

CMS Draft Analysis Note

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

2020/04/13

Archive Hash: a6fc868

Archive Date: 2020/04/13

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

Kevin Lannon¹, Andrew Wightman¹, Geoffrey Smith¹, Reza Goldouzian¹, Michael Hildreth¹, Kelci Mohrman¹, Brian Winer², Wuming Luo², Anthony Lefeld², and Brent R. Yates²

¹ University of Notre Dame

² The Ohio State University

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 pp collision dataset collected by the CMS experiment in 2017.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor:	The TOP EFT working group
PDFTitle:	Probing Effective Field Theory Models using Associated Top Quark Production in Multiple Lepton Final States at 13 TeV.
PDFSubject:	CMS
PDFKeywords:	CMS, physics, top, EFT

Please also verify that the abstract does not use any user defined symbols

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	5
6	3	Data and MC Samples	7
7	3.1	Data Samples and Triggers	8
8	3.2	Signal Samples	9
9	3.3	Background Samples	12
10	4	Object reconstruction and identification	13
11	4.1	Lepton Identification	13
12	4.2	Jets and B-tagging	17
13	5	Event selection and categorization	19
14	5.1	$2\ell ss$ category	19
15	5.2	3ℓ category	19
16	5.3	4ℓ category	20
17	5.4	Separation into sub-categories	20
18	6	Data-to-MC corrections	20
19	6.1	Pileup reweighting	21
20	6.2	Trigger efficiency	22
21	6.3	Identification and isolation efficiency for e and μ	22
22	6.4	b-tag efficiency and mistag rate	23
23	6.5	ECAL prefiring correction	23
24	7	Background estimation	25
25	7.1	“Fake” background	25
26	7.2	Charge “flip” background	29
27	8	Signal extraction and EFT model parameters fitting	32
28	9	Systematic uncertainties	35
29	10	Results	42
30	10.1	Standard Model Signals	42
31	10.2	EFT Signals	43
32	A	Synchronization Plots	80
33	B	MC Validation	83
34	C	Systematics Checks	94
35	D	A comparison of the Limits	96
36	E	Additional plots requested during review	98
37			

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 the full list nature's constituents has not yet been discovered. Likewise, an explanation of 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 violations, this analysis neglects these operators in favor of dimension-six operators as the leading contributors. The first sub-leading contributions that conserve baryon and lepton number arise from dimension-eight operators and are not considered in this analysis.

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 an 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 directly test 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 b 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 (b-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 physical 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 physical processes contribute to the signal region, 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 $t\bar{t}W$ and $t\bar{t}H$ production. Each of these physical 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 charge 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 sigma 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:

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

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:

$$\sigma \propto |\mathcal{M}|^2 \propto \sum_{j,k=0}^N s_{jk} c_j c_k. \quad (4)$$

The s_{jk} are structure constants of a multi-dimensional quadratic function where N is the number of WC considered and the $s_{00}c_0^2$ term is understood to correspond to the purely SM piece. Though the interference terms are less suppressed by Λ , we also include the quadratic contributions since their effects can be large (especially in cases when the limits on the WC are not strong). We can solve for the structure constants by evaluating the cross section at multiple points in WC space. Once the constants are known, the fit can be used to determine σ for any combination of WC values. This procedure is valid for 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; the validation of the reweighting procedure is discussed in Appendix B. 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 is the dim6TopEFT model [4]. Some of the important aspects of this implementation are listed below.

- The degrees of freedom implemented in the dim6TopEFT model are derived from the Warsaw basis of dimension 6 operators [5]. A map between the Wilson Coefficients considered in this analysis and the corresponding Warsaw basis operators is shown in Table 2.
- 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.

Since only tree-level simulation is possible with this model, we include an extra final state parton in our privately produced signal samples. To allow MADGRAPH to properly account for the effects of this extra parton, some modifications to the dim6TopEFT model were needed; the modifications are described in Section 3.2.

2.3 Operator Selection

Barring flavor structure, there are 59 independent dimension-six operators that conserve baryon and lepton number [5]. In order to reduce the phase space of all possible NP coupling strengths, we consider only a subset of the dimension-six operators containing at least one top quark that are described in the dim6TopEFT model. The excluded operators have their couplings fixed to their SM values of zero.

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 have 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 any signal process by more than a factor of five and also not 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. The number of terms needed to parameterize the cross section dependence on WCs scales as $\mathcal{O}(n^2)$. In order to keep the computational costs of extracting and storing the parameterization reasonable, only the top 16 WCs were chosen. The WC cQq83 was excluded in favor of ctZ in order to round out the list of 'two heavy + boson' and 'two heavy + two lepton' WCs. The only two WCs from these two categories not considered are cpb, which involves no top, and cblSi, which this analysis is not expected to be very sensitive to.

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, ctIS, and ctIT. 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.

WC	FOM	$t\bar{t}H$	$t\bar{t}l\bar{l}$	$t\bar{t}l\nu$	$t\bar{t}lq$	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 $\sigma_{SM+NP}/\sigma_{SM} < 5$ 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.

Operators involving two quarks and one or more bosons		
Operator	Definition	Wilson coefficient
$\dagger O_{u\varphi}^{(ij)}$	$\bar{q}_i u_j \tilde{\varphi} (\varphi^\dagger \varphi)$	$c_{t\varphi} + ic_{t\varphi}^I$
$O_{\varphi q}^{1(ij)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi) (\bar{q}_i \gamma^\mu q_j)$	$c_{\varphi Q}^- + c_{\varphi Q}^3$
$O_{\varphi q}^{3(ij)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu^I \varphi) (\bar{q}_i \gamma^\mu \tau^I q_j)$	$c_{\varphi Q}^3$
$O_{\varphi u}^{(ij)}$	$(\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi) (\bar{u}_i \gamma^\mu u_j)$	$c_{\varphi t}$
$\dagger O_{\varphi ud}^{(ij)}$	$(\tilde{\varphi}^\dagger i D_\mu \varphi) (\bar{u}_i \gamma^\mu d_j)$	$c_{\varphi tb} + ic_{\varphi tb}^I$
$\dagger O_{uW}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} \tau^I u_j) \tilde{\varphi} W_{\mu\nu}^I$	$c_{tW} + ic_{tW}^I$
$\dagger O_{dW}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} \tau^I d_j) \varphi W_{\mu\nu}^I$	$c_{bW} + ic_{bW}^I$
$\dagger O_{uB}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} u_j) \tilde{\varphi} B_{\mu\nu}$	$(c_{tW} - c_{tZ}) / + i(c_{tW}^I - c_{tZ}^I) /$
$\dagger O_{uG}^{(ij)}$	$(\bar{q}_i \sigma^{\mu\nu} T^A u_j) \tilde{\varphi} G_{\mu\nu}^A$	$c_{tG} + ic_{tG}^I$
Operators involving two quarks and two leptons		
Operator	Definition	Wilson coefficient
$O_{\ell q}^{1(ijkl)}$	$(i \gamma^\mu \ell_j) (\bar{q}_k \gamma^\mu q_\ell)$	$c_{Q\ell}^{-(\ell)} + c_{Q\ell}^{3(\ell)}$
$O_{\ell q}^{3(ijkl)}$	$(i \gamma^\mu \tau^I \ell_j) (\bar{q}_k \gamma^\mu \tau^I q_\ell)$	$c_{Q\ell}^{3(\ell)}$
$O_{\ell u}^{(ijkl)}$	$(i \gamma^\mu \ell_j) (\bar{u}_k \gamma^\mu u_\ell)$	$c_{t\ell}^{(\ell)}$
$O_{e\bar{q}}^{(ijkl)}$	$(\bar{e}_i \gamma^\mu e_j) (\bar{q}_k \gamma^\mu q_\ell)$	$c_{Qe}^{(\ell)}$
$O_{eu}^{(ijkl)}$	$(\bar{e}_i \gamma^\mu e_j) (\bar{u}_k \gamma^\mu u_\ell)$	$c_{te}^{(\ell)}$
$\dagger O_{\ell equ}^{1(ijkl)}$	$(i e_j) \varepsilon (\bar{q}_k u_\ell)$	$c_t^{S(\ell)} + ic_t^{SI(\ell)}$
$\dagger O_{\ell equ}^{3(ijkl)}$	$(i \sigma^{\mu\nu} e_j) \varepsilon (\bar{q}_k \sigma_{\mu\nu} u_\ell)$	$c_t^{T(\ell)} + ic_t^{TI(\ell)}$

Table 2: Operators that have effects on $t\bar{t}H$, $t\bar{t}l\bar{l}$, $t\bar{t}l\nu$, $t\bar{t}lq$ and tHq processes at order $1/\Lambda^2$ and are considered in this analysis. The couplings are assumed to involve only third-generation quarks. The field φ ($\tilde{\varphi} = \varepsilon\varphi^*$) is the Higgs boson doublet. The ℓ and q represent the left-handed lepton and quark doublets respectively, while e represent the right-handed lepton, and u and d represent the right-handed quark singlets. The quantity $T^A = \frac{1}{2}\lambda^A$ denotes the eight Gell-Mann matrices. $D_\mu = \partial_\mu - ig_s \frac{1}{2}\lambda^A G_\mu^A - ig \frac{1}{2}\tau^I W_\mu^I - ig' Y B_\mu$ is the covariant derivative. $W_{\mu\nu}^I = \partial_\mu W_\nu^I - \partial_\nu W_\mu^I + g\varepsilon_{IJK} W_\mu^J W_\nu^K$ 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. The abbreviations and denote the sine and cosine of the weak mixing angle (in the unitary gauge). More details about the operators can be found in Ref. [4].

3 Data and MC Samples

In the following we list as reference the set of samples of the RunII Fall17 MiniAOD (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 dataset as due to 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 $t\bar{t}H$ multilepton group for the lepton MVA, including scale factors and fake rates, and those are not available for the 2018 dataset at this time.

On 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 analyses. We anticipate that this first result will generate important discussions 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 time to incorporate feedback received from the HEP community on the first version of the analysis. During the 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 41.53 fb^{-1} [6]. We use the 31Mar2018 version of the datasets. Details on the various datasets contributing to our analysis are given in Table 4. Dataset overlaps are removed by iteratively restricting the list of allowed triggers for a given event to be included. The overlap removal proceeds as follows: For dataset A, if an event passes any dataset A dependent trigger, then it is included; For dataset B, if an event passes any dataset B dependent trigger and does not pass any dataset A dependent trigger, then it is included; For dataset C, an event must pass triggers from C and not pass any triggers from A or any triggers from B. The process continues for every dataset considered in the analysis.

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 two-lepton-same-sign ($2\ell ss$) channels, while events in the three-lepton (3ℓ) and four-lepton (4ℓ) channels are recorded using

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.

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, which pass any of the HLT paths that are included in the mix are selected. The effect of triggers that are disabled or prescaled in some part of the analyzed data are 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 tHq . An additional set of $t\bar{t}H$, $t\bar{t}Z$, $t\bar{t}W$, tZq , and tHq are simulated with aMC@NLO [8] and matched to PYTHIA [9] for the parton shower. These centrally produced samples are only used for validation of the privately produced samples. Details on the centrally produced samples are in Table 5.

In order to improve the accuracy of our privately produced LO samples and facilitate better agreement with the centrally produced NLO samples, we include an extra parton in the final states of the $t\bar{t}H$, $t\bar{t}l\bar{l}$, and $t\bar{t}l\nu$ samples; however, this required modifications to the dim6TopEFT model. First, in order to allow MADGRAPH to properly handle the emission of a gluon from non-QCD vertices (specifically, the vertices arising from the O_{tG} operator), an extra factor of the strong coupling was applied to the operator's coefficients. This was necessary because the MADGRAPH program assumes that gluon emission from quarks is due to QCD, but the vertices arising from the O_{tG} operator include (non QCD) emission of gluons, so MADGRAPH incorrectly groups diagrams containing these vertices. However, with an extra factor of the strong coupling included in the definition of the problematic vertices (i.e. ctG and $ctGI$ vertices), MADGRAPH's classification of diagrams should again be appropriate. Next, it was found that the dim6TopEFT model does not include any five-particle vertices. This presented an issue for our $t\bar{t}H$ sample since the operator gives rise to a five-particle vertex (involving top quarks, gluons, and a Higgs boson) relevant to the $t\bar{t}H$ process. The model file was modified to include this missing vertex. Additionally, the five-particle vertices for ctW , ctZ , and cbW were

Dataset	Run Range	Int. Lumi (fb^{-1})
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.82 fb^{-1}
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.66 fb^{-1}
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.25 fb^{-1}
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.28 fb^{-1}
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.54 fb^{-1}
Total DoubleMuon	297046–306462	41.53 fb^{-1}
/DoubleEG/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.82 fb^{-1}
/DoubleEG/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.66 fb^{-1}
/DoubleEG/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.25 fb^{-1}
/DoubleEG/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.28 fb^{-1}
/DoubleEG/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.54 fb^{-1}
Total DoubleEG	297046–306462	41.53 fb^{-1}
/MuonEG/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.82 fb^{-1}
/MuonEG/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.66 fb^{-1}
/MuonEG/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.25 fb^{-1}
/MuonEG/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.28 fb^{-1}
/MuonEG/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.54 fb^{-1}
Total MuonEG	297046–306462	41.53 fb^{-1}
/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.82 fb^{-1}
/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.66 fb^{-1}
/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.25 fb^{-1}
/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.28 fb^{-1}
/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.54 fb^{-1}
Total SingleMuon	297046–306462	41.53 fb^{-1}
/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD	297046–299329	4.82 fb^{-1}
/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD	299368–302029	9.66 fb^{-1}
/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD	302030–303434	4.25 fb^{-1}
/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD	303824–304797	9.28 fb^{-1}
/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD	305040–306462	13.54 fb^{-1}
Total SingleElectron	297046–306462	41.53 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 [7].

also included, though these vertices do not affect the processes considered in our analysis. The modified dim6TopEFT model file can be found here [10]. While these modifications allowed us to generate $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ samples with an extra final state parton, the $t\bar{t}lq$ and tHq signal samples are not generated with an extra final state parton due to complications associated with correctly performing the jet matching between the matrix element and the parton shower for t-channel single top processes (such as $t\bar{t}lq$ and tHq).

Each of the signal samples is produced using leading-order (LO) matrix elements implemented in the MADGRAPH5_aMC@NLO 2.6.0 program [8] with the modified version of the dim6TopEFT physics model. In order to populate as much of the relevant phase space as possible, each signal sample was generated at a non-SM point; each sample's starting point is listed in Appendix B. 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 7.

The $t\bar{t}t\bar{t}$ sample is excluded from the analysis because the operators considered are expected to

Process	Xsec (pb)	Sample
tH ($H \rightarrow \text{non-bb}$)	0.5085 ¹ (1-0.577)	/tHJetToNonbb.M125_TuneCP5.13TeV-amcatnloFXFX-madspin-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018_new.pmx.94X_mc2017_realistic_v14-v1/MINIAODSIM
tW ($W \rightarrow \ell\nu$)	0.2043	/TTWJetsToLNu.TuneCP5.PSweights.13TeV-amcatnloFXFX-madspin-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018_new.pmx.94X_mc2017_realistic_v14-v1/MINIAODSIM
tZ ($Z \rightarrow \ell\bar{\ell}, \nu\bar{\nu}$)	0.2529	/TTZToLNuNu.M-10.TuneCP5.PSweights.13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic_v14-v1/MINIAODSIM
tZq ($Z \rightarrow \ell\bar{\ell}$)	0.0942	/tZq.JL4f_ckm.NLO.TuneCP5.PSweights.13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X_mc2017_realistic_v14-v1/MINIAODSIM

Table 5: Centrally produced signal samples used only for validation. Note there is no appropriate central sample for the tHq process.

MG Setting	EFT ttX	EFT tXq	tH	tZ	tW	tZq
p_T^j	auto	$> 0.1 \text{ GeV}$	—	—	—	$> 0.1 \text{ GeV}$
p_T^ℓ	*	*	—	*	*	*
$ \eta^j $	*	*	—	—	—	*
$ \eta^\ell $	*	*	—	*	*	*
$\Delta R_{\ell\ell}$	*	*	—	*	*	*
$m_{\ell\ell}$	$> 10 \text{ GeV}^1$	$> 30 \text{ GeV}^1$	—	$> 10.0 \text{ GeV}$	—	$> 30.0 \text{ GeV}$

¹ The $m_{\ell\ell}$ cut only applies to $t\bar{t}l\bar{l}$ and $t\bar{t}lq$ samples.

Table 6: Comparison of the different cuts between the private EFT and central samples. p_T^j/p_T^ℓ is the jet/lepton p_T , η^j/η^ℓ is the jet/lepton pseudorapidity, $\Delta R_{\ell\ell}$ is the distance between any pair of leptons, and $m_{\ell\ell}$ is the invariant mass of any SFOS leptons. A (*) indicates that the cut is not applied and a (—) indicates that the cut is not applicable for that process. There is no appropriate tHq central sample with which to compare, so there is no central tHq in this table.

have a negligible impact on the $t\bar{t}t\bar{t}$ process relative to the other signal processes. The expected impact on the $t\bar{t}t\bar{t}$ 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 $t\bar{t}t\bar{t}$ 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, the re-weighting procedure described in Section 2.1 is used. The re-weighted points are 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.

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. Appendix B describes this validation, and Figure 45 (in Appendix B) shows the comparison to the 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 [11] set of parton distri-

Process	Xsec (pb)	Location	Events
tHJ	0.2151	/store/user/kmohrman/FullProduction/Round6/Batch7/postLHE_step/v2/mAOD_step.tHJJet.*	12.8M
tHnu	0.2043	/store/user/kmohrman/FullProduction/Round6/Batch1/postLHE_step/v1/mAOD_step.tHnuJet.*	4.9M
		/store/user/kmohrman/FullProduction/Round6/Batch2/postLHE_step/v1/mAOD_step.tHnuJet.*	2.9M
		/store/user/kmohrman/FullProduction/Round6/Batch5/postLHE_step/v1/mAOD_step.tHnuJet.*	2.9M
		/store/user/kmohrman/FullProduction/Round6/Batch6/postLHE_step/v2/mAOD_step.tHnuJet.*	4.9M
		/store/user/kmohrman/FullProduction/Round6/Batch3/postLHE_step/v1/mAOD_step.tHq4f.*	10.0M
tHq	0.0771	/store/user/kmohrman/FullProduction/Round6/Batch1/postLHE_step/v1/mAOD_step.tHq4fNoSchanWNoHiggs0p.*	5.6M
		/store/user/kmohrman/FullProduction/Round6/Batch2/postLHE_step/v1/mAOD_step.tHq4fNoSchanWNoHiggs0p.*	3.4M
		/store/user/kmohrman/FullProduction/Round6/Batch5/postLHE_step/v1/mAOD_step.tHq4fNoSchanWNoHiggs0p.*	3.4M
		/store/user/kmohrman/FullProduction/Round6/Batch6/postLHE_step/v2/mAOD_step.tHq4fNoSchanWNoHiggs0p.*	5.6M
tHl	0.2529	/store/user/awightma/postLHE_step/2019.04.19/tHlq4f-tch-NoHiggs.0partons.xqcut0/v2/mAOD_step.tHlq4fNoSchanWNoHiggs0p.*	0.5M
		/store/user/awightma/postLHE_step/2019.04.19/tHlq4f-tch-NoHiggs.0partons.xqcut0.extra/v2/mAOD_step.tHlq4fNoSchanWNoHiggs0p.*	2.0M
		/store/user/awightma/postLHE_step/2019.04.19/tHlq4f-tch-NoHiggs.0partons.xqcut0.extra2/v1/mAOD_step.tHlq4fNoSchanWNoHiggs0p.*	5.0M
		/store/user/kmohrman/FullProduction/Round6/Batch4/postLHE_step/v1/mAOD_step.tHlq4fNoSchanWNoHiggs0p.*	5.0M
		/store/user/kmohrman/FullProduction/Round6/Batch6/postLHE_step/v2/mAOD_step.tHlq4fNoSchanWNoHiggs0p.*	5.0M

Table 7: Privately produced signal samples used in this analysis, located on the Notre Dame T3.

Sample	Xsec (pb)
/WZTo3LNU_TuneCP5.13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.new.pmx.94X.mc2017.realistic.v14-v1/MINIAODSIM	4.42965
/WWTo2L2Nu_NNPFD31_TuneCP5.13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.new.pmx.94X.mc2017.realistic.v10-v1/MINIAODSIM	10.481
/ZZTo4L.13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.new.pmx.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
/WWW.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
/TTGJets_TuneCP5.13TeV-amcatnloFXFX-madspin-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X.mc2017.realistic.v14-v1/MINIAODSIM	3.697
/TTJets_TuneCP5.13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X.mc2017.realistic.v14-v1/MINIAODSIM	831.76
/ST.4-channel.top.4f.inclusiveDecays.TuneCP5.13TeV-powhegV2-madspin-pythia8/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X.mc2017.realistic.v14-v1/MINIAODSIM	136.02
/ST.4-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/RunIIFall17MiniAODv2-PU2017.12Apr2018.94X.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 8: 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.

but ion functions. Parton shower and hadronization processes are modelled using the generator PYTHIA [9] with the CP5 tune [12].

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 [13], and are reconstructed using the same version of the CMS event reconstruction software as used for data. Each signal sample was validated by comparing gen-level and basic reconstructed objects to the corresponding centrally produced sample, details can be found in Appendix B.

3.3 Background Samples

The simulated samples used to estimate the backgrounds include $t\bar{t}\gamma$, diboson production, and triboson production. Additional samples are also generated to simulate SM background processes that are estimated from data, and are used for validation purposes (see 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 [8, 14, 15] and POWHEG [16–21] programs. No dim6 EFT effects are included in any of the background samples. The same PYTHIA chain that was outlined in section 3.2 is also used for every background sample. See Table 8 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 the particle-flow (PF) algorithm [22–26] 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 the $t\bar{t}H$ multilepton analysis [27, 28], discussed in Section 4.1.4.

4.1.1 Basic electron identification

The identification of electrons is performed by a multivariate algorithm [29, 30], based on a BDT [31], which is 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. [30].

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 [32]. 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 [33].

379 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. [34, 35]. The size of the cone is given by:

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

380 The effective area \mathcal{A} is obtained from the simulation, by studying the correlation between I_ℓ
 381 and ρ , and is determined in bins of η , separately for electrons and muons. Numerical values
 382 are given in Table 9.

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 9: 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 [36, 37]. 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 $t\bar{t}H$ multilepton analysis. The method is described briefly below; for more information please see here [27, 28].

