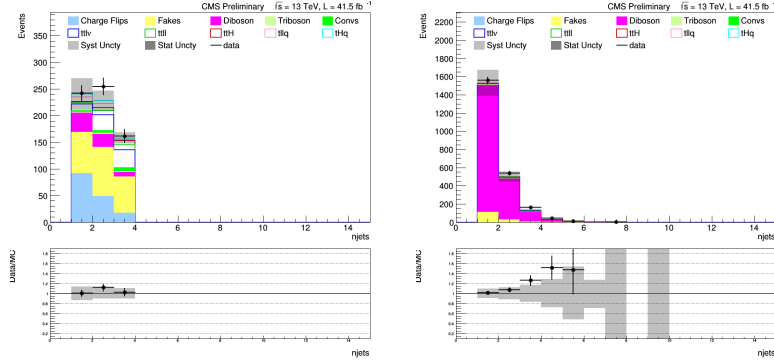


Figure 30: We looked at various methods of extending the control region plots of figs 4 and 5. Adding the 3-jet bin to the 2lss control region was investigated, but was found to have high signal contamination, so this was not adopted (left plot). Methods of forming a non-prompt CR in 3l were also investigated. We attempted to look for bins enriched in non-prompt leptons coming from Z+jets events by requiring two same-flavor, opposite-sign leptons plus an additional lepton, and looking across jet multiplicities. However, we could not find a jet multiplicity bin that was significantly enriched in fakes (right plot). We are constrained to exactly 0 b-tagged jets in the 3l CR in order to avoid overlap with the 3l SRs, so we are not able to look for a 3l non-prompt CR in higher btag multiplicities.

References

- [1] CMS Collaboration, “Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”, arXiv:1711.02547.
- [2] CMS Collaboration, “Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”, Technical Report CMS-PAS-TOP-18-008, CERN, Geneva, 2018.
- [3] CMS Collaboration, “Observation of $t\bar{t}H$ production”, arXiv:1804.02610.
- [4] D. Barducci et al., “Interpreting top-quark LHC measurements in the standard-model effective field theory”, arXiv:1802.07237.
- [5] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, 2017.
- [6] C. Collaboration, “PdmV 2017 Analysis Recipes”.
<https://twiki.cern.ch/twiki/bin/view/CMS/PdmV2017Analysis>.
- [7] 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.
- [8] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A brief introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [9] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [10] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, Technical Report CMS-PAS-GEN-17-001, CERN, Geneva, 2018.
- [11] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [12] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [13] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [14] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (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* **11** (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* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [17] R. Frederix, E. Re, and P. Torrielli, “Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO”, *JHEP* **09** (2012) 130, doi:10.1007/JHEP09(2012)130, arXiv:1207.5391.
- [18] 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.
- [19] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, “W+W-, WZ and ZZ production in the POWHEG BOX”, *JHEP* **11** (2011) 078, doi:10.1007/JHEP11(2011)078, arXiv:1107.5051.
- [20] CMS Collaboration, “Particle-flow event reconstruction in CMS and performance for jets, taus, and E_T^{miss} ”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.
- [21] CMS Collaboration, “Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector”, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.
- [22] 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, CERN, 2010.
- [23] CMS Collaboration, “Particle-flow commissioning with muons and electrons from J/ ψ (1S) and W events at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-003, CERN, 2010.
- [24] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [25] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in pp collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [26] EGamma POG Collaboration.
https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2#Recommended_Muon
- [27] A. Hocker et al., “TMVA - toolkit for multivariate data analysis”, *PoS ACAT* (2007) 040, arXiv:physics/0703039.
- [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] Muon POG Collaboration.
https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Medium_Muon
- [30] M. Cacciari, G. P. Salam, and G. Soyez, “The catchment area of jets”, *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