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 jets of $p_T > 15$ GeV are 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` are set to zero. The p_T of the jet is computed by applying the “lepton-aware” jet energy corrections (JEC) [38], 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 $t\bar{t}H$ events (signal) and non-prompt leptons in $t\bar{t}$ +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 11 and 12. 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. The primary difference between the two is whether the leptons pass (tight) or fail (fakeable) the BDT cut. In order to reduce potential biases of the background estimation procedure, the p_T of leptons in the fakeable selection 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 ($1/E - 1/p$). These 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.

EGamma POG MVA cuts			
	$ \eta < 0.8$	$0.8 < \eta < 1.44$	$ \eta > 1.44$
$p_T < 10 \text{ GeV}$	-0.13	-0.32	-0.08
$p_T > 10 \text{ GeV}$	-0.86	-0.81	-0.72

Table 10: Cuts on the EGamma POG MVA for different bins in electron $|\eta|$ and p_T .

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	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	See Table 10	> 0.50	See Table 10
Conversion rejection	—	✓	✓
$\sigma_{i\eta i\eta}$	—	$< \{0.011 / 0.011 / 0.030\}$	$< \{0.011 / 0.011 / 0.030\}$
H/E	—	< 0.10	< 0.10
$1/E - 1/p$	—	$> -0.04 \frac{1}{\text{GeV}}$	$> -0.04 \frac{1}{\text{GeV}}$
p_T^e / p_T^j	—	> 0.6	—
DeepCSV(b) of nearby jet	—	< 0.07	< 0.4941
Prompt-e MVA	—	< 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 11: 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. A hyphen (—) indicates selection criteria that are not applied.

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	—
p_T^μ/p_T^j	—	> 0.6	—
DeepCSV(b) of nearby jet	—	< 0.07	< 0.4941
Prompt- μ MVA	—	< 0.90	> 0.90

Table 12: Loose, fakeable, and tight selection criteria for muons. 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 [39, 40]. The charged hadrons 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 [41]. The prescribed jet energy corrections in Table 13 are applied as a function of the jet E_T and η [42]. 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 13: 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 [43] 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 ($\text{DeepCSV(b)} > 0.1522$) has approximately 85% efficiency to tag jets from b quarks and a 10% chance to tag jets from light quarks or gluons. The *medium* working point ($\text{DeepCSV(b)} > 0.4941$) has approximately 70%

467 efficiency for tagging jets from b quarks and 1.0% chance to tag jets from light quarks or gluons
468 [43].
469

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. Table 14 summarises the category and subcategory divisions described in the following sections.

5.1 $2\ell ss$ category

The two lepton same-sign ($2\ell ss$) 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 $t\bar{t} + \text{jets}$ background. Residual $t\bar{t} + \text{jets}$ background contributions are further suppressed by requiring that the charge is well-measured for all electrons and muons in the event which pass the fakeable object selection criteria. 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 [44]. 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. Including events with three jets was found to introduce large contributions from the data-driven backgrounds and so is excluded. 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 $p_T > 25 \text{ GeV}$, $> 15 \text{ GeV}$, and $> 10 \text{ GeV}$, respectively. In the case that the trailing lepton is an electron, the requirement on it is instead $p_T > 15 \text{ GeV}$. Two or more jets of $p_T > 30 \text{ GeV}$ 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, as well as both t quarks, 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 $p_T > 25 \text{ GeV}$, $> 15 \text{ GeV}$, $> 10 \text{ GeV}$, and $> 10 \text{ GeV}$, 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 $p_T > 15 \text{ GeV}$. Two or more jets of $p_T > 30 \text{ GeV}$ 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 total lepton charge. This 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. 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 lepton charges, 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 $t\bar{t}l\bar{l}q$ is enhanced in the former case. The more stringent b-tag requirement ensures better separation between $t\bar{t}l\bar{l}q$ and $t\bar{t}l\bar{l}$ events by keeping $t\bar{t}l\bar{l}q$ events from being promoted to to 2b events due to misidentification of b-jets.

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 14 summarises the subcategory divisions.

Table 15 summarizes the expected number of events in each category, Table 16 summarizes the expected number of events in each category after fitting for the various nuisance parameters, and Figures 1 and 2 show the expected number of events per category as a function of jet multiplicity.

6 Data-to-MC corrections

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

- Pileup reweighting

	$2\ell ss$	3ℓ	$\geq 4\ell$
Tight Leptons	2	3	≥ 4
Jets	≥ 4	≥ 2	≥ 2
DeepCSV(b) Medium Jets	≥ 1	1, ≥ 2	≥ 1
DeepCSV(b) Loose Jets	≥ 2	-	≥ 2
nJet Subcategories	4, 5, 6, ≥ 7	2, 3, 4, ≥ 5	2, 3, ≥ 4
Other Subcategories	Lepton Charge	Sign of Net Lepton Charge "SFOS Z"	- -

Table 14: Object requirements for the different event categories. Requirements separated by commas indicate a division into subcategories.

	$2\ell ss$ ("p")	$2\ell ss$ ("m")	3ℓ (1b "p")	3ℓ (1b "m")	3ℓ ($\geq 2b$ "p")	3ℓ ($\geq 2b$ "m")	3ℓ (SFOS Z, 1b)	3ℓ (SFOS Z, $\geq 2b$)	$\geq 4\ell$
Diboson	1.6	1.2	5.9	4.7	0.4	0.3	52.1	4.1	0.6
Triboson	0.5	0.5	0.2	0.2	0.0	0.1	3.5	0.6	0.1
Charge Flips	8.5	8.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fakes	25.6	26.8	11.3	13.0	3.3	2.5	16.9	3.8	0.0
Conversions	10.9	9.2	2.3	2.6	1.7	1.9	0.8	0.4	0.0
Sum Background	47.0	46.2	19.7	20.5	5.4	4.8	73.3	8.9	0.7
$t\bar{t}l\nu$	68.7	37.1	14.4	8.0	10.8	5.9	2.9	2.3	0.0
$t\bar{t}l\bar{l}$	19.3	19.0	12.7	13.3	9.1	8.5	95.5	63.2	9.4
$t\bar{t}H$	24.7	24.1	7.9	7.6	5.1	5.2	3.2	2.2	1.0
$t\bar{t}lq$	2.7	1.5	3.5	1.8	1.2	0.6	39.8	13.3	0.0
tHq	0.8	0.4	0.3	0.2	0.2	0.1	0.1	0.1	0.0
Sum Signal	116.2	82.2	38.9	30.8	26.3	20.3	141.6	81.1	10.4
Total Expected	163 ± 20	128 ± 15	59 ± 7	51 ± 6	32 ± 4	25 ± 3	215 ± 25	90 ± 13	11 ± 2
Data	192	171	85	64	32	28	239	95	12

Table 15: SM expected yields after the selection in all categories and final states. These categories are further broken down based on the number of jets in the event.

- 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 [45]. 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. [46], 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 [47]. We work around the problem using the distribution

	$2\ell ss$ ("p")	$2\ell ss$ ("m")	3ℓ (1b "p")	3ℓ (1b "m")	3ℓ ($\geq 2b$ "p")	3ℓ ($\geq 2b$ "m")	3ℓ (SFOS Z, 1b)	3ℓ (SFOS Z, $\geq 2b$)	$\geq 4\ell$
Diboson	1.7	1.5	6.6	5.5	0.5	0.4	61.1	4.6	0.1
Triboson	0.6	0.5	0.0	0.0	0.0	0.0	0.5	0.1	0.0
Charge Flips	8.8	8.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fakes	37.0	39.6	17.2	18.4	4.6	3.2	24.2	5.5	0.0
Conversions	12.3	10.4	2.8	2.8	2.0	2.1	1.1	0.5	0.0
Sum Background	60.4	60.7	26.6	26.7	7.1	5.8	86.9	10.7	0.1
$t\bar{t}l\nu$	81.1	44.0	16.6	9.1	12.1	6.7	3.4	2.5	0.0
$t\bar{t}l\bar{l}$	22.6	22.5	14.2	14.7	10.1	9.4	106.5	70.9	10.4
$t\bar{t}H$	28.6	27.9	8.5	8.1	5.5	5.6	3.5	2.4	1.1
$t\bar{t}lq$	2.9	1.7	3.8	1.9	1.3	0.6	42.1	14.1	0.0
tHq	0.9	0.5	0.3	0.2	0.2	0.1	0.1	0.1	0.0
Sum Signal	136.1	96.5	43.4	34.0	29.1	22.4	155.6	90.0	11.6
Total Expected	197 ± 9	157 ± 7	70 ± 4	61 ± 4	36 ± 2	28 ± 2	242 ± 10	101 ± 6	12 ± 1
Data	192	171	85	64	32	28	239	95	12

Table 16: SM expected yields after fitting for nuisance parameters. These categories are further broken down based on the number of jets in the event.

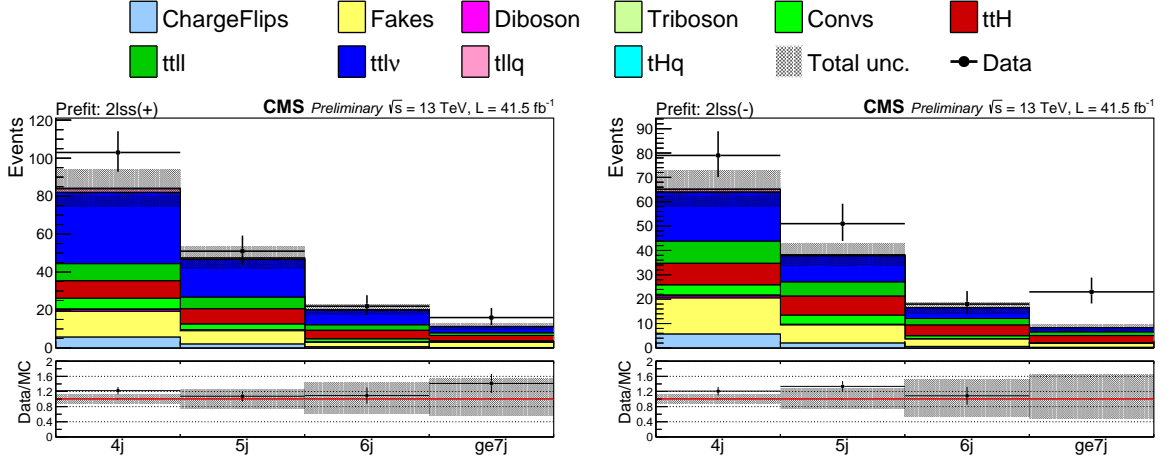


Figure 1: SM Expected number of signal and background events in the $2\ell ss$ categories. From left to right, the plots show events in the $2\ell ss("p")$ and $2\ell ss("m")$ categories.

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

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 17 are corrected by applying the ratio of the efficiency in data to the efficiency in the MC simulation as weights 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 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. This extra uncertainty is in addition to the ones shown in Table 17.

The SFs 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 [27] using this same method are re-used here. The SFs are given in Table 17.

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 11 and 12 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. [48] has been re-

peated for data recorded in 2017 (see Ref. [28] 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 a 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 DeepCSV algorithm, and in the mistag rate for light flavor (u, 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 [49]. The per-jet SFs 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 at 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 [50]. 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 varying the correction up/down 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.

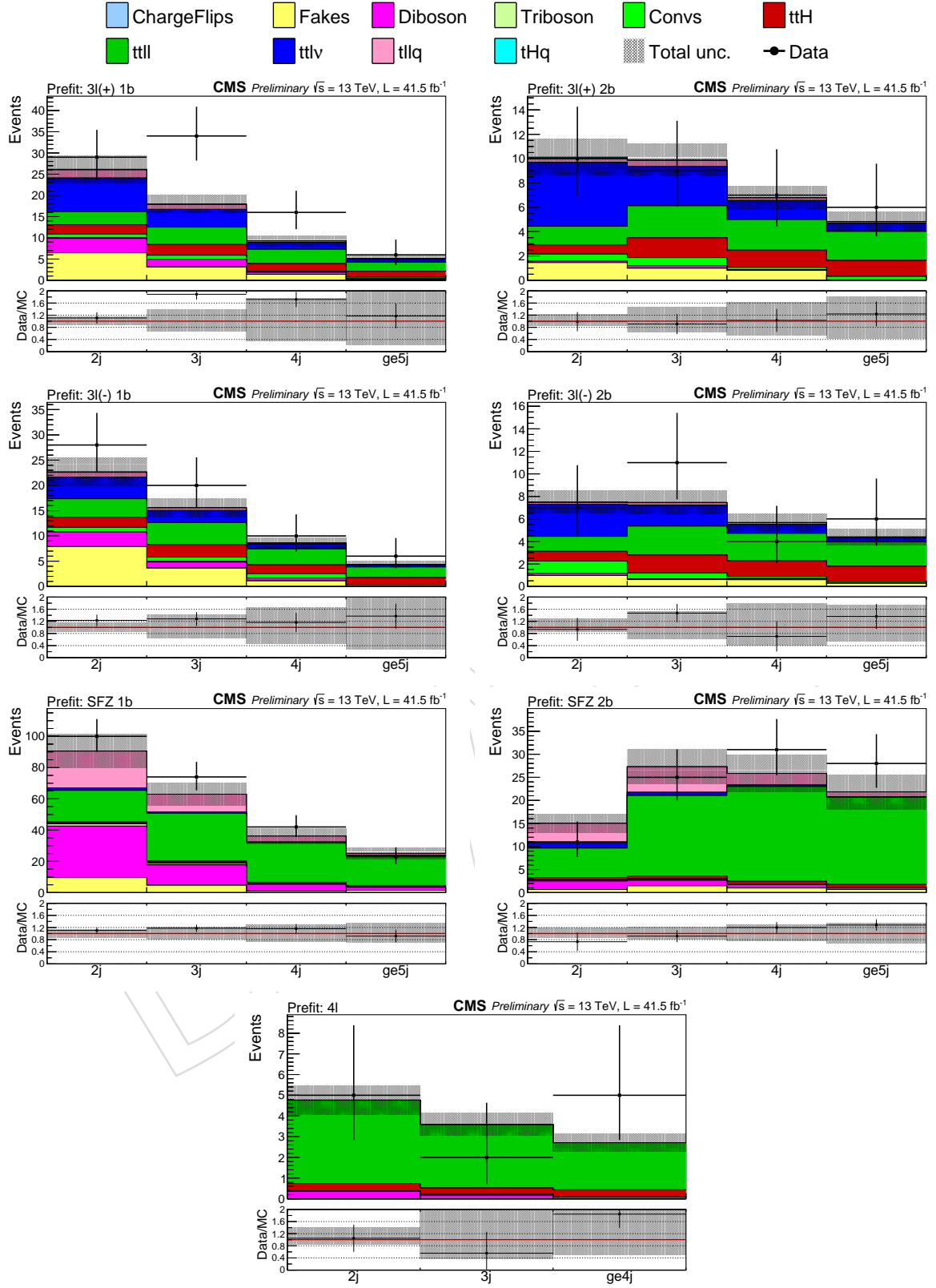


Figure 2: SM Expected number of signal and background events in the 3ℓ and 4ℓ categories. Going from left to right and top to bottom on a per-row basis, the plots show events in the $3\ell(1b \text{ "p"})$, $3\ell(\geq 2b \text{ "p"})$, $3\ell(1b \text{ "m"})$, $3\ell(\geq 2b \text{ "m"})$, $3\ell(\text{SFOS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

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 $t\bar{t}l\bar{l}$ 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. [48, 51–53]. 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 11). 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 fakes background is based on a fake-factor (FF) method. Briefly, the FF method first uses events in data containing non-prompt leptons to determine the FF as a ratio of the number of leptons passing the tight selection to the number passing the fakeable selection. The FF is then applied to other data in an event selection named the application region (AR), which uses our standard event selection except both fakeable and tight leptons are allowed. This is made orthogonal to our signal region (SR) by vetoing events where all of the leptons are tight. The FF application determines the number of fakes that are in the SR based on the number of events in the AR.

This method is identical to the one used by the CMS $t\bar{t}H$ multilepton analysis [27]. We synchronize our object and event selections with this analysis and use the same FF. Information about this synchronization can be found in Appendix A. A more complete description of the FF method is described below.

The collection of real events used to derive the FF must pass either the single electron or single muon triggers. The selected events must contain either exactly one tight or exactly one fakeable lepton and at least one jet which is separated from the lepton by $\Delta R > 0.7$. This collection, called the measurement region (MR), is enriched in multijet events, which contain only non-prompt leptons. The backgrounds for this measurement contains prompt leptons are W+jets, Z+jets, $t\bar{t}$, and diboson processes. In each bin, the number of multijet events contributing to

the total event yield in the MR is determined using a maximum-likelihood fit of a modified invariant transverse mass of the lepton:

$$m_T^{\text{fix}} = \sqrt{2 \cdot p_{T\ell}^{\text{fix}} E_T^{\text{miss}} (1 - \cos \Delta\phi)}. \quad (9)$$

Here $\Delta\phi$ refers to the transverse angle between the lepton momentum and the E_T^{miss} vector, while $p_{T\ell}^{\text{fix}}$ is a constant value of 35 GeV. Using a constant value for the lepton momentum has been found to give a more accurate result for the fake rate than using the magnitude of the momentum [37]. The templates used in the fit are explicitly validated in a dedicated $2\ell\text{ss}$ control region seen in Figure 5. The dependence of m_T^{fix} on E_T^{miss} produces a rapidly-falling distribution for multijet events, whose E_T^{miss} is mainly due to resolution effects and is quite small. In contrast, the largest background for this measurement is W +jets, whose m_T^{fix} distribution peaks at $m_W = 80$ GeV and is very broad.

The FFs are derived in bins of the lepton $|\eta|$ and cone- p_T rather than the standard “reco- p_T ” as they will be applied to the AR which contains fakeable leptons. The condition on the cone- p_T is applied to reduce biases in the measurement of the lepton mis-id rates. The issue is that the p_T thresholds that are applied on trigger level impose an implicit condition on the isolation of the lepton, because the lepton mis-id rates are parametrized by the cone- p_T , and not the reco- p_T , of the lepton. Such a 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.

In each bin, the fits determine the number of fakeable lepton events and tight lepton events simultaneously. These numbers are used to determine the FFs seen in Figure 3 individually known as f_i . However, the ratio of tight to fakeable leptons F_i is more directly used to determine the fake rate:

$$f_i = \frac{N_{\text{tight}}}{N_{\text{tight}} + N_{\text{fakeable}}} \quad (10)$$

$$F_i = \frac{f_i}{1 - f_i} = \frac{N_{\text{tight}}}{N_{\text{fakeable}}}. \quad (11)$$

The fake rate in the signal region is calculated using the following equations which correspond to the $2\ell\text{ss}$, 3ℓ , and 4ℓ categories:

$$N_{\text{II}}^{\text{fake}} = \sum_{\text{fl}} F_1 + \sum_{\text{lf}} F_2 - \sum_{\text{ff}} F_1 F_2 \quad (12)$$

$$N_{\text{III}}^{\text{fake}} = \sum_{\text{fll}} F_1 + \sum_{\text{lfl}} F_2 + \sum_{\text{llf}} F_3 - \sum_{\text{ffl}} F_1 F_2 - \sum_{\text{flf}} F_1 F_3 - \sum_{\text{lff}} F_2 F_3 + \sum_{\text{fff}} F_1 F_2 F_3 \quad (13)$$

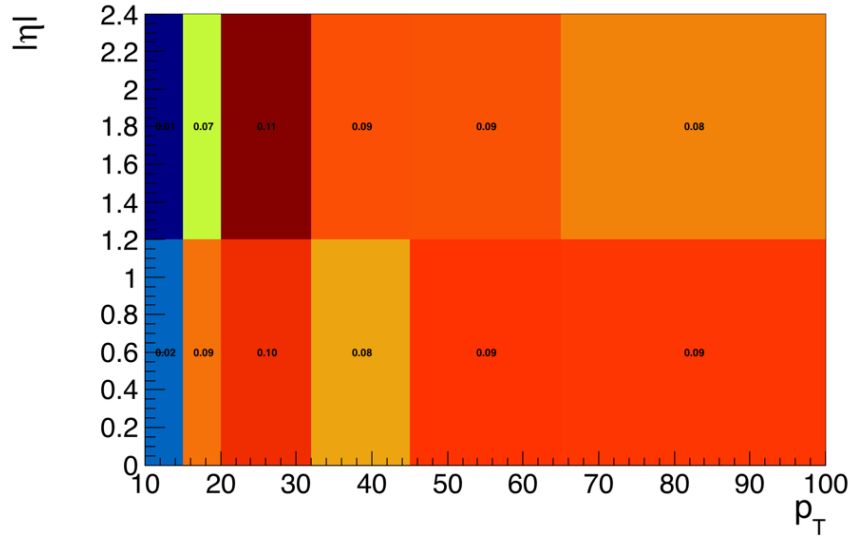
$$N_{\text{IIII}}^{\text{fake}} = \sum_{\text{flll}} F_1 + \sum_{\text{lflf}} F_2 + \sum_{\text{llfl}} F_3 + \sum_{\text{lllf}} F_4 - \sum_{\text{ffll}} F_1 F_2 - \sum_{\text{flfl}} F_1 F_3 - \sum_{\text{flff}} F_1 F_4 - \sum_{\text{lffl}} F_2 F_3 - \sum_{\text{llff}} F_2 F_4 - \sum_{\text{llff}} F_3 F_4 + \sum_{\text{fffl}} F_1 F_2 F_3 + \sum_{\text{fflf}} F_1 F_2 F_4 + \sum_{\text{flff}} F_1 F_3 F_4 + \sum_{\text{lfff}} F_2 F_3 F_4 - \sum_{\text{ffff}} F_1 F_2 F_3 F_4 \quad (14)$$

The labels of the sums indicate which leptons are fakeable and which ones are tight in descending order of cone- p_T . For example, \sum_{lfl} indicates events with three leptons where the first and third-highest cone- p_T leptons pass the tight selection and the second-highest cone- p_T lepton passes the fakeable selection. The fake ratio factors F_i are present for each fakeable lepton in the event, and their indices reference which leptons in the cone- p_T ordering the fake ratio corresponds to. The signs of the terms alternate based on the number of fakeable leptons in the events to eliminate double-counting.

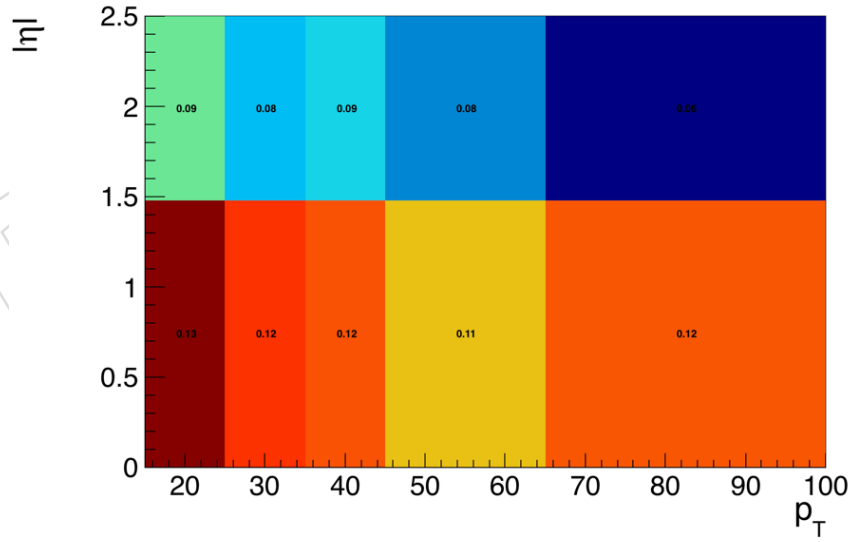
One danger of using this method is the difference in implicit isolation requirements in the MR and AR. The MR only contains one lepton per event meaning every lepton passes the trigger requirements on the reco- p_T . However, in the AR not every lepton must pass these requirements. The FFs are determined in bins of cone- p_T , and since the cone- p_T is generally higher than the reco- p_T , this results in a bias. While the cone- p_T difference does not directly affect the fake rate estimate, it does effect the calculation of the prompt- e and prompt- μ BDT used in the tight and fakeable selections. If not counteracted, this would cause an overestimate on the fake rate. To remove the bias in the cone- p_T , we require it to exceed roughly double the trigger reco- p_T requirement so that with the isolation requirement $I_\ell < 0.4 \cdot \text{reco-}p_T$, all fakeable lepton should inherently pass the trigger reco- p_T requirement. The trigger requirements on the variables $\sigma_{i\eta i\eta}$, H/E , and $1/E - 1/p$ are applied to the fakeable electron selection as well to reduce a similar but smaller bias on the electron isolation (see Table 11).

The photon conversion background has some contribution to this fake rate. The contribution is determined using a MC sample of multijet events. The numbers N_{tight} and N_{fakeable} are scaled by the fraction of events containing an electron that matches within $\Delta R < 0.3$ to a prompt photon. A scaling factor is determined for each bin in cone- p_T and $|\eta|$ to eliminate this double-counting. This lowers the electron FF by 10-20%.

The fake rate determined using multijet events differs somewhat from the fake rate caused by non-prompt leptons in $t\bar{t}$ events. Since much of our SR contains events with a $t\bar{t}$ pair, we must correct for this effect as well. The fake rates have been roughly matched by tuning the DeepCSV(b) requirement in the fakeable selection. The remaining difference is accounted for by the uncertainty on the FF.



(a)



(b)

Figure 3: Maps of the cone- p_T and $|\eta|$ -dependent FF f_i for muons (a) and electrons (b). The FFs 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 $t\bar{t} \rightarrow bW^+ \bar{b}W^- \rightarrow b\ell^+ \nu \bar{b}\ell^- \bar{\nu}$, 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 [27] 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 bins (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 18.

	$15 \leq p_T < 25 \text{ GeV}$	$25 \leq p_T < 50 \text{ GeV}$	$50 \text{ GeV} \leq p_T$
$0 \leq \eta < 1.479$	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

Table 18: 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.

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 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% [27]. 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 [44]. 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.

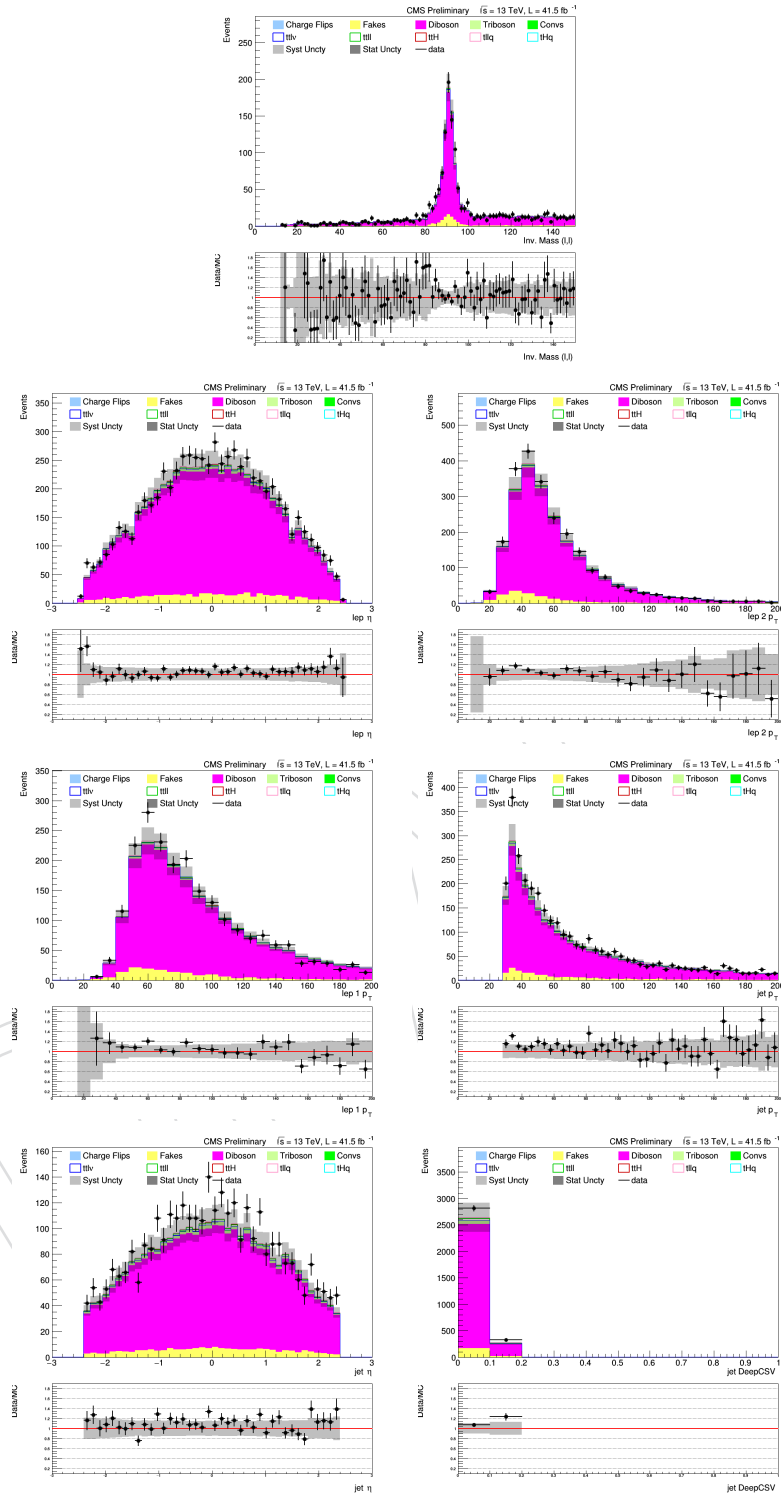


Figure 4: Control region plots (3 leptons). These plots show the good agreement between data and MC after all corrections have been applied, for a selection enriched in Diboson events (the dominant background in the 3 lepton signal regions). In this control region we require exactly 3 tight leptons and 1 or more jets, with exactly 0 jets 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 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 for 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 jet multiplicity 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 ctW^2 would have a string pair $\langle \text{“ctW”, “ctW”} \rangle$. Linear terms are of the form $\langle \text{“sm”, “ctW”} \rangle$ or $\langle \text{“ctW”, “sm”} \rangle$. The constant term, or in other words the standard model yield, is given by the pair $\langle \text{“sm”, “sm”} \rangle$. 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 page¹. The first is a fork of the CombineHarvester package [54]. The scripts there build the single datacard needed for use in the Higgs PAG “Combine” tool [55]. The other “cms-govner” repository named “EFTFit” provides the scripts and modules necessary to run combine with our specific physics model.

Fitting the event yields is done through the open source HiggsAnalysis-CombinedLimit software². This is based on the RooStats [56] framework developed for use by the CMS and ATLAS groups. A suite of statistical methods is provided by this software. The data yields are provided to the software as an observed yield in each category. The Monte Carlo yields of each process in each category are provided to the software along with the possible fluctuations caused by each systematic uncertainty. These are taken at their Standard Model values (no EFT effects present). For the fit of the Wilson coefficients, these expected yields are morphed during the

¹<https://github.com/cms-govner>

²<https://github.com/cms-analysis/HiggsAnalysis-CombinedLimit>

fit according to the parameterizations described in Section 2 to determine the best fit values of the WC. The robustness of the event reweighting provides an accurate description of each process's dependence on the EFT operators.

The fit of the Wilson coefficients is done through a maximization of the likelihood, or rather a minimization of the negative log-likelihood (NLL). Due to the complexity of the fit, a global NLL minimization varying all Wilson coefficients simultaneously is liable to find a local minimum rather than the global minimum. To avoid this, fits are performed while scanning over a large range of Wilson coefficient values. The ranges of the Wilson coefficients considered matches the narrowest ranges used in the determination of their figure of merit in Table 1. These ranges scale the inclusive $t\bar{t}$ cross section by no more than 15%, and scale no signal cross section to more than five times the Standard Model value. These scans are performed by stepping through values of one or two WC while the rest are profiled (or are fixed to the SM value depending on the type of scan). The best fit value of each WC is reported as its value at the global NLL minimum in the one-dimensional scan over that WC. The other points in the scan are described by the difference between each point's NLL value and the value at the global minimum (ΔNLL). Sixteen one-dimensional scans are required, then, to fully determine the best fit values for all Wilson coefficients.

Confidence intervals are determined graphically by analyzing the one-dimensional WC scans. For these scans in one parameter, the 1σ (68.27%) and 2σ (95.45%) confidence intervals include all points whose $2\Delta\text{NLL}$ is less than one and four, respectively. The intersection of the scans with one and four (linearly interpolated) define the endpoints of the certainty intervals.

In certain cases, the 1D scans are not smooth distributions as they should be. This is caused by a technical limitation in the fit software. The discontinuities occur when there are multiple NLL minima in the multidimensional space that the software does not elegantly switch between to find the true 1D NLL minima. To correct for this flaw, we find the confounding variables which contain the problematic minima. A multidimensional scan in each parameter plus the confounding parameters is performed. The minimum NLL values are extracted in slices of these scans to extract effective 1D scans. These new 1D scans are then used in place of the originals to extract the best fit values and uncertainty intervals.

Additionally, a fit is performed to determine the Standard Model signal strengths without regard to EFT effects. This fit is performed on both the true data and Asimov dataset as a consistency check for our analysis strategy. In the former, a comparison is made to the fit using signal samples provided directly by CMS as an additional validation of our signal samples. In these fits, each of the five signal processes are scaled linearly by separate signal strengths μ_{signal} . Due to the significant insensitivity of this analysis to tHq , this process is instead scaled by μ_{tH} . For these variables, a value of '1' corresponds to the standard model prediction. Fits are performed both simultaneously for the four signals as well as for each signal individually while the other three are fixed to $\mu_{\text{signal}} = 1$. Ideally, the fitted signal strengths for the asimov dataset should be $\mu_{\text{signal}} = 1$ with errors similar to or somewhat larger than results reported by other CMS analyses. For the fits to real data, we expect some variation due to excesses and deficits reported by the dedicated analyses for these processes.

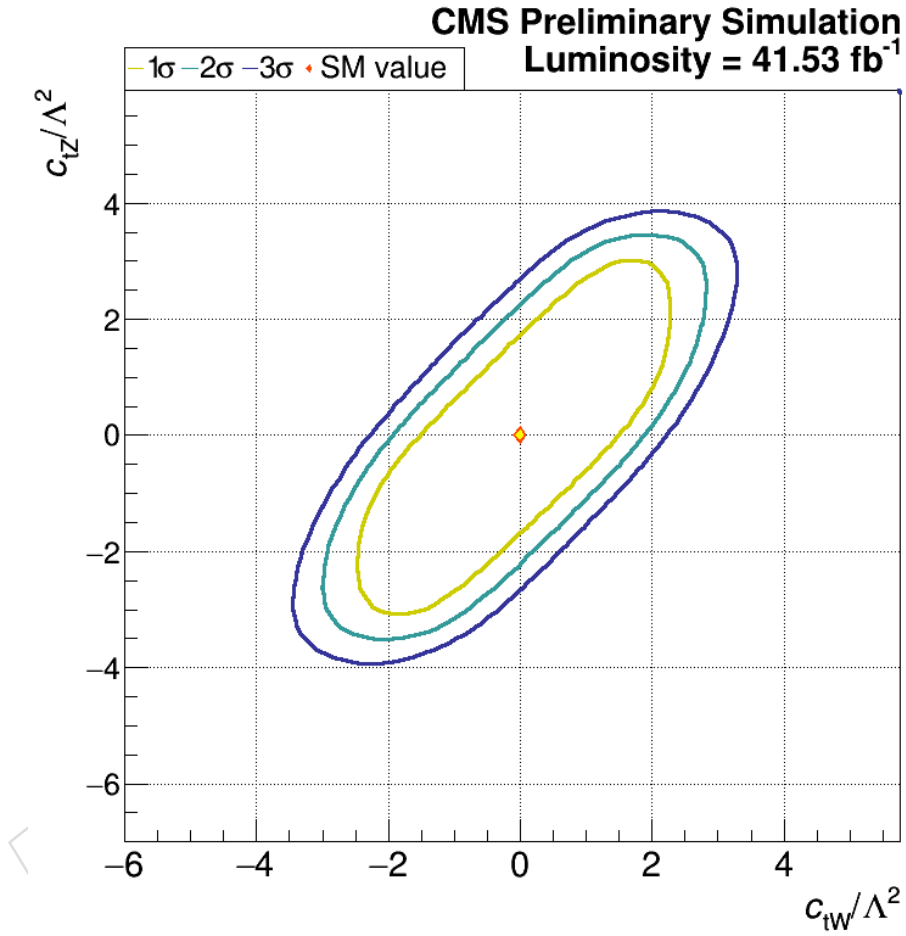


Figure 6: An examination of a 2D confidence-level contour. The axes are values of the Wilson coefficients of interest. The solid yellow, teal, and blue lines correspond to the 1σ , 2σ , and 3σ contours, respectively. The diamond marker marks the origin, that is, the SM-expected values of the Wilson coefficients. The dashed lines are useful for determining the 1D limits. The minimum/maximum values along an axis of the dashed yellow, teal, and blue contours are the 1D 1σ , 2σ , and 3σ limits for that Wilson coefficient. The 1D limits are also shown in dedicated 1D scans as well as a summary table and plot.

9 Systematic uncertainties

There are three types of systematic effects considered in this analysis: those that affect only the event counts (“rates”) of signal or background processes, those that affect only the shapes of the jet multiplicity distributions for signal or background processes, and those that affect both the rate and the shape. Since this analysis is performed as a counting experiment, here “shape” effects refer to event migrations between categories that only differ by the number of jets required. The overall normalization of each jet multiplicity distribution remains unchanged. The effects of “rate+shape” systematics are applied simultaneously so that they are completely correlated. Unless otherwise noted, each of the uncertainties listed here apply equally to signal and background and are treated as 100% correlated across all processes and categories. When stated, the size of an uncertainty refers to the typical 1σ uncertainty on each category’s yield. Below is a list of the systematic effects considered for this analysis.

Luminosity: Rate systematic. The uncertainty on the luminosity estimate is 2.3%. [6].

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 [57] 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% isolation 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 [58]. 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 17. This effect ranges from 2–5%.

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, which corresponds to a 1σ fluctuation. This effect typically ranges from <1% to 3%.

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 [59]. The events are then reanalyzed, including reapplying the jet-based selection and recalculating all kinematic quantities. After this, the direction of the correlations were often confused. This was due to the net number of jets entering or leaving a bin being small while the error on the shifts being relatively large. To correct this, the systematic was transformed in to a symmetric nuisance for each bin using the average error. From here, the correlations were set manually by analyzing a toy model for how JES shifts should behave. Bins below the peak in the njets distribution have a positive theta value correpond to a decrease in the bin yield. Bins above the peak in the njets distribution have a positive theta value correpond to an increase in the bin yield. The peak bins themselves follow the direction of the UP fluctuation. This effect typically ranges from 5–10%.

b-tag scale factors: Rate + shape systematics. We are using the differential b-tagging scale factors described in [43]. 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 SFs, 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 by $\pm 1\sigma$ 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 [60]. This effect ranges from 2–6% for the heavy flavor purity and <1% to 2% for the light flavor purity.

sample statistics: Finally, the impact of statistical uncertainties associated with the size of the samples used to derive the SFs 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 SFs, 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. Each of these effects ranges from < 1% to 2%.

For charm-jet 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-jet scale factors will be treated independently with respect to all the uncertainties for the heavy-flavor and light-flavor scale factors. Their effects range from 1–7%. More details on these uncertainties are given in Ref. [60].

Uncertainty on the Lepton Fake Rate estimate: Rate + shape systematic. We estimate the uncertainty on the lepton fake rate following the procedure described in [48]. 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 ≈ 25 –30%, depending on the jet multiplicity bin (see Figs. 7 and 8). 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 the jet multiplicity bin, and amounts to ≈ 10 –30%. 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 19 [61]. 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 long 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.

Process	pdf				QCD Scale						
	gg_ttH	gg	q \bar{q}	qg_tHq	t \bar{t} V	t \bar{t} γ	ttH	tHq	V	VV	VVV
t \bar{t} H	3.6%						-9.2%/+5.8%				
t \bar{t} lq			4%						1%		
t \bar{t} l ν			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} γ +jets		5%				10%					

Table 19: 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.

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 NNPDF31_nnlo_hessian.pdfas PDF parameterization for the 5f samples and NNPDF31_nnlo.as_0118_nf_4 for the 4f samples. 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 [62, 63].

Renormalization and factorization scale: Shape systematic. Uncertainties due to the renormalization scale μ_R and the factorization scale μ_F in the 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. From here, we calculated the uncertainty envelope from these two systematics. The up and down fluctuations are taken to be the maximum of the up and down fluctuations of the factorization uncertainty, renormalization uncertainty, and their sum. This is accommodated via weights obtained directly from the generator information. Since the normalization uncertainties of the matrix-element generator are covered by the cross section uncertainties (listed in Tab. 19), only the shape variation of discriminant distributions is considered, i.e. they are scaled to retain the overall normalization. After the size of each effect is determined, the two are combined into a single uncertainty envelope. The bounds of the envelope are determined by taking the maximum of the factorization scale errors, the renormalization scale errors, and their sum. The total effect in each bin is 1–5%. Additional checks to determine the impact of using different EFT starting points to calculate the uncertainties is also considered and covered in Appendix C. We found little to no effect due to picking different starting points, within the precision available in this analysis.

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 $\sqrt{2}$. A dedicated SM sample (produced without EFT effects) is used to determine this systematic; the values obtained through this study are then applied to the full analysis samples.

Matching uncertainty: Shape systematic. This uncertainty only applies to $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ since matching is only performed for processes that include an extra parton. Determined by varying the matching scale to valid values above and below the nominal matching scale, the uncertainty is computed bin by bin. A dedicated SM sample (produced without any EFT effects) was used to perform this study.

Missing parton uncertainty: Shape systematic. Since an extra final-state parton was not able to be included in the privately produced LO single top processes ($t\bar{t}l\bar{q}$ and tHq), they are not expected to be as accurate as the privately produced LO $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ samples. Comparing our privately produced LO $t\bar{t}l\bar{q}$ sample to the centrally produced NLO tZq sample (as described in Section 3.2), there existed a discrepancy that was not covered by our existing systematic uncertainties. We therefore introduced a new systematic uncertainty specifically for the $t\bar{t}l\bar{q}$ sample to cover this disagreement. The same systematic is applied to tHq .

Table 20 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.

DRAFT

Table 20: 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 jet multiplicity 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	Applied to signal and all MC backgrounds
Lepton ID/Iso	rate+shape	Applied to signal and all MC backgrounds
Trigger efficiency	rate+shape	Applied to signal and all MC backgrounds
Pileup	rate+shape	Applied to signal and all MC backgrounds
JES	rate+shape	Applied to signal and all MC backgrounds
b-Tag HF purity	rate+shape	Applied to signal and all MC backgrounds
b-Tag HF stats (linear)	rate+shape	Applied to signal and all MC backgrounds
b-Tag HF stats (quadratic)	rate+shape	Applied to signal and all MC backgrounds
b-Tag LF purity	rate+shape	Applied to signal and all MC backgrounds
b-Tag LF stats (linear)	rate+shape	Applied to signal and all MC backgrounds
b-Tag LF stats (quadratic)	rate+shape	Applied to signal and all MC backgrounds
b-Tag Charm (linear)	rate+shape	Applied to signal and all MC backgrounds
b-Tag Charm (quadratic)	rate+shape	Applied to signal and all MC backgrounds
Fakes FF measurement	rate+shape	Related to FF measurement for the fakes background.
Fakes AR stats	rate+shape	Due to statistics in the AR for the fakes background. Uncorrelated across categories.
Charge Flip	rate	Uncertainty on the 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 $t\bar{t}\gamma$ +jets prediction
QCD Scale ($t\bar{t}V$)	rate	Scale uncertainty for NLO $t\bar{t}l\nu$ and $t\bar{t}l\bar{l}$ 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 $t\bar{t}H$ ($t\bar{t}l\bar{l}$, $t\bar{t}\gamma$ +jets)
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 ($t\bar{t}lq$, $t\bar{t}l\nu$, Diboson, Triboson)
pdf (qg , tHq)	rate	Pdf uncertainty for qg -initiated processes (tHq)
PDF 4f	shape	Applied to signal and all MC backgrounds that use 4f PDF
PDF 5f	shape	Applied to signal and all MC backgrounds that use 5f PDF
Q^2 Scale Renorm. + Fact.	shape	Renormalization and factorization scale uncertainties: Applied to signal and all MC backgrounds. Uncorrelated between all processes
Parton Shower	shape	Parton-shower scale uncertainties: Applied to signal and all MC backgrounds. Uncorrelated between all processes
Matching Uncertainty	shape	Parton-shower matching uncertainty: Applied to signal samples produced with an extra parton ($t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$). Uncorrelated between all processes
Missing parton uncertainty	shape	Introduced to account for the fact that an extra parton could not be included in the privately produced single top samples: Applied only to $t\bar{t}lq$ and tHq .