- [31] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659** (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
- [32] CMS Collaboration, “Search for $t\bar{t}H$ production in multilepton final states at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-15-008, CERN, 2015.
- [33] CMS Collaboration, “Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-16-022, CERN, 2016.
- [34] Banerjee, S. and others, “Search for the associated production of a Higgs boson with a top-quark pair in final states with electrons, muons, and hadronically decaying tau leptons with data recorded at $\sqrt{s} = 13$ TeV in 2017.”, *CMS Analysis Note* **2018/098** (2018).
- [35] CMS Collaboration, “Search for the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV (2017 data)”, CMS Physics Analysis Summary CMS-PAS-HIG-18-019, CERN, 2018.
- [36] C. Botta, M. Peruzzi, and G. Petrucciani.
<https://indico.cern.ch/event/446448/contributions/1952877/attachments/1161380/1672225/lepaware>
- [37] C. Mueller et al., “Search for $t\bar{t}H$ in multilepton final states with 2016 data”, *CMS Analysis Note* **2016/211** (2016).
- [38] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [39] M. Cacciari, G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- [40] JetMET POG. <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2017>.
- [41] 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.
- [42] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [43] A. Tiko.
<https://indico.cern.ch/event/606232/contributions/2444209/attachments/1398539/2133074/chargefli>
- [44] M. Hildreth.
https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVPileUpDescription#Reweighting_Method.
- [45] K. Ehatäht et al.
https://gitlab.cern.ch/ttH_leptons/doc/blob/master/2017/datasets.md#52-data-samples.
- [46] T. Boccali et al.
<https://indico.cern.ch/event/695872/contributions/2877123/attachments/1593469/2522749/pileup-p>
- [47] Bloch, D. and others, “Search for $t\bar{t}H$ in multilepton final states with the full 2016 dataset”, *CMS Analysis Note* **2017/029** (2017).

- [48] BTV POG Collaboration. <https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMETHODS>.
- [49] L. T. Group.
<https://twiki.cern.ch/twiki/bin/view/CMS/L1ECALPrefiringWeightRecipe>.
- [50] A. Brinkerhoff et al., “Search for the standard model Higgs boson produced in association with top quarks and decaying to leptons”, *CMS Analysis Note* **2013/159** (2013).
- [51] B. Stieger et al., “Search for ttH in multilepton final states at 13 TeV”, *CMS Analysis Note* **2015/321** (2016).
- [52] C. H. project, ““Combine Harvester” [software], commit 4adba2b7fcf993ba1304fbddb66abae197549880”.
- [53] H.-C. project, ““HiggsAnalysis-CombineLimit” [software], commit a1dc956452f5b8e055de43bfb093e5243a1e99d9”.
- [54] Muon POG Collaboration.
<https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017>.
- [55] EGamma POG Collaboration.
https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale.
- [56] JetMET POG.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections>.
- [57] N. Bartosik et al., “Calibration of the Combined Secondary Vertex b-Tagging discriminant using dileptonic ttbar and Drell-Yan events”, *CMS Analysis Note* **2013/130** (2013).
- [58] ATLAS, CMS Collaboration, “Search for triboson $W^{\pm}W^{\pm}W^{\mp}$ production in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **77** (2017) 141, doi:10.1140/epjc/s10052-017-4692-1, arXiv:1610.05088.
- [59] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *Eur. Phys. J. C* **75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.
- [60] F. Demartin et al., “The impact of PDF and alphas uncertainties on Higgs Production in gluon fusion at hadron colliders”, *Phys. Rev. D* **82** (2010) 014002, doi:10.1103/PhysRevD.82.014002, arXiv:1004.0962.
- [61] N. P. Hartland et al., “A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector”, arXiv:1901.05965.
- [62] CMS Collaboration, “Search for new physics via top quark production in dilepton final state at 13 TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-17-020, CERN, 2018.
- [63] CMS Collaboration, “Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s}=13$ TeV”, CMS Physics Analysis Summary CMS-PAS-TOP-18-009, CERN, 2019.