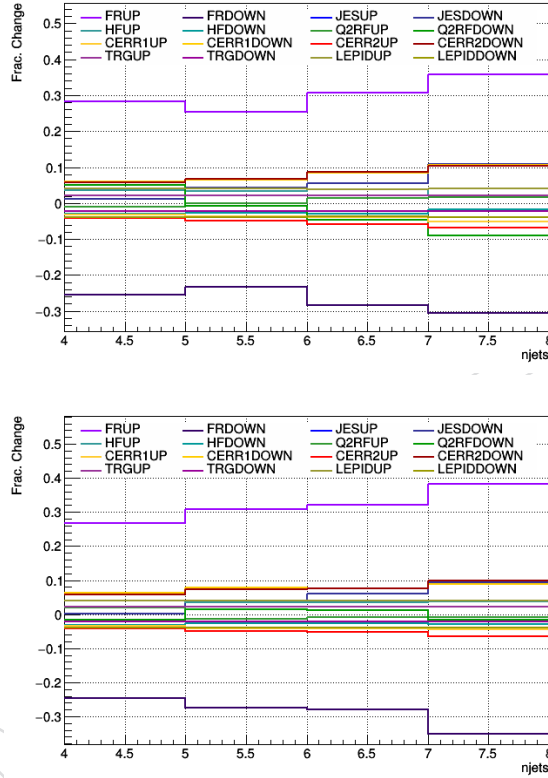


Figure 7: Fractional change in the expected yield w.r.t. the nominal yields for signal and background events, as a function of n_{jets} , as different systematic variations are considered. The systematics are a subset chosen to highlight those that have a relatively large effect on the shape of the n_{jets} 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 $2\ell_{\text{ss}}$ categories are shown. From left to right, the plots show events in the $2\ell_{\text{ss}}$ ("p") and $2\ell_{\text{ss}}$ ("m") categories.

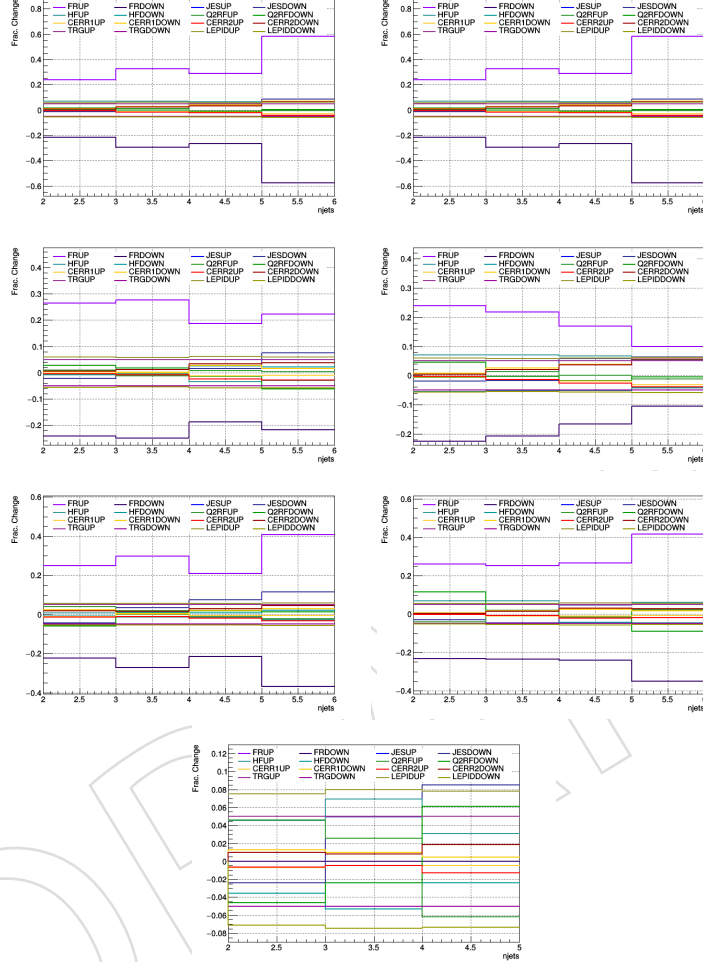


Figure 8: Fractional change in the expected yield w.r.t. the nominal yields for signal and background events, as a function of n_{jets} , as different systematic variations are considered. The systematics are a subset chosen to highlight those that have a relatively large effect on the shape of the n_{jets} 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). Due to this analysis' insensitivity to $t\bar{t}Hq$, we scale the $t\bar{t}Hq$ process by the $t\bar{t}H$ signal strength. The fit was carried out simultaneously in each of the four signal strengths, as well as individually with other processes fixed to $\mu = 1$. The results of both checks are shown in Table 21 for the Asimov dataset, and Table 22 for the 2017 data, 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
$t\bar{t}l\bar{l}$	1.00 ± 0.21	1.00 ± 0.18	$1.00^{+0.15}_{-0.13}$ (TOP-18-009)
$t\bar{t}l\nu$	1.00 ± 0.33	1.00 ± 0.24	1.00 ± 0.22 (TOP-17-005)†
$t\bar{t}H$	1.00 ± 0.58	1.00 ± 0.45	$1.00^{+0.39}_{-0.35}$ (HIG-18-019)
$t\bar{t}lq$	1.00 ± 0.49	1.00 ± 0.44	$1.00^{+0.23}_{-0.21}$ (TOP-18-008)

Table 21: Best fit values of the different SM signal strengths for each of this analysis' signal processes. Here the $t\bar{t}Hq$ process has been scaled by $\mu_{t\bar{t}H}$ instead due to this analysis's insensitivity to $\mu_{t\bar{t}Hq}$. The second column contains the results from a simultaneous fit of the four signal strengths. The third column contains the results from a fit on each signal strength individually (others fixed 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 the one 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. 9). 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 $t\bar{t}Zq$ 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 $t\bar{t}H$ 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. When comparing to the dedicated $t\bar{t}W$ and $t\bar{t}Z$ 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 $t\bar{t}W$, 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 $t\bar{t}W$ and $t\bar{t}Z$ make up a large fraction of our yield, with $t\bar{t}Z$ 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 in the dedicated analyses therefore do not have as large of an effect in this

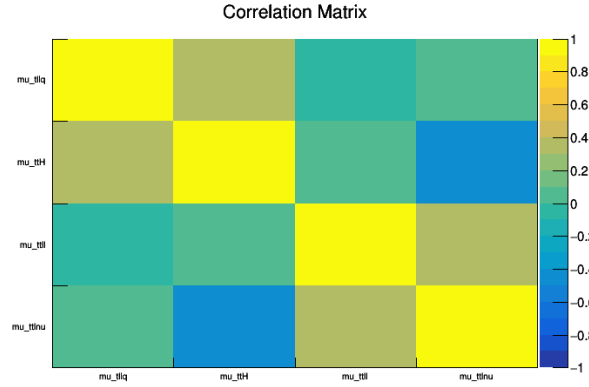


Figure 9: Correlation Matrix of Standard Model signal strengths. This is a result of a simultaneous fit of the four signal strengths on the SM dataset. The processes $t\bar{t}l\nu$ and $t\bar{t}l\bar{l}$ are moderately correlated, as are $t\bar{t}H$ and $t\bar{t}lq$ and $t\bar{t}lq$ and $t\bar{t}l\nu$. 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.

case.

A fit of the aforementioned SM signal strenghts to the data reveals the follwing results:

SM Signal Strength Results (Unblinded Dataset)			
Signal Process	Best Fit μ (Private)	Best Fit μ (Central)	Dedicated Analysis
$t\bar{t}l\bar{l}$	$1.05^{+0.25}_{-0.19}$	$1.20^{+0.27}_{-0.21}$	$1.28^{+0.14}_{-0.13}$ (TOP-18-009)
$t\bar{t}l\nu$	$0.79^{+0.37}_{-0.32}$	$0.75^{+0.36}_{-0.32}$	$1.23^{+0.31}_{-0.28}$ (TOP-17-005)+
$t\bar{t}H$	$2.52^{+0.85}_{-0.70}$	$2.57^{+0.87}_{-0.70}$	$0.75^{+0.46}_{-0.43}$ (HIG-18-019)
$t\bar{t}lq$	$1.11^{+0.58}_{-0.56}$	$0.46^{+0.53}_{-0.47}$	$1.01^{+0.23}_{-0.21}$ (TOP-18-008)

Table 22: Best fit values of the different signal strengths for each of this analysis' signal processes. Here the $t\bar{t}H$ signal strength also controls the $t\bar{t}Hq$ processes due to this analysis' relative insensitivity to $t\bar{t}Hq$. The second column contains the results from a simultaneous fit of the four signal strengths using our private signal samples. The third column contains the results from a simultaneous fit of the four signal strengths using central samples for $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}H$, and $t\bar{t}Zq$. There is no equivalent $t\bar{t}Hq$ central sample, so here our private $t\bar{t}Hq$ sample is used again. 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 the one marked with a †, which use the 2016 dataset.

Figures 10–13 show the pull and impacts for each signal strength.

10.2 EFT Signals

10.2.1 Standard Model Expectation

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

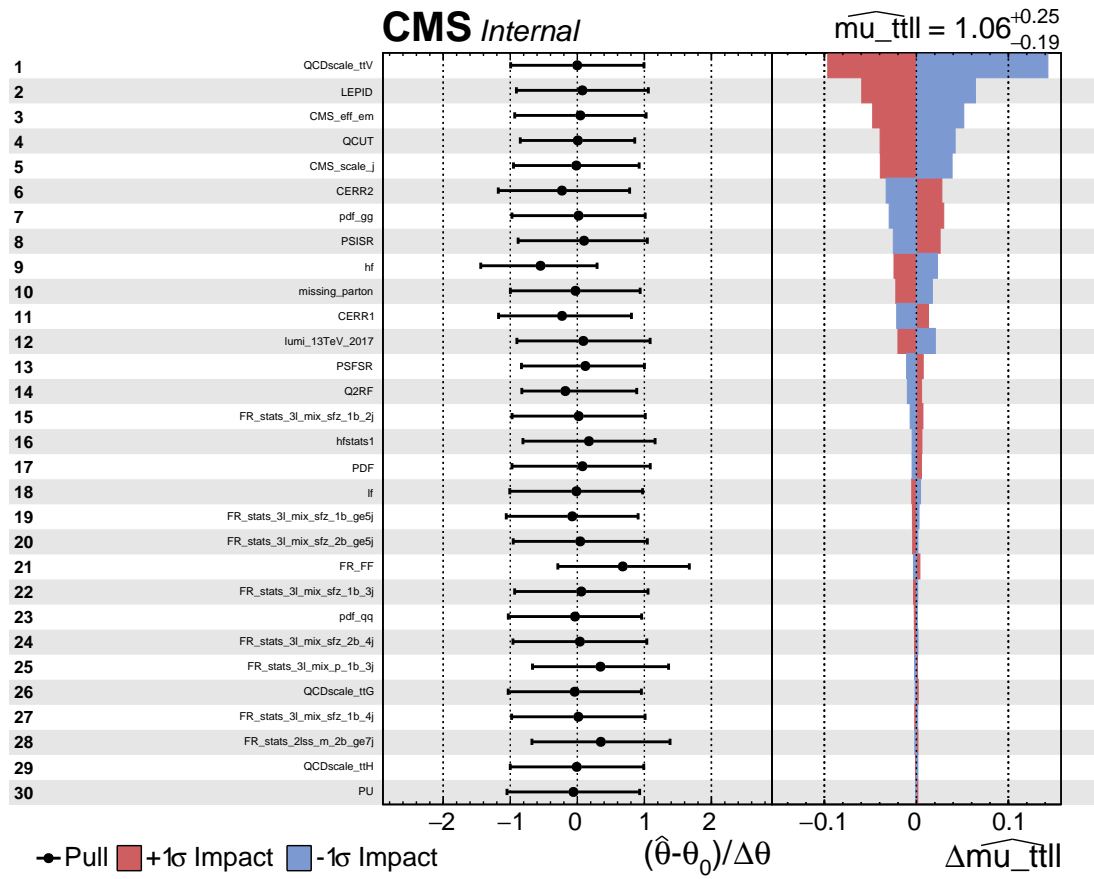


Figure 10: The impacts and pulls of the largest systematics for $t\bar{t}l\bar{l}$, ranked by impact.

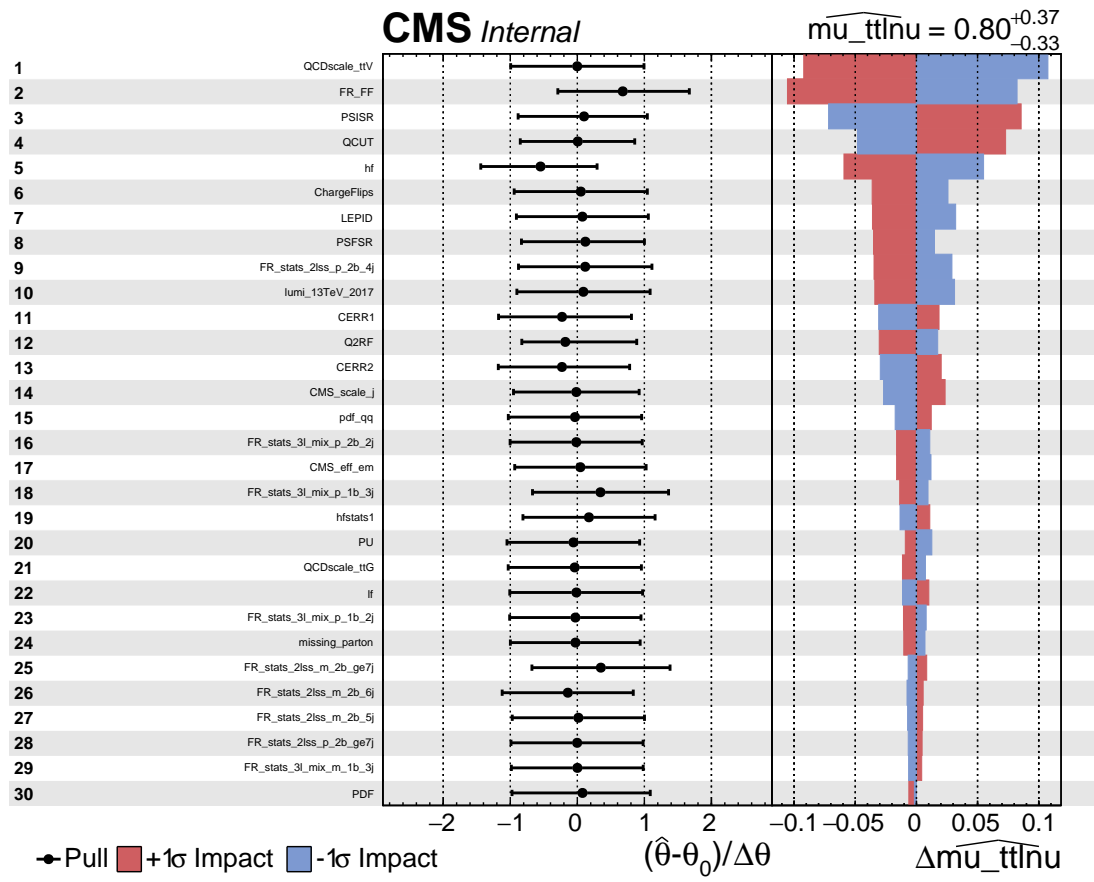


Figure 11: The impacts and pulls of the largest systematics for ttlnu , ranked by impact.

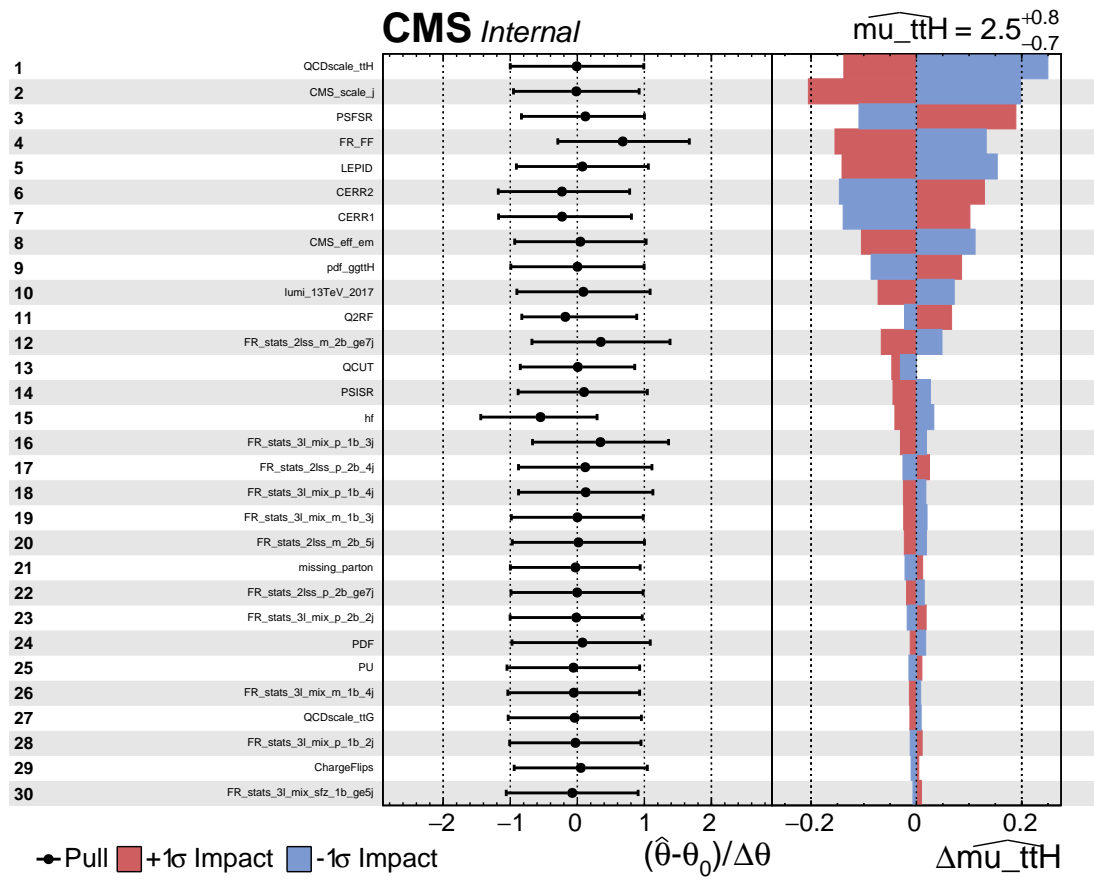


Figure 12: The impacts and pulls of the largest systematics for $t\bar{t}H$, ranked by impact.

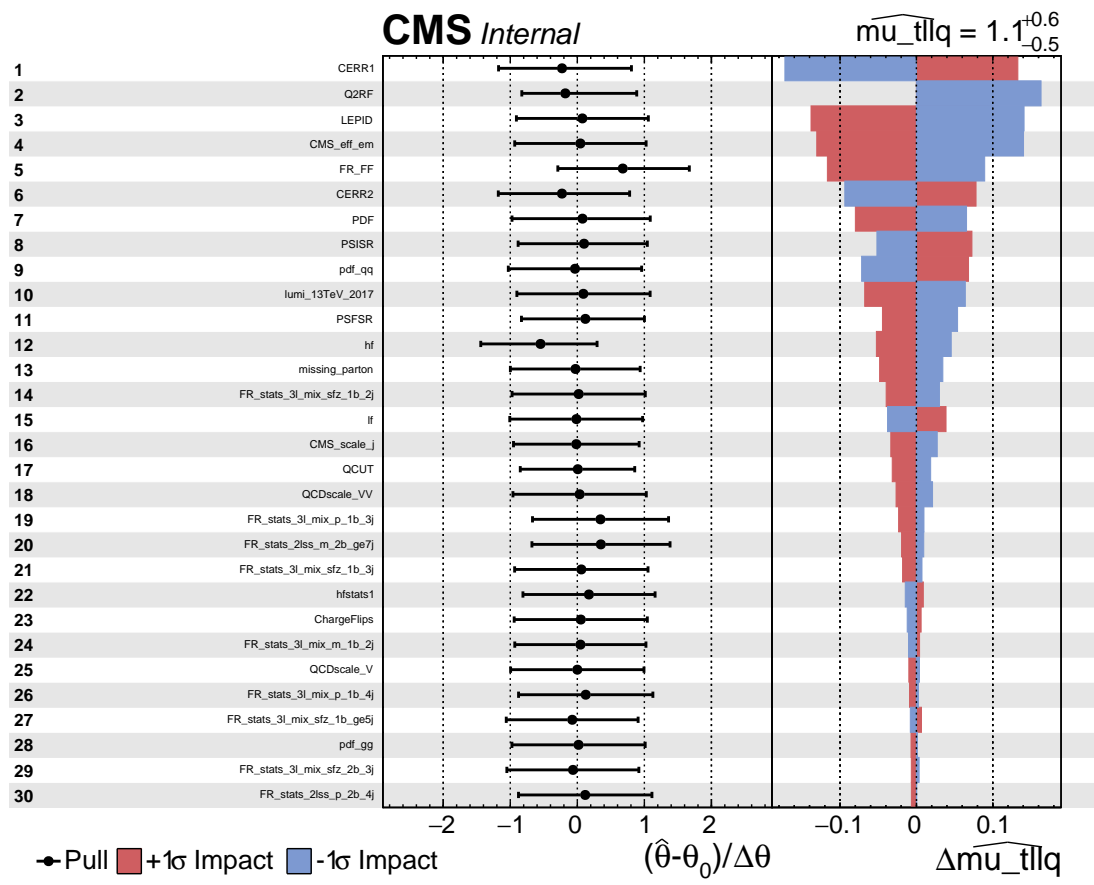


Figure 13: The impacts and pulls of the largest systematics for $t\bar{l}lq$, ranked by impact.

as a 16-dimensional hypersurface, so we also present a series of 2D projections of this hypersurface along with 1D projections. All results are obtained using the the HiggsAnalysis-CombinedLimit tool as described in Section 8.

Table 23 shows the Standard Model-expected 95% confidence intervals. These results are displayed graphically in Figure 14, where 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 fixed to their Standard Model values of 0. Wilson Coefficients with multiple, distinct uncertainty intervals have multiple intervals listed.

Summary Table		
Wilson Coefficient	2σ Interval (others profiled)	2σ Interval (others fixed to SM)
c_{tW} / Λ^2	[-2.75, 2.58]	[-1.41, 1.29]
c_{tZ} / Λ^2	[-3.31, 3.29]	[-1.57, 1.62]
$c_{t\varphi} / \Lambda^2$	[-10.14, 39.02]	[-6.86, 19.65] and [11.39, 37.12]
$c_{\varphi Q}^- / \Lambda^2$	[-8.04, 26.49]	[-3.78, 3.20]
c_{tG} / Λ^2	[-1.02, 0.92]	[-0.92, 0.44]
c_{bW} / Λ^2	[-4.06, 4.03]	[-3.36, 3.33]
$c_{\varphi Q}^3 / \Lambda^2$	[-8.12, 3.60]	[-7.13, 3.10]
$c_{\varphi tb} / \Lambda^2$	[-14.74, 14.57]	[-12.26, 12.09]
$c_{\varphi t} / \Lambda^2$	[-22.34, 13.86]	[-5.52, 4.90]
$c_{Q\ell}^{3(\ell)} / \Lambda^2$	[-7.93, 7.28]	[-7.30, 6.91]
$c_{Q\ell}^{-(\ell)} / \Lambda^2$	[-4.30, 5.39]	[-3.56, 4.58]
$c_{Qe}^{(\ell)} / \Lambda^2$	[-4.76, 4.84]	[-3.97, 4.15]
$c_{t\ell}^{(\ell)} / \Lambda^2$	[-4.65, 5.29]	[-3.88, 4.29]
$c_{te}^{(\ell)} / \Lambda^2$	[-4.69, 5.27]	[-3.73, 4.56]
$c_t^{S(\ell)} / \Lambda^2$	[-6.84, 6.84]	[-5.79, 5.79]
$c_t^{T(\ell)} / \Lambda^2$	[-0.90, 0.90]	[-0.77, 0.77]

Table 23: Standard Model-expected 2σ uncertainty intervals. The intervals are found by scanning over a single Wilson coefficient while either treating the other 15 as unconstrained nuisance parameters (profiled), or fixing the other 15 to the SM value of zero.

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 profiling the other WCs, or by fixing 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 Standard Model values of the coefficients. These are split into two sets based on their treatment of the other 15 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 profiled, the contours represent the projection of the 16D hyper-surface onto this 2D plane.

10.2.2 Fits to Data

Table 24 shows the 95% confidence intervals for each Wilson Coefficient. These results are displayed graphically in Figure 17, where again 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 fixed to their Standard Model values of 0. Wilson Coefficients with multiple, distinct uncertainty intervals have multiple intervals listed.

Summary Table		
Wilson Coefficient	2σ Interval (others profiled)	2σ Interval (others fixed to SM)
c_{tW} / Λ^2	[-3.08, 2.87]	[-2.15, 0.21] and [-0.29, 1.96]
c_{tZ} / Λ^2	[-3.32, 3.15]	[-2.14, 2.19]
$c_{t\varphi} / \Lambda^2$	[-16.98, 44.26]	[-14.12, 32.30] and [-1.48, 44.48]
$c_{\varphi Q}^- / \Lambda^2$	[-7.59, 21.65]	[-3.45, 3.33]
c_{tG} / Λ^2	[-1.38, 1.18]	[-1.26, 0.08] and [-0.69, 0.79]
c_{bW} / Λ^2	[-4.95, 4.95]	[-4.12, 4.09]
$c_{\varphi Q}^3 / \Lambda^2$	[-7.37, 3.48]	[-7.21, 2.25]
$c_{\varphi tb} / \Lambda^2$	[-12.72, 12.63]	[-9.87, 9.67]
$c_{\varphi t} / \Lambda^2$	[-18.62, 12.31]	[-20.91, -6.52] and [-14.10, 4.24]
$c_{Q\ell}^{3(\ell)} / \Lambda^2$	[-9.67, 8.97]	[-9.91, 9.50]
$c_{Q\ell}^{- (\ell)} / \Lambda^2$	[-4.02, 4.99]	[-4.76, 5.83]
$c_{Qe}^{(\ell)} / \Lambda^2$	[-4.38, 4.59]	[-5.20, 5.36]
$c_{t\ell}^{(\ell)} / \Lambda^2$	[-4.29, 4.82]	[-5.15, 5.51]
$c_{te}^{(\ell)} / \Lambda^2$	[-4.24, 4.86]	[-4.97, 5.80]
$c_t^{S(\ell)} / \Lambda^2$	[-6.52, 6.52]	[-7.70, 7.70]
$c_t^{T(\ell)} / \Lambda^2$	[-0.84, 0.84]	[-1.01, 1.01]

Table 24: Wilson coefficient 2σ uncertainty intervals. The intervals are found by scanning over a single Wilson coefficient while either treating the other 15 as unconstrained nuisance parameters (profiled), or fixing the other 15 to the SM value of zero.

Figure 18 shows the 1D likelihood scans of each of the 16 Wilson coefficients using the 2017 CMS dataset. These fits were made by either treating the other coefficients as profiled nuisance parameters or by fixing 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 19 are the 2D likelihood scans of pairs of coefficients for fits to the 2017 CMS dataset. The plots show the 1σ , 2σ , and 3σ contours in this cross section of the phase space. Also shown are markers for the Standard Model values of the coefficients. These are split into two sets based on their treatment of the other 15 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 profiled, the contours represent the projection of the 16D hyper-surface onto this 2D plane.

10.2.3 Other Results and Cross Checks

The pulls and impacts of the top-30 systematics for each Wilson coefficients are show in Figures 20–34. The number of events selected in different categories is compared to the postfit

1054 expected contributions of the signal processes and of the different background processes in
1055 Figures 35 and 36. Figures 37 compare the pre and postfit expected yields, as well as the ex-
1056 pected yields with one WC set to its $\pm 2\sigma$ limits.

DRAFT

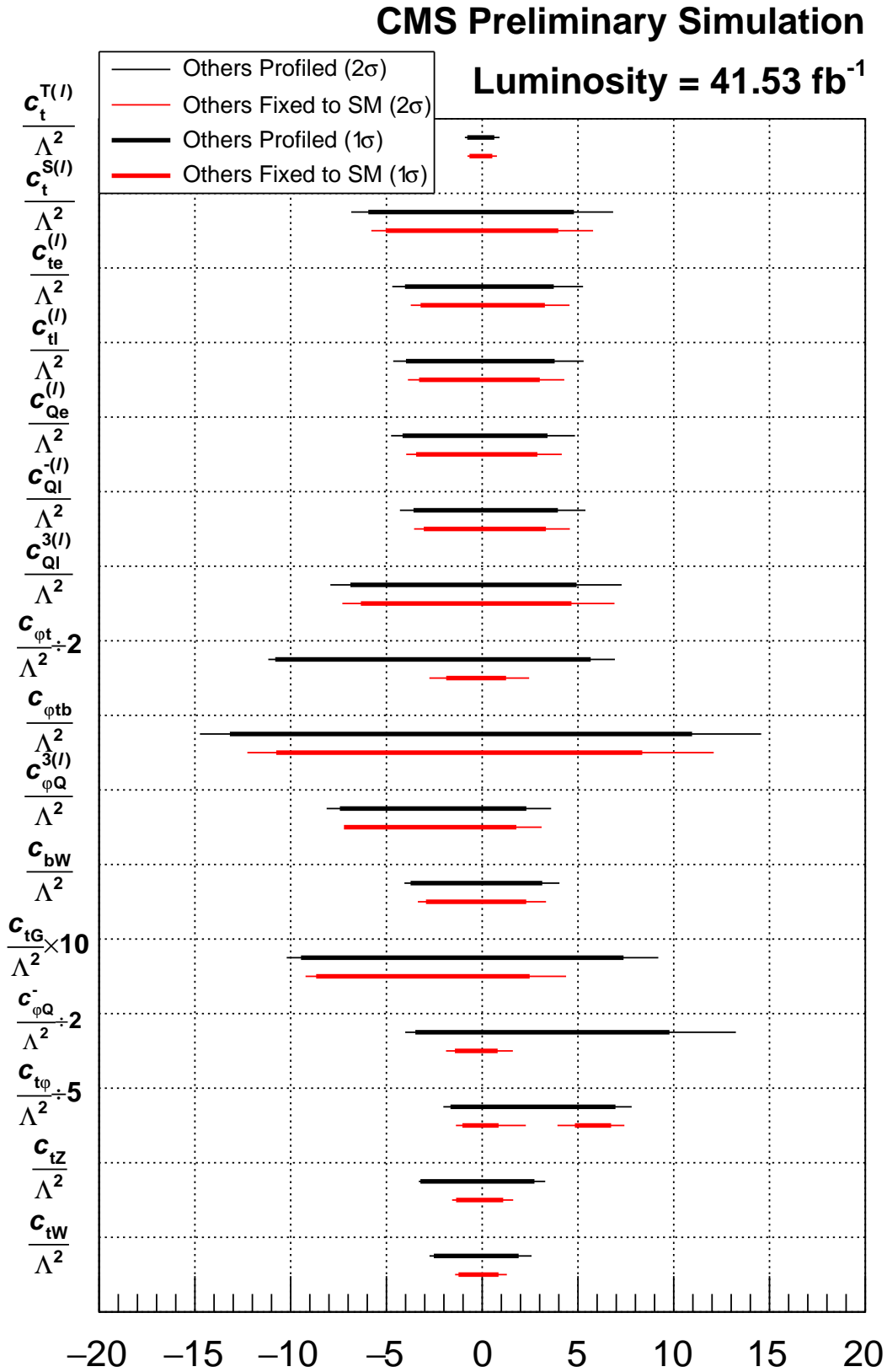
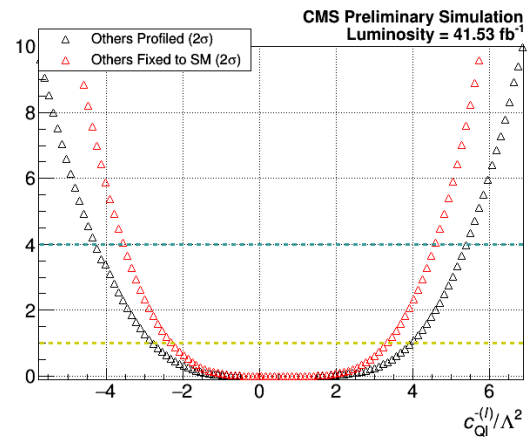
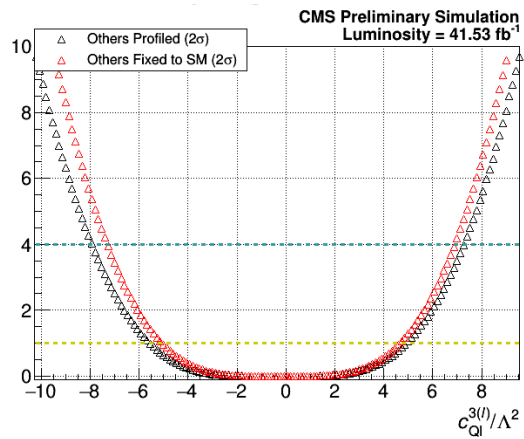
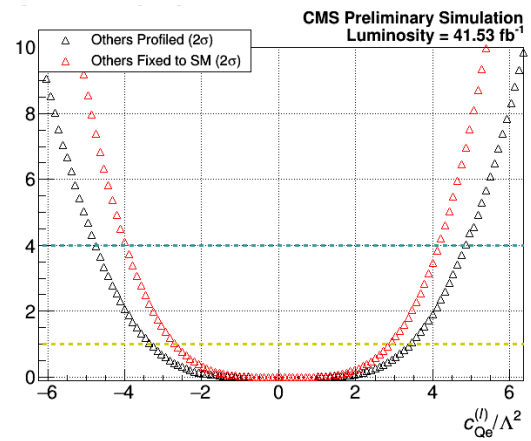
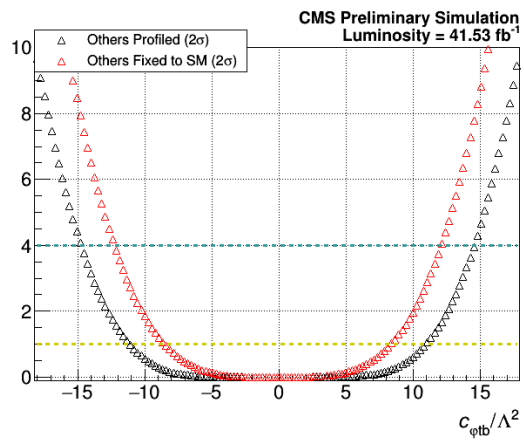
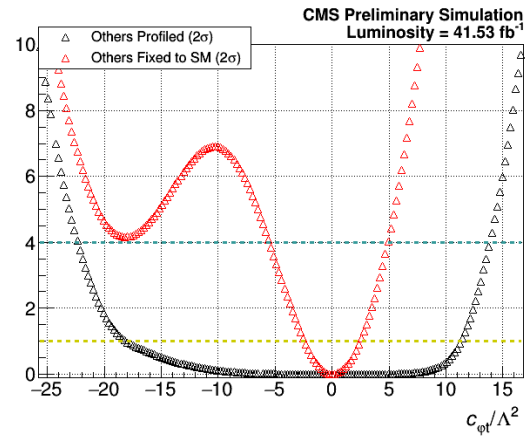
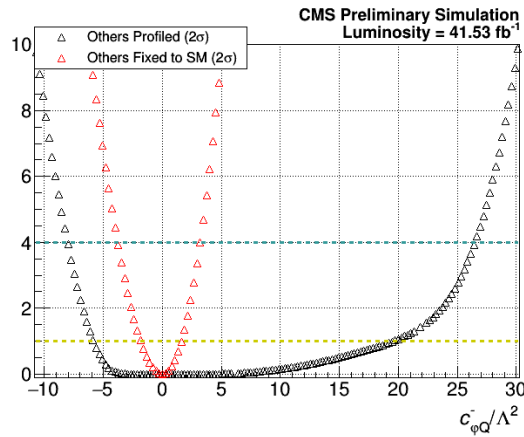
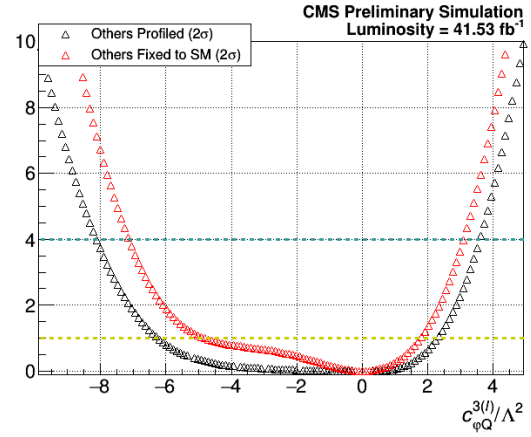
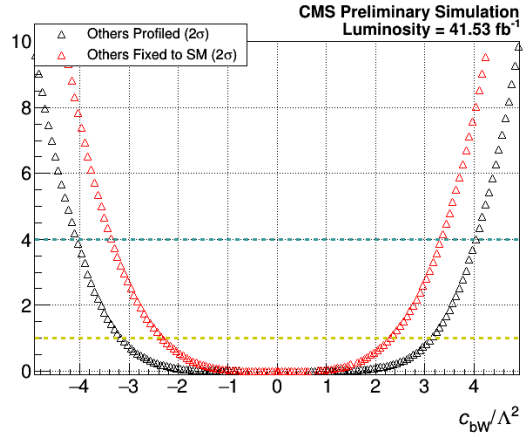


Figure 14: Standard Model-expected uncertainty intervals. WC 1σ (thick line) and 2σ (thin line) uncertainty intervals. Black lines correspond to the other WCs profiled, while red lines correspond to the other WCs fixed to the SM value of zero. Taken from 1D scans with other Wilson coefficients profiled or fixed to the SM value of zero.



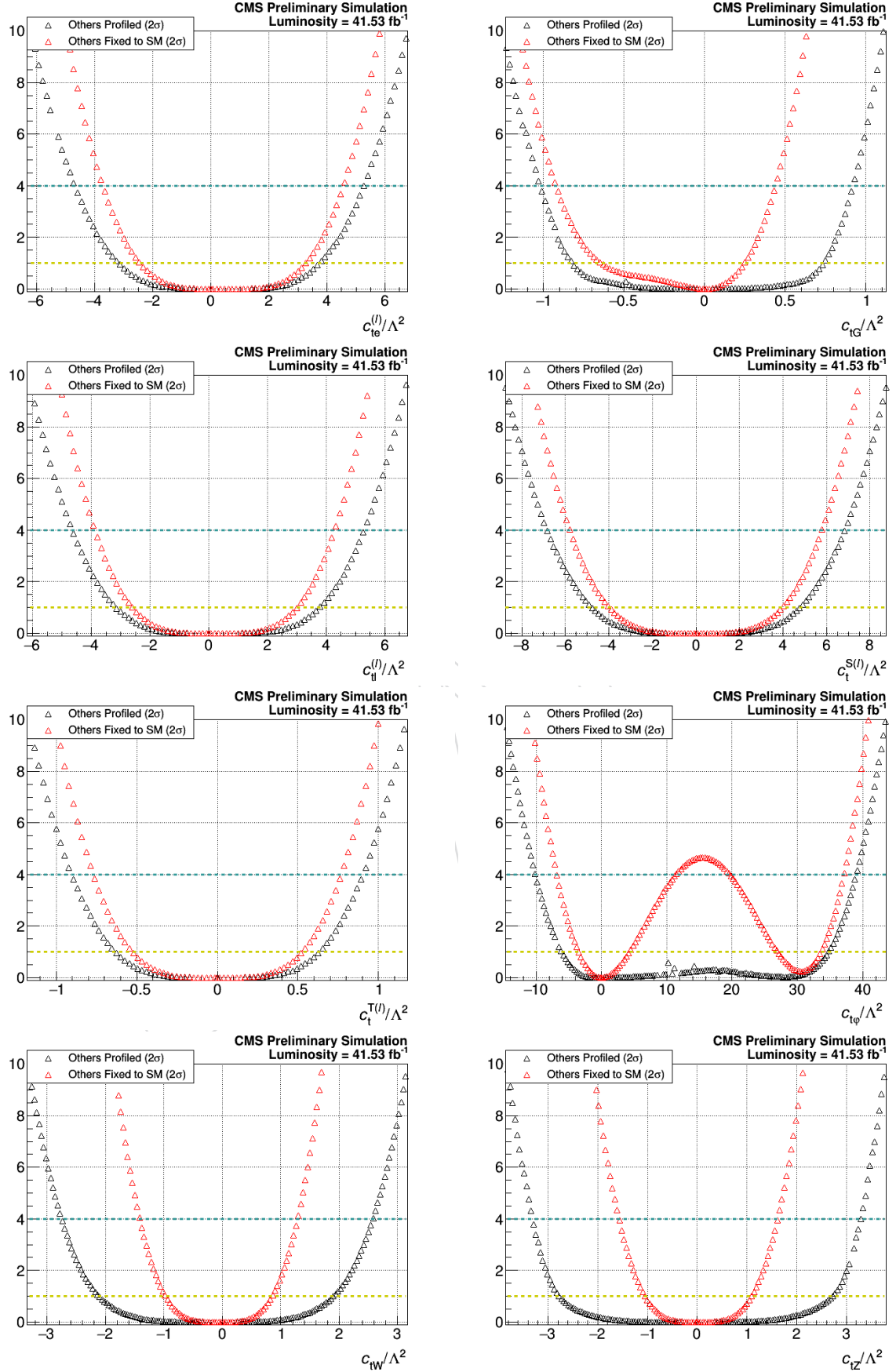
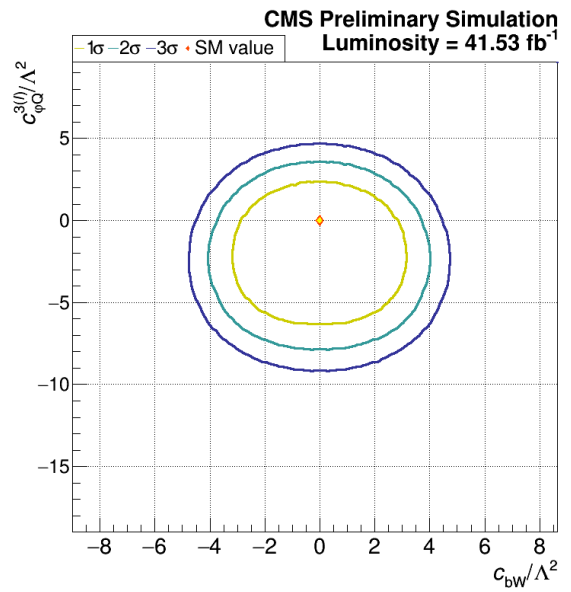
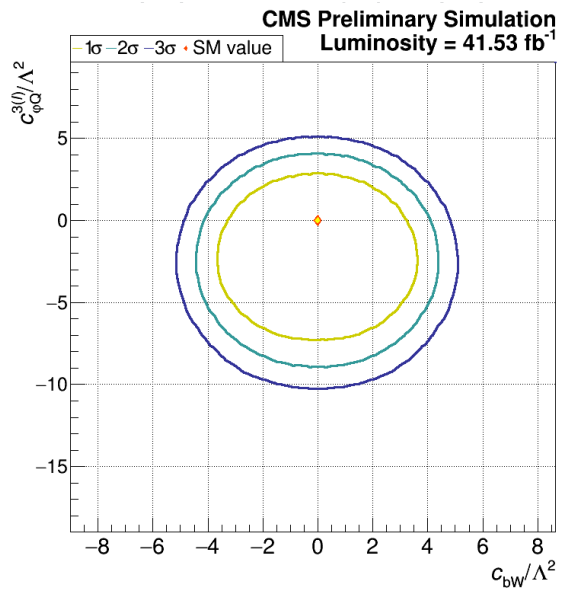
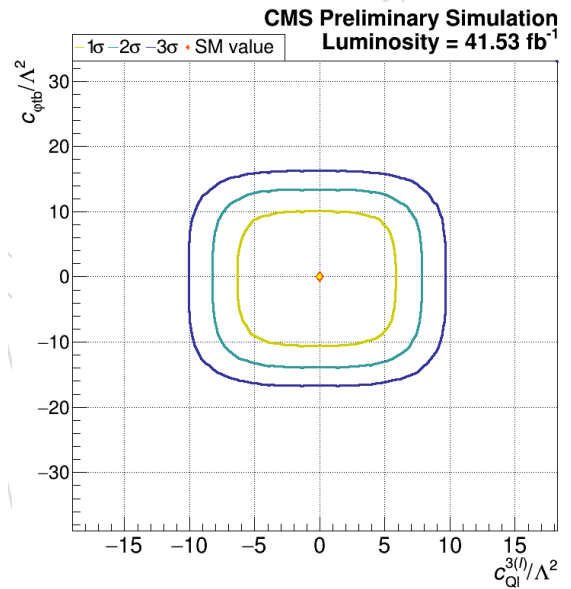
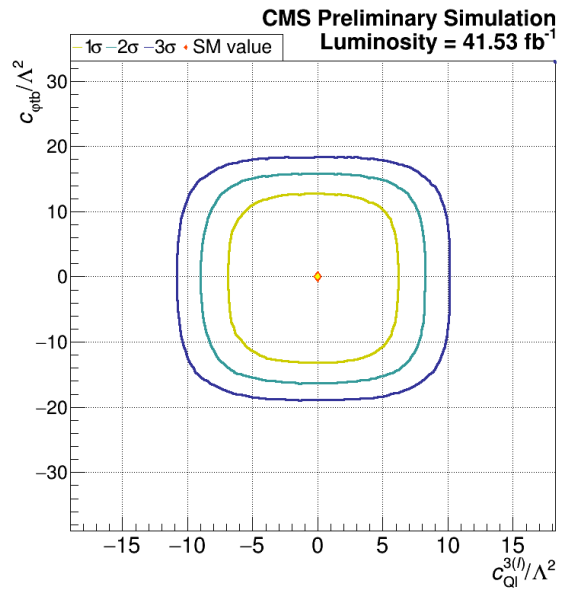
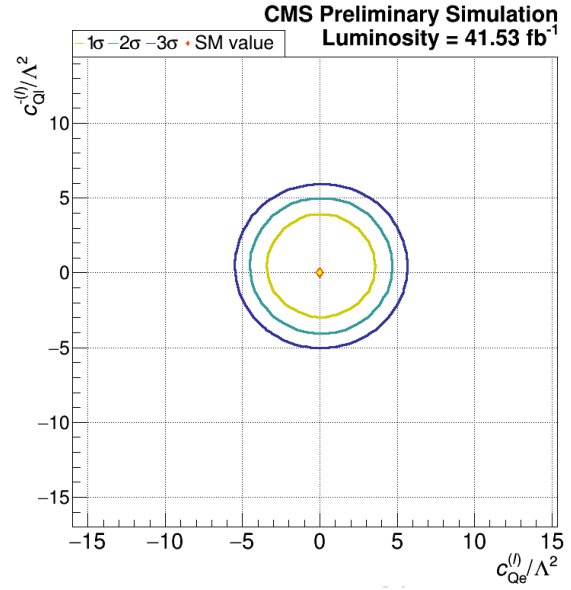
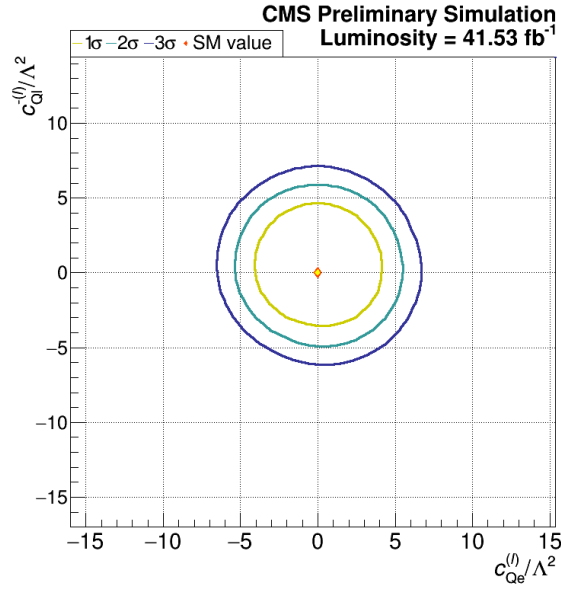
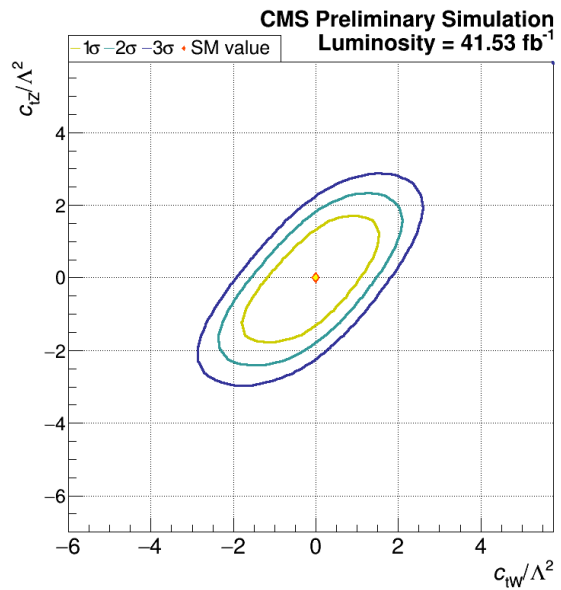
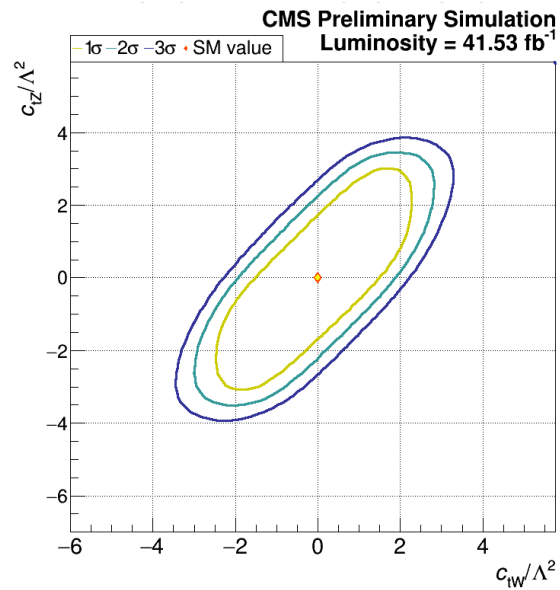
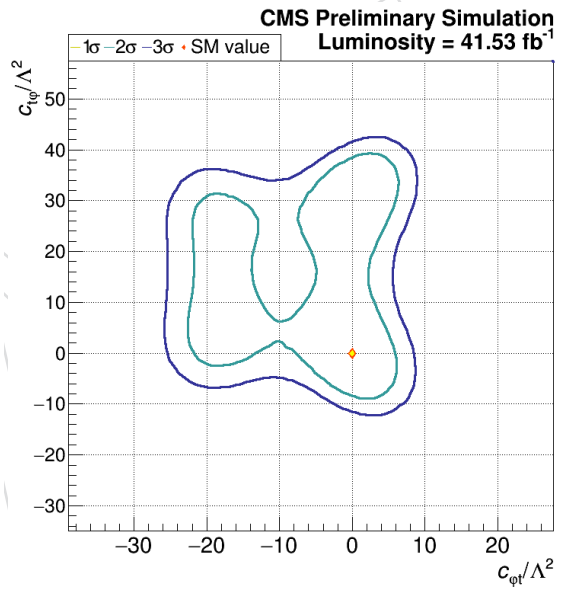
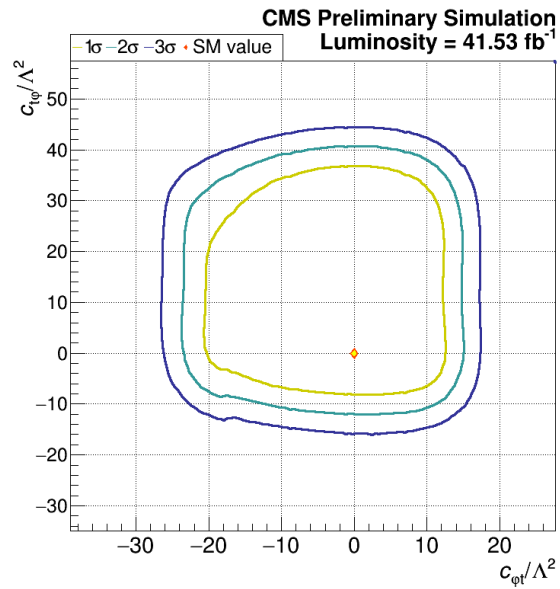
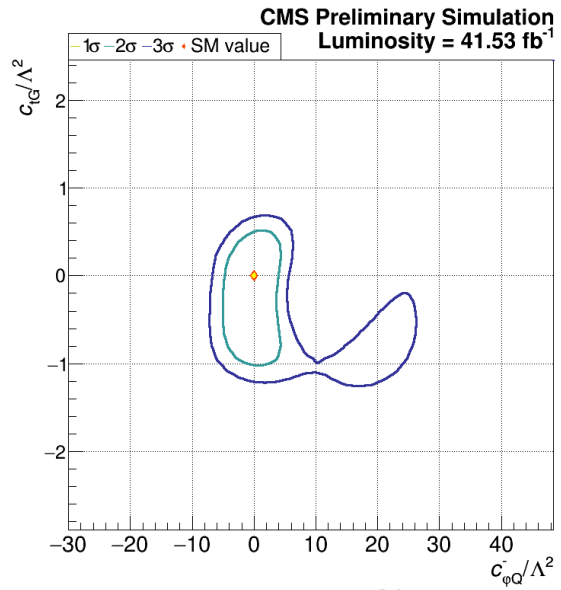
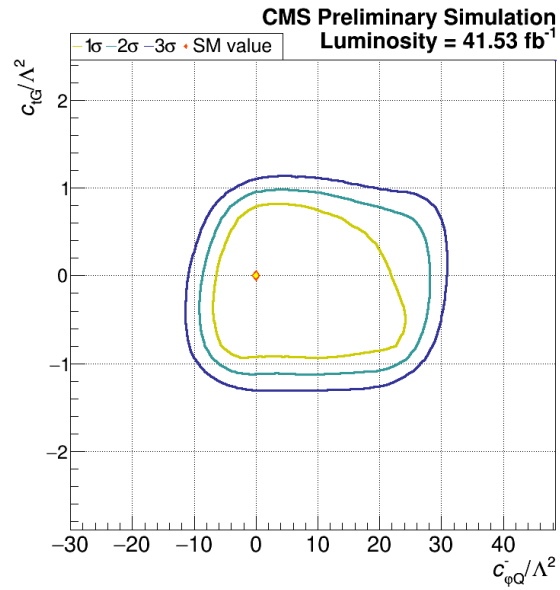


Figure 15: Scans over values of single Wilson Coefficients. Standard Model-expected results. Black points are profiled scans while the other 15 Wilson Coefficients are profiled. Red points are scans while the other 15 Wilson Coefficients are fixed to their Standard Model values.





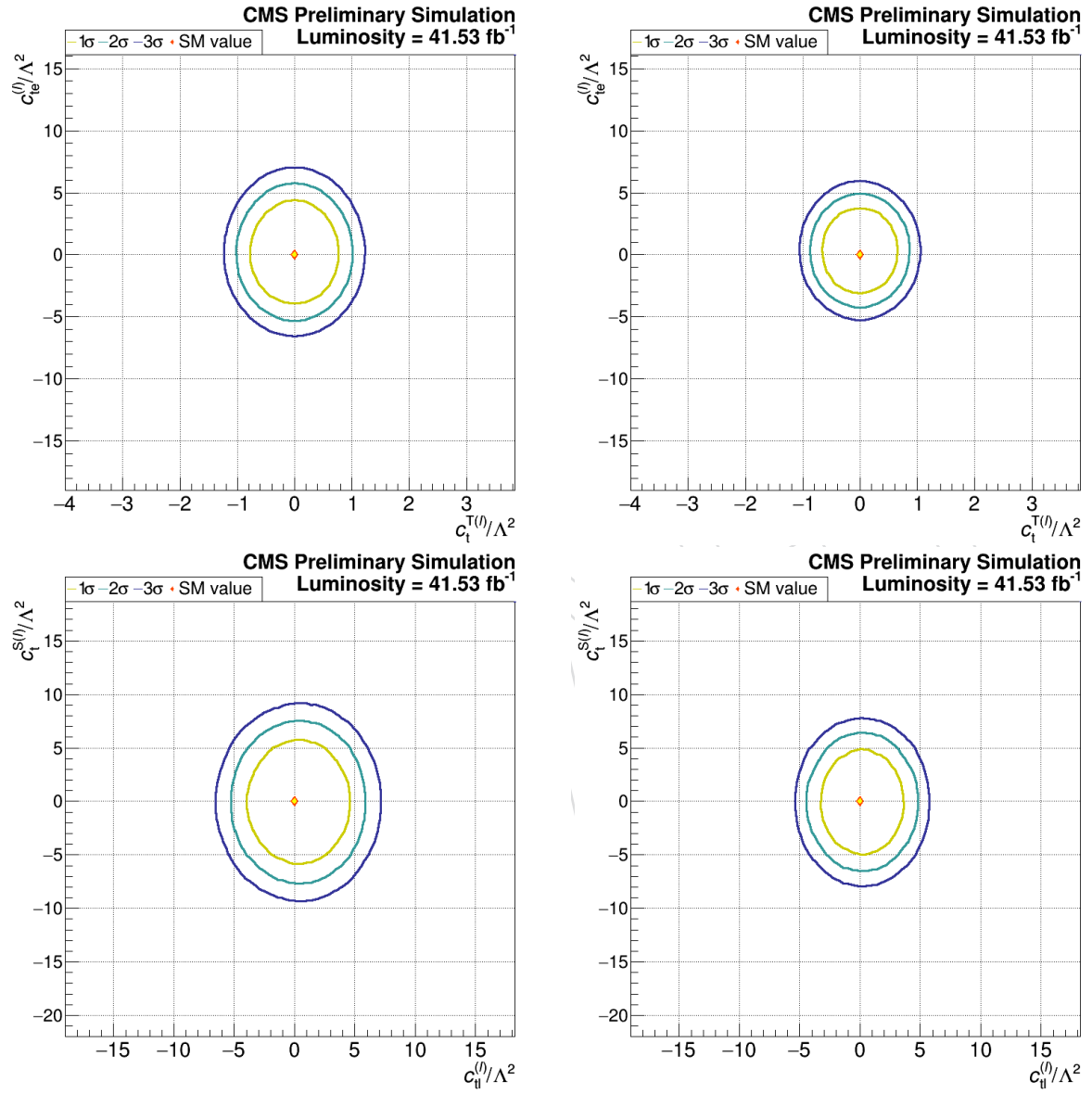


Figure 16: Left column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients profiled. Right column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients fixed to their Standard Model value.

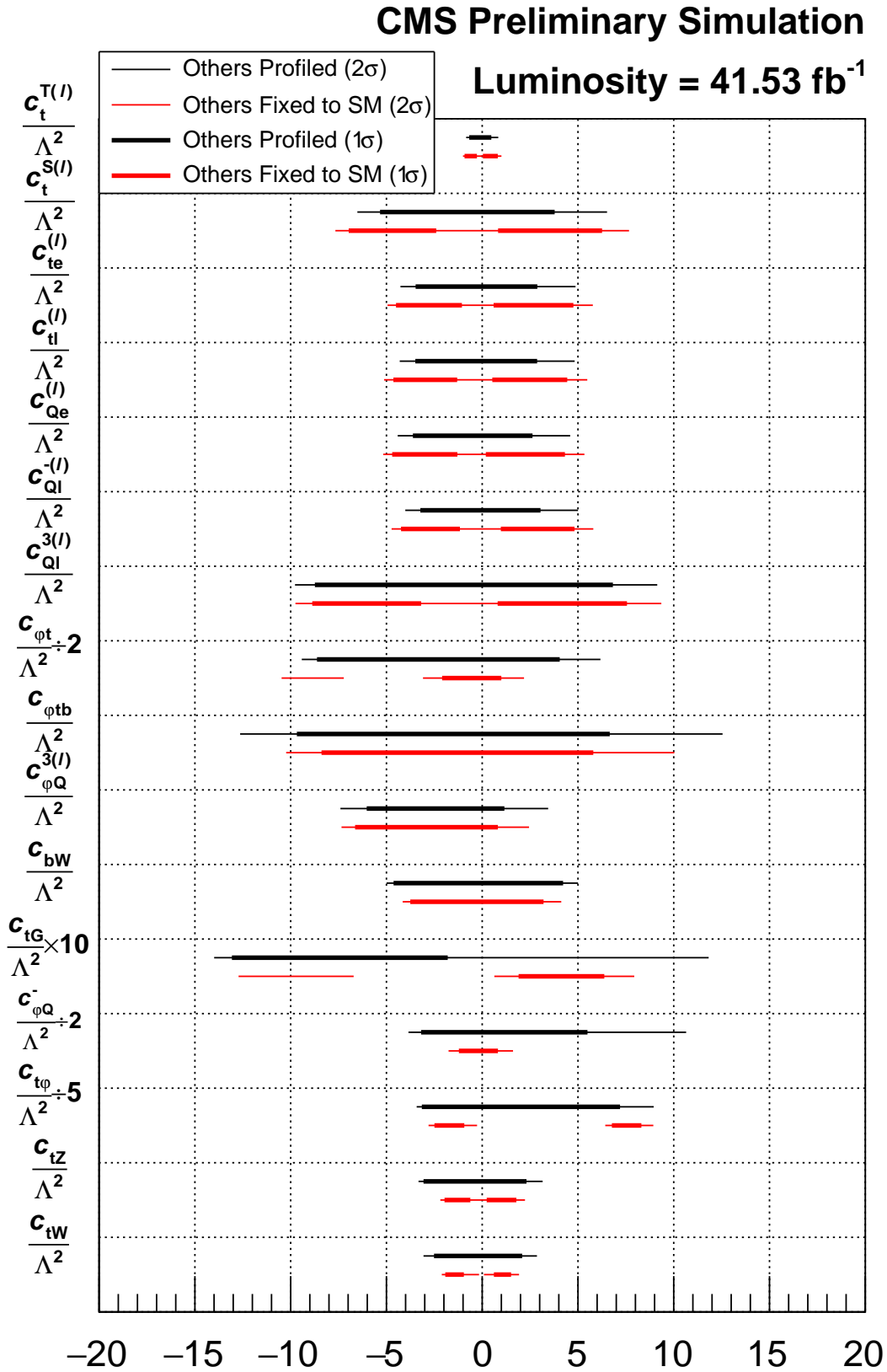
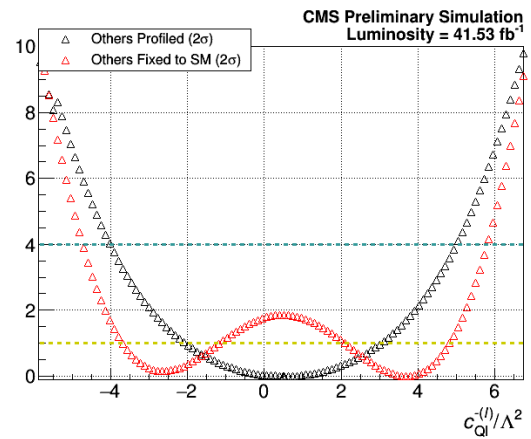
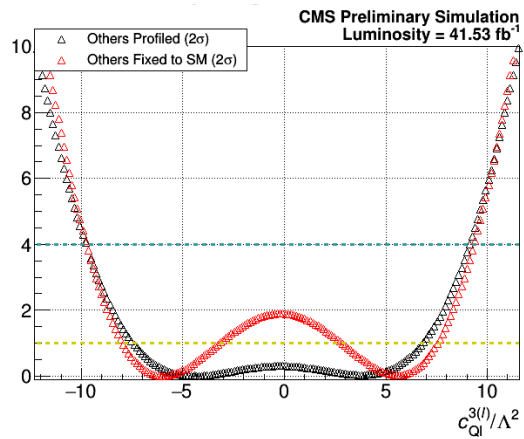
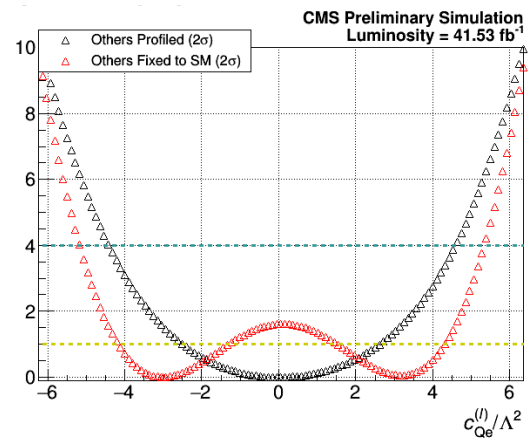
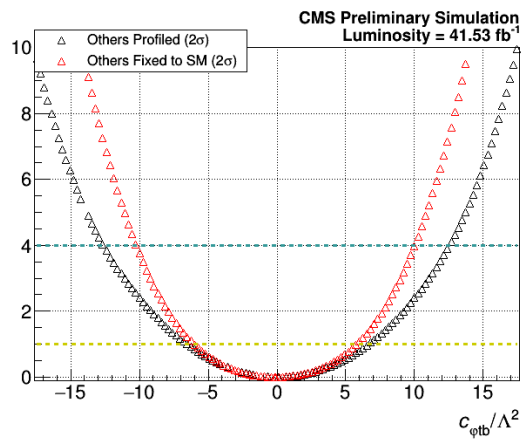
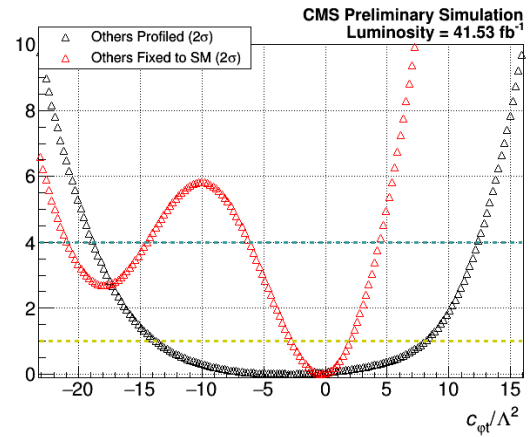
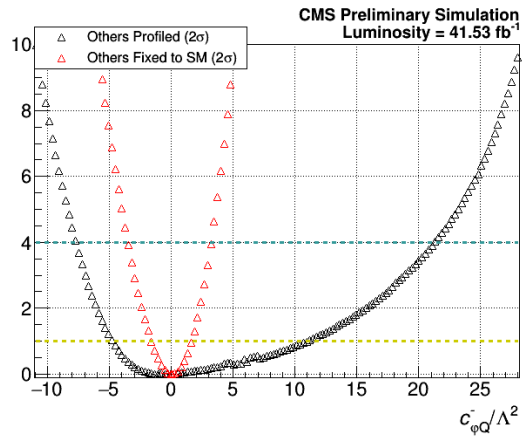
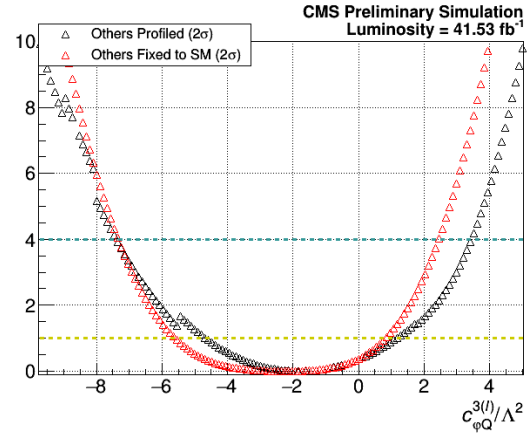
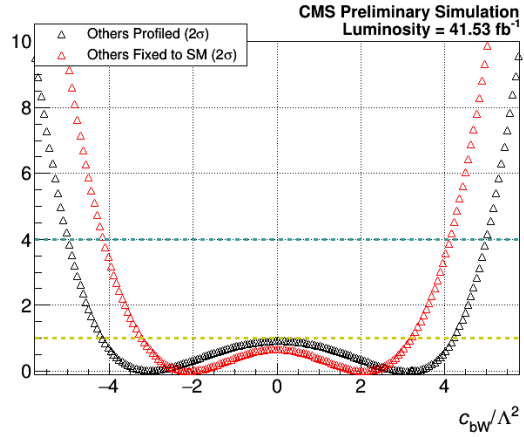


Figure 17: Results for fits to the 2017 CMS dataset. WC 1σ (thick line) and 2σ (thin line) uncertainty intervals. Black lines correspond to the other WCs profiled, while red lines correspond to the other WCs fixed to the SM value of zero. Taken from 1D scans with other Wilson coefficients profiled or fixed to the SM value of zero.



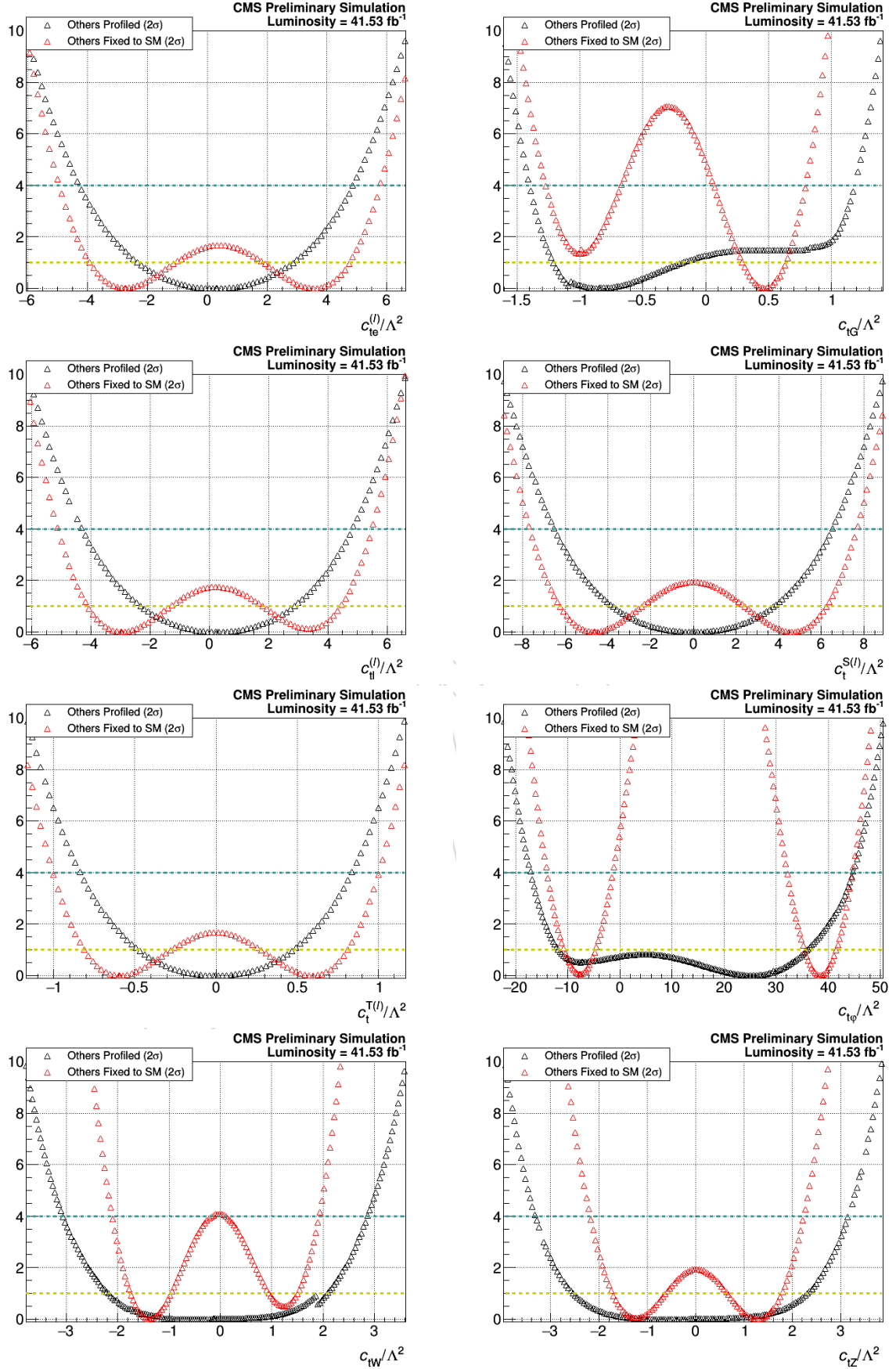
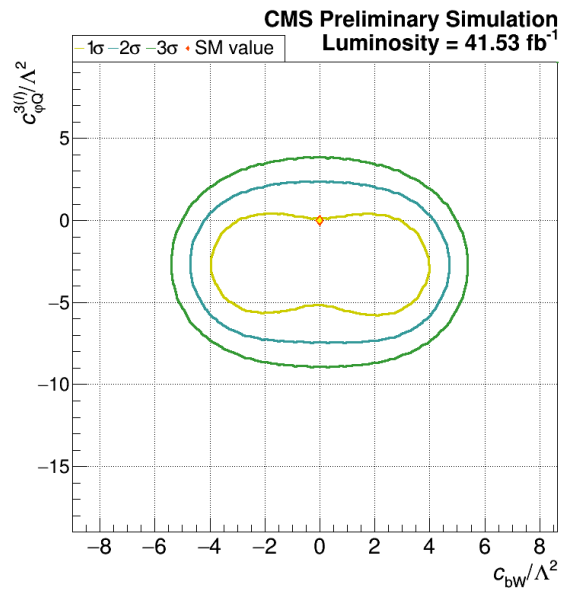
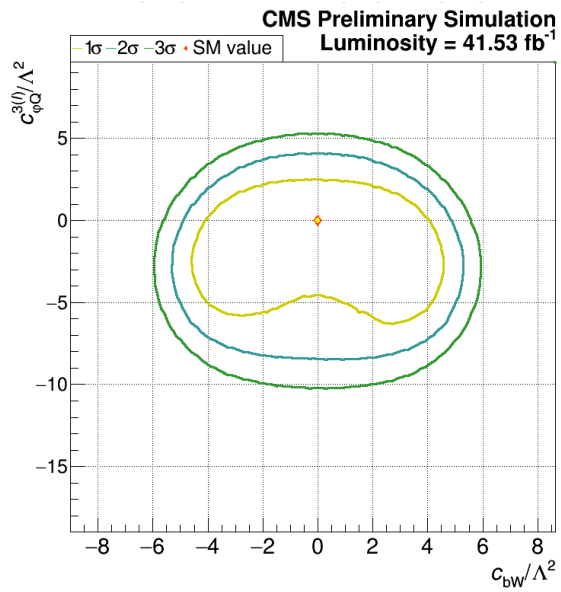
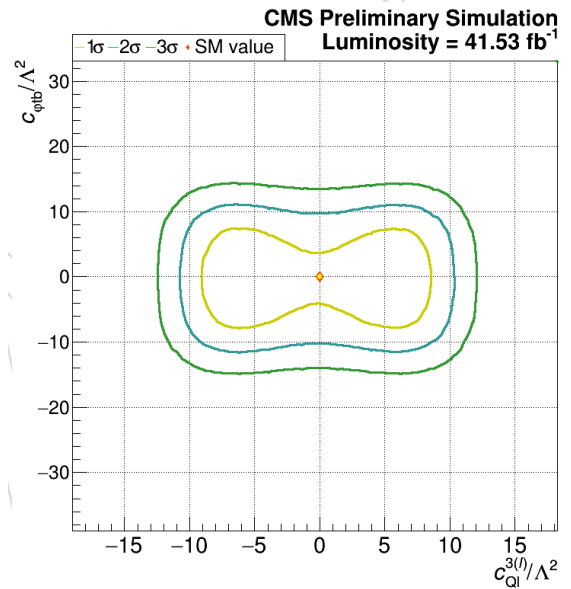
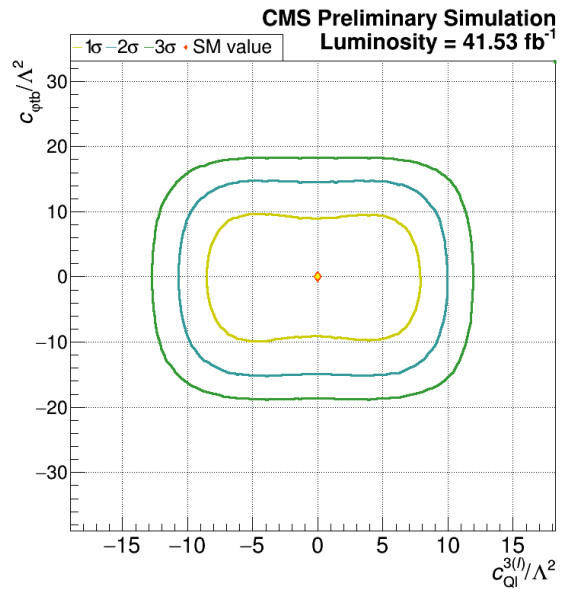
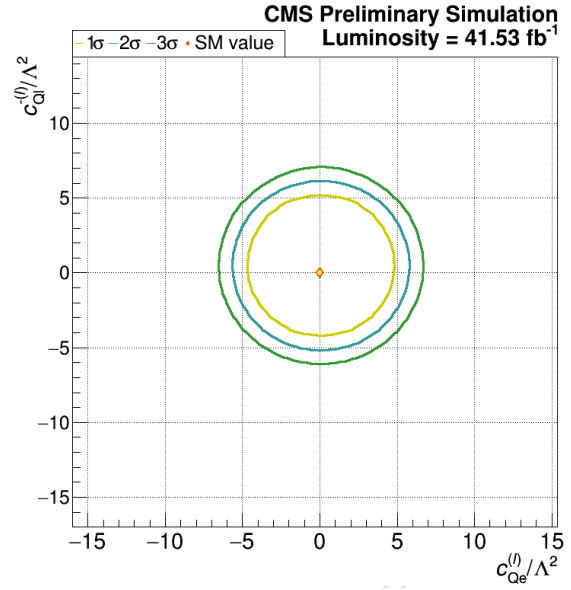
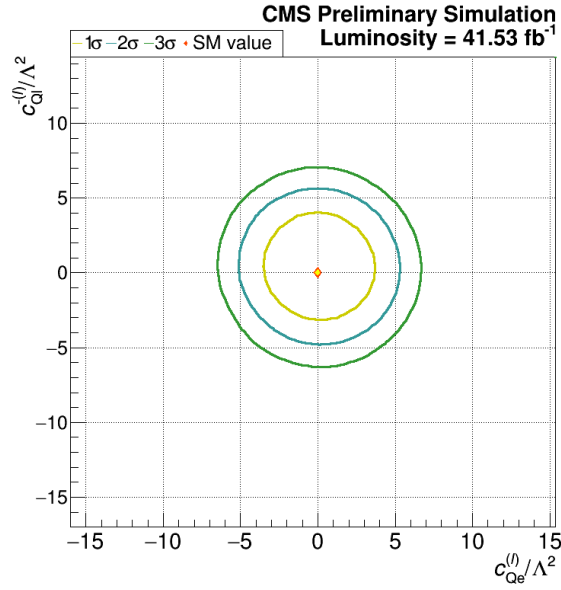
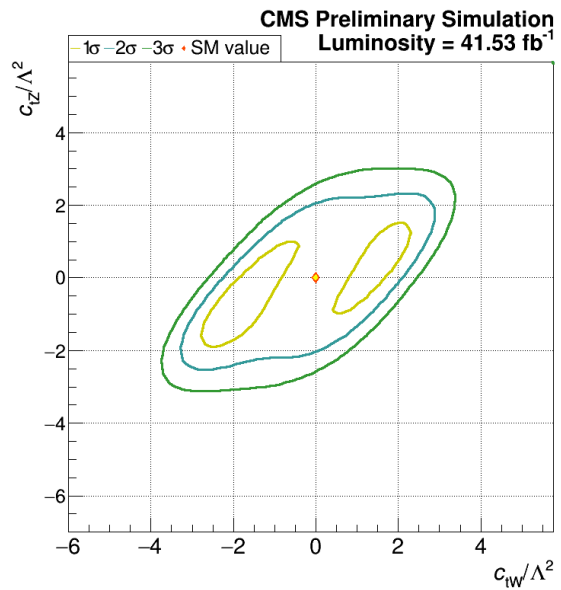
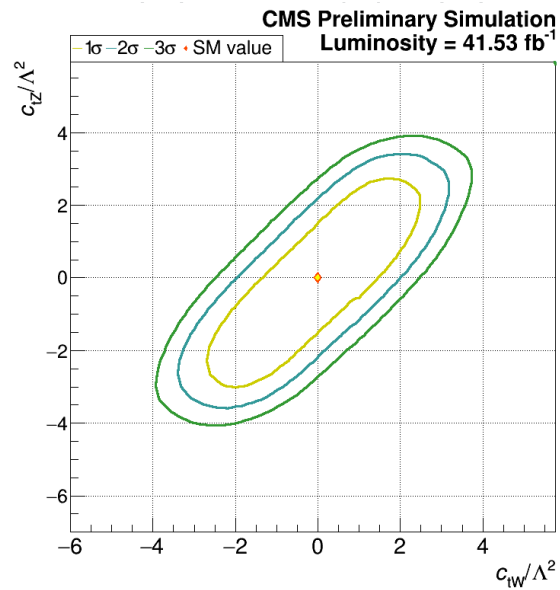
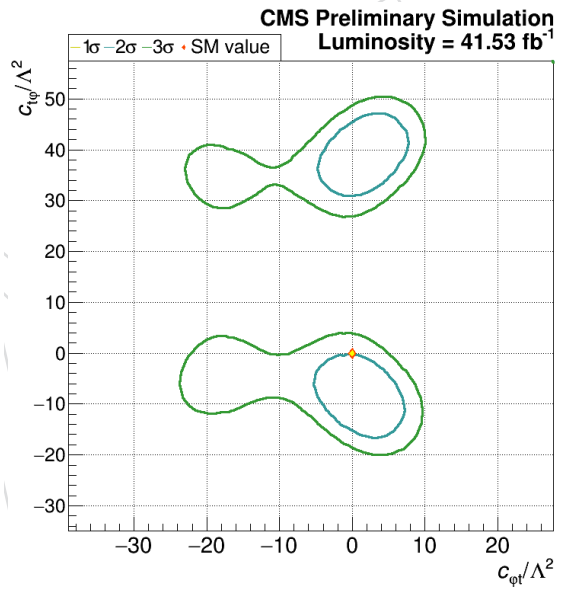
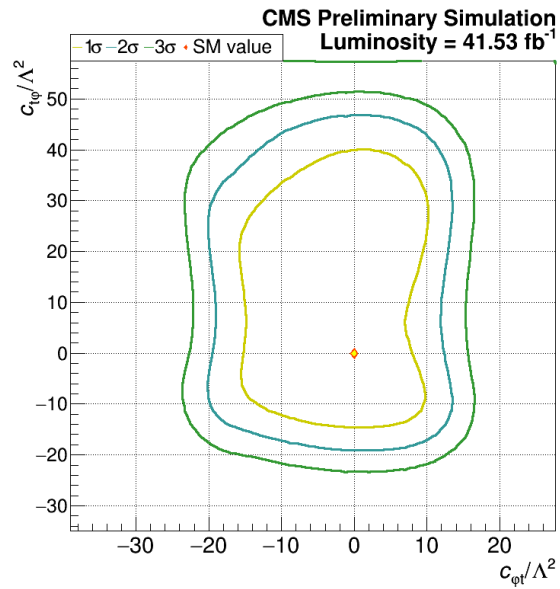
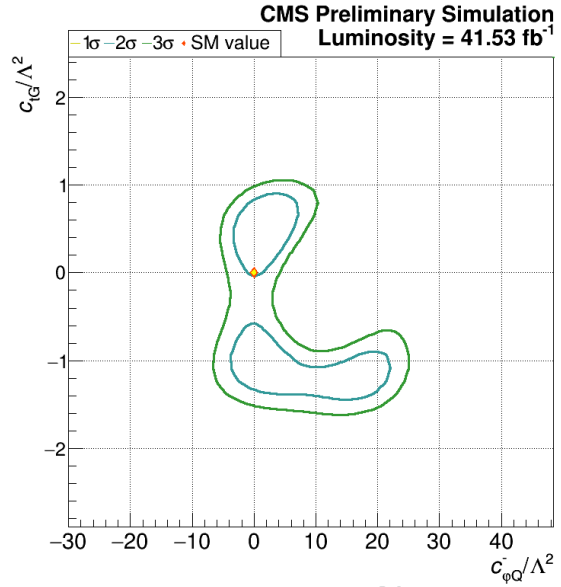
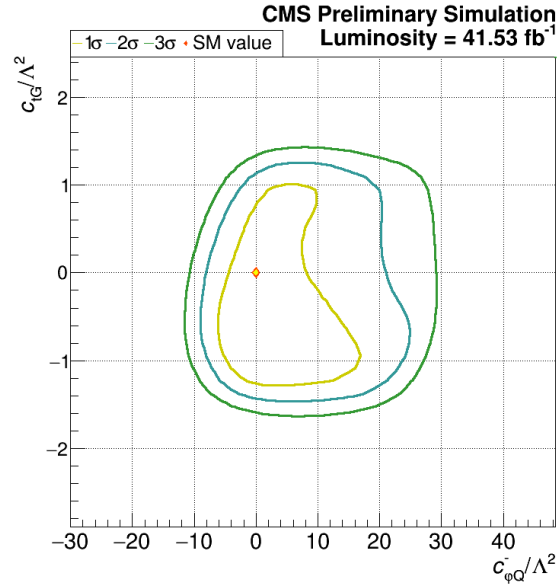


Figure 18: Scans over values of single Wilson Coefficients. Results from fits to the 2017 CMS dataset. Black points are profiled scans while the other 16 Wilson Coefficients are profiled. Red points are scans while the other 15 Wilson Coefficients are fixed to their Standard Model values.





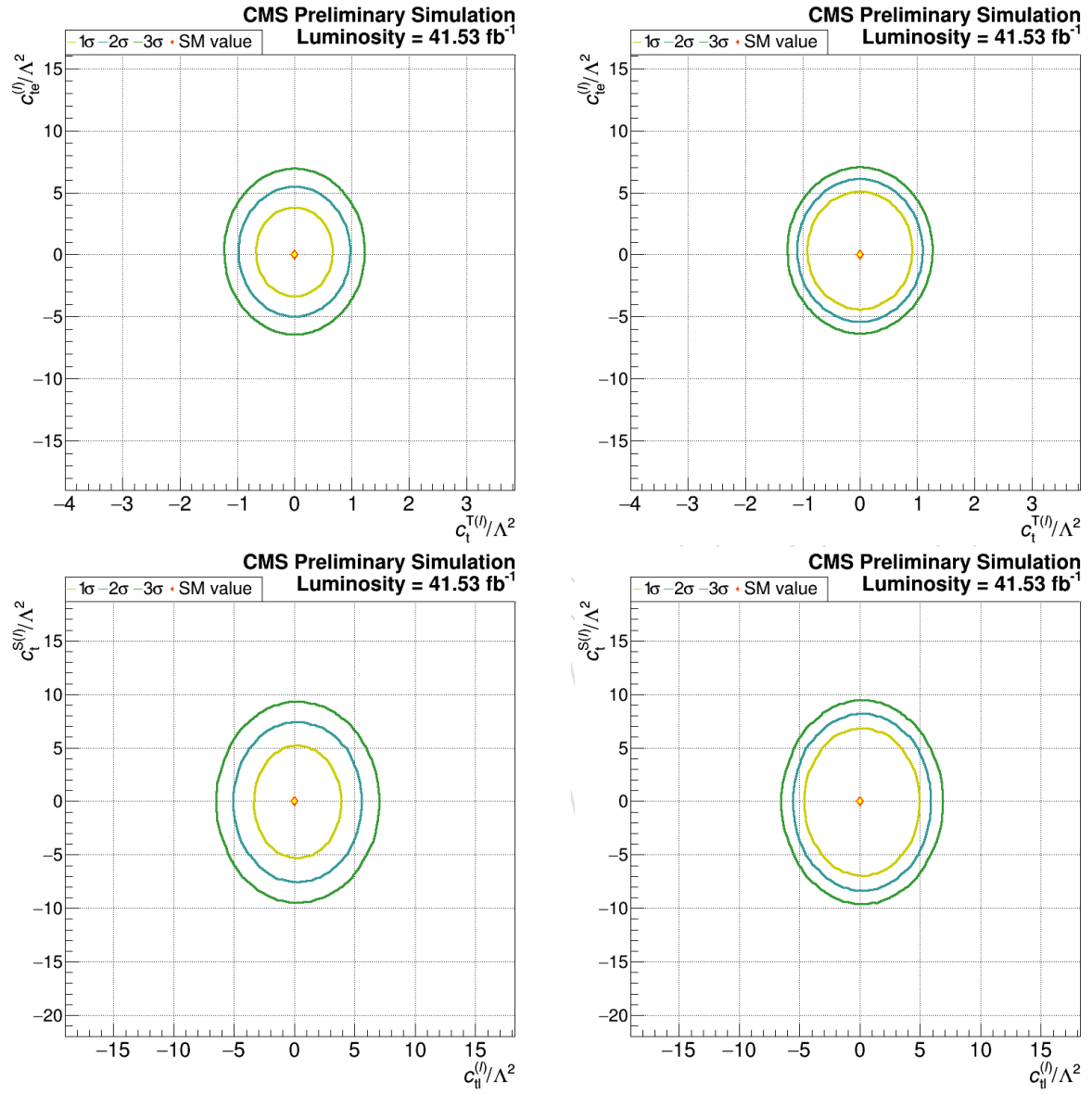


Figure 19: Left column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients profiled. Right column: The Standard Model-predicted 1σ , 2σ , and 3σ contours of a 2D scan with other Wilson Coefficients fixed to their Standard Model value.

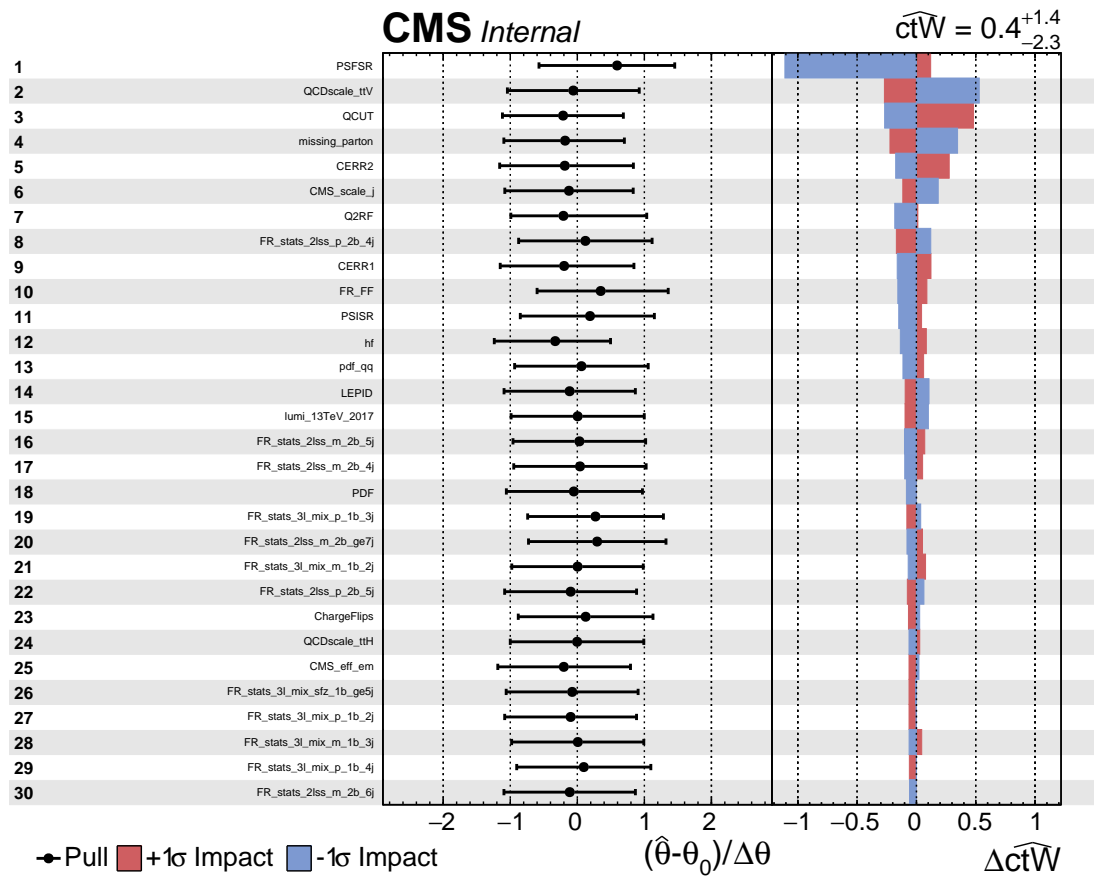


Figure 20: The impacts and pulls of the largest systematics for c_{tW} , ranked by impact.

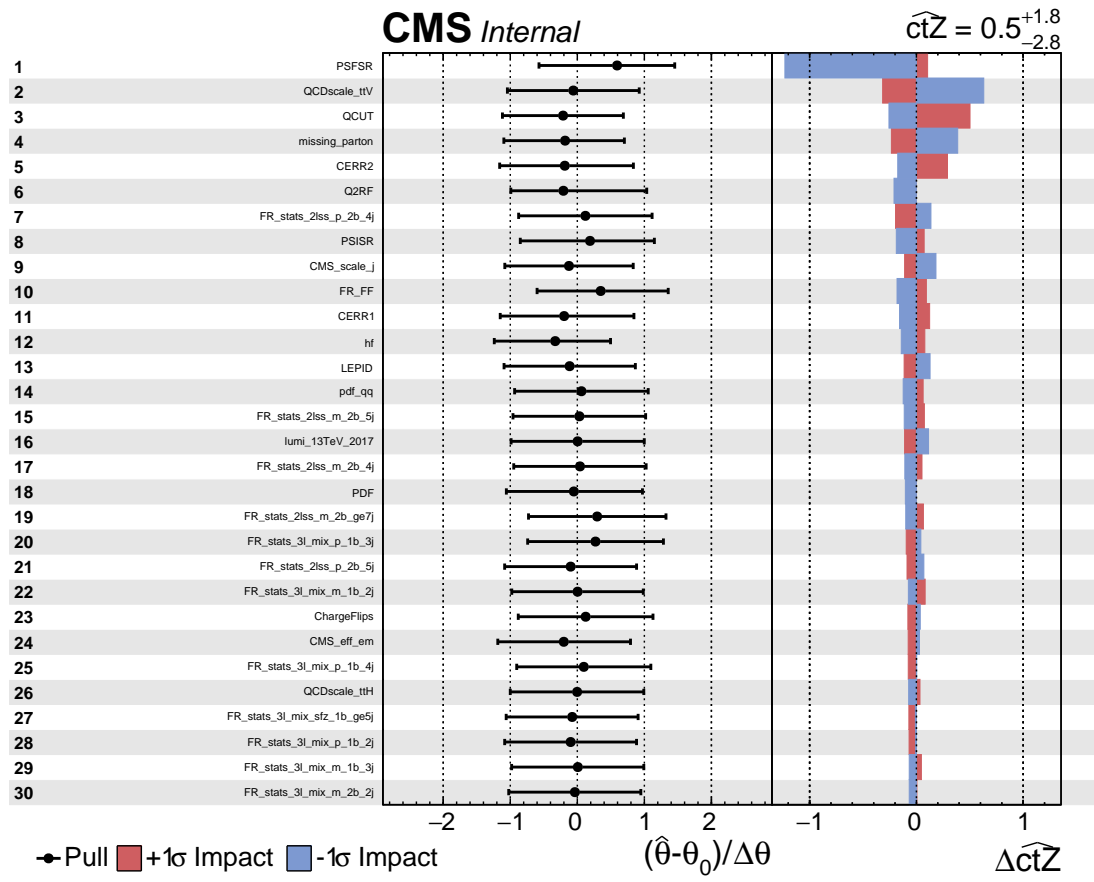


Figure 21: The impacts and pulls of the largest systematics for c_{tZ} , ranked by impact.

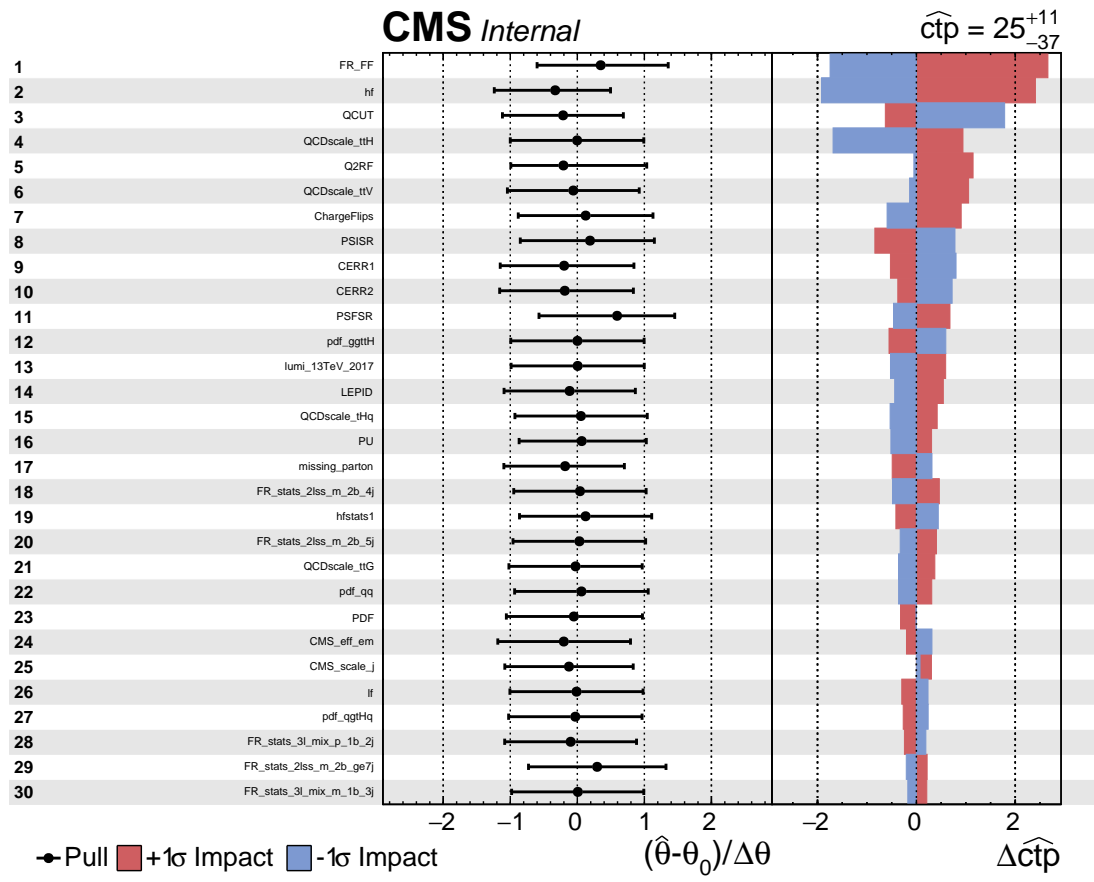


Figure 22: The impacts and pulls of the largest systematics for $c_{t\varphi}$, ranked by impact.

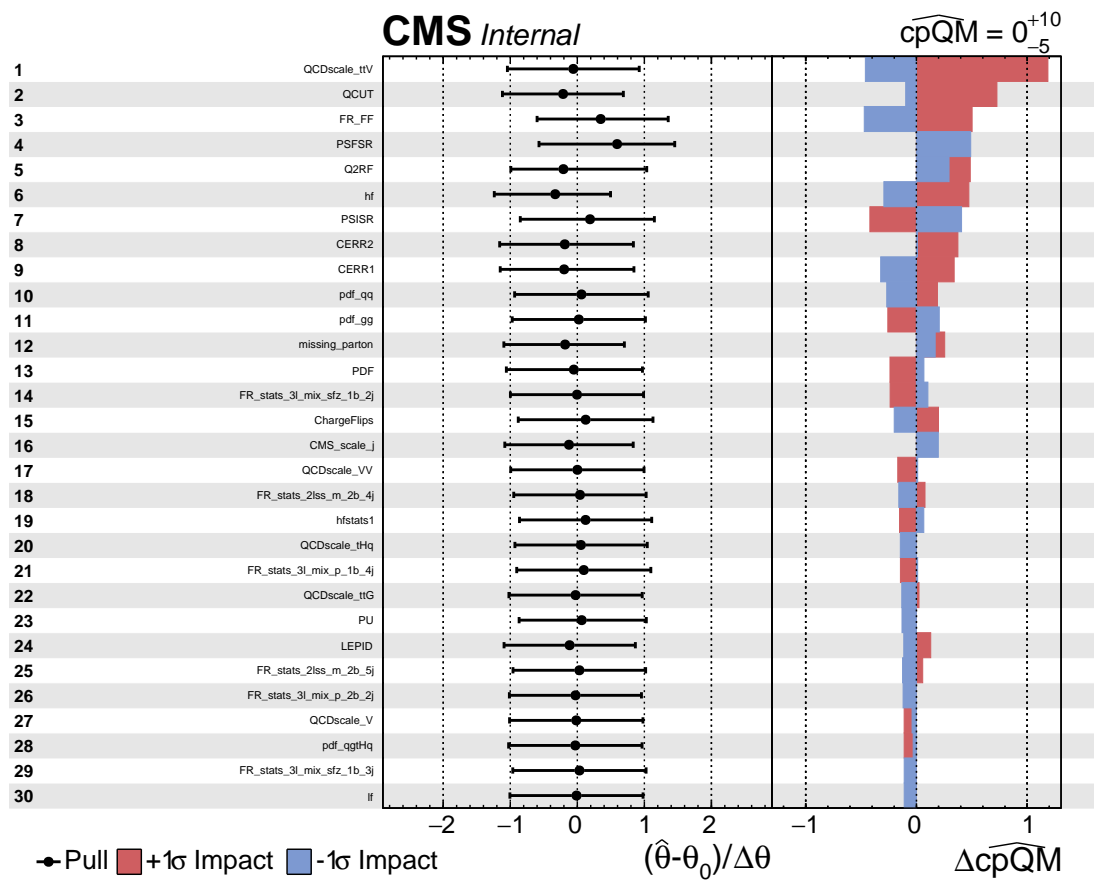


Figure 23: The impacts and pulls of the largest systematics for $c_{\varphi Q}^-$, ranked by impact.

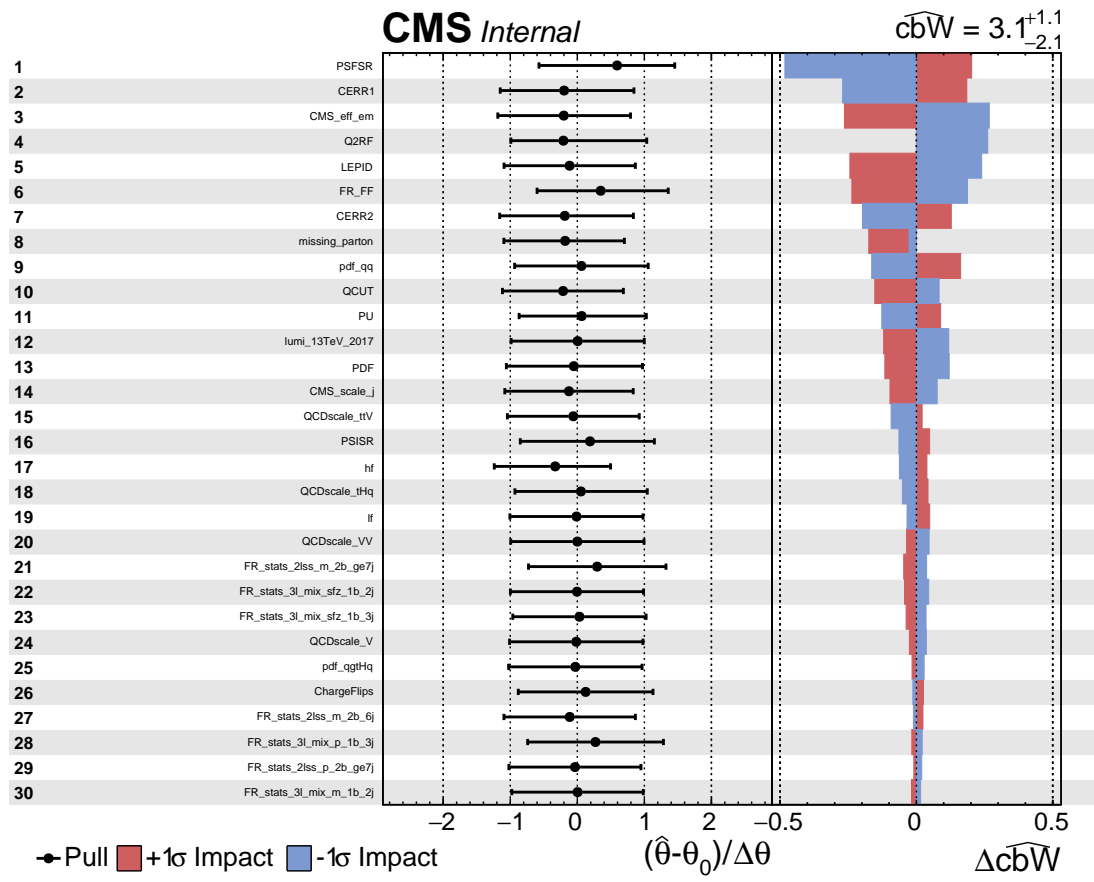


Figure 24: The impacts and pulls of the largest systematics for c_{bW} , ranked by impact.

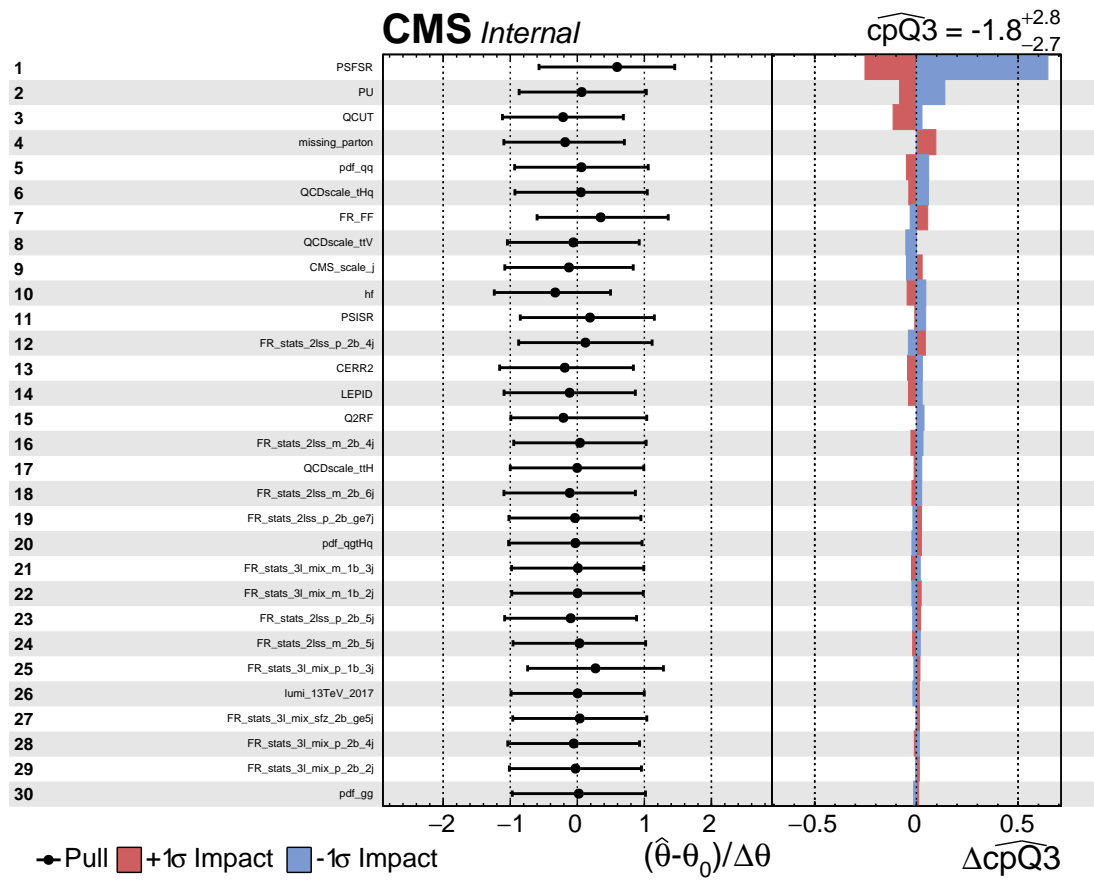


Figure 25: The impacts and pulls of the largest systematics for $c^3_{\varphi Q}$, ranked by impact.

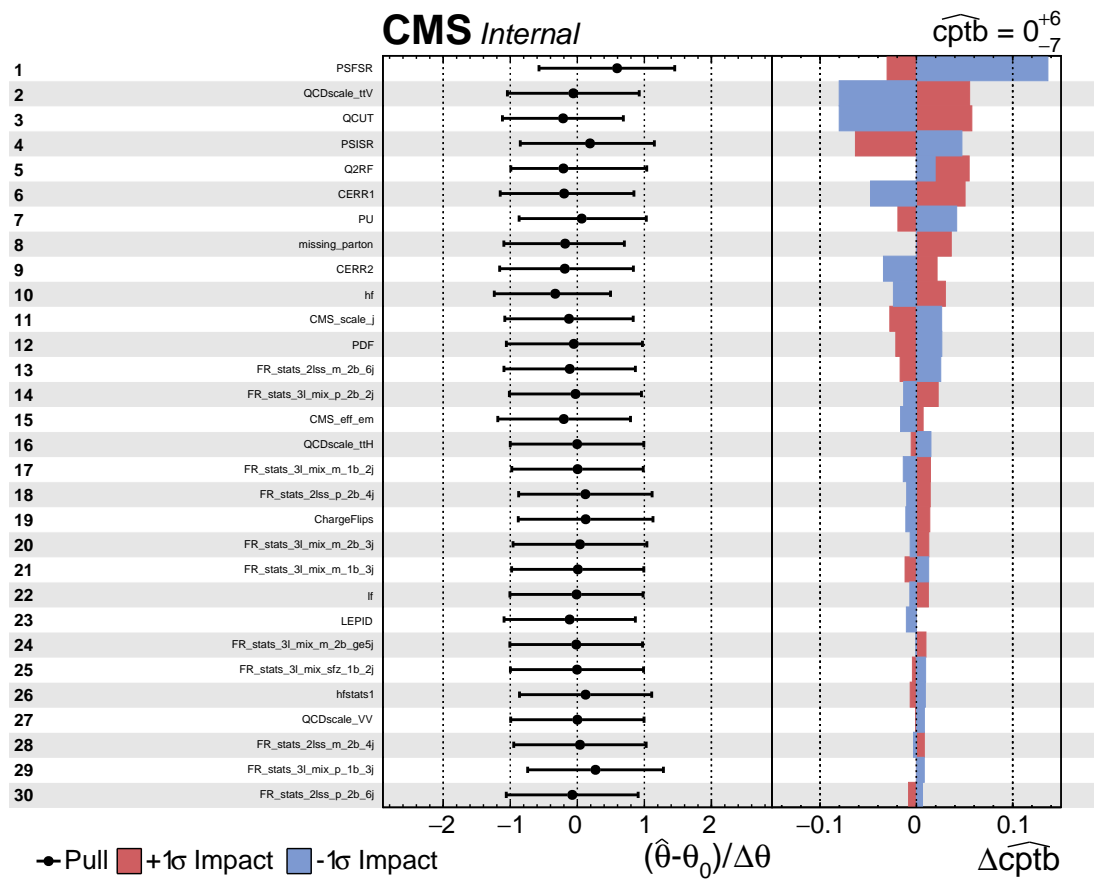


Figure 26: The impacts and pulls of the largest systematics for c_{ptb} , ranked by impact.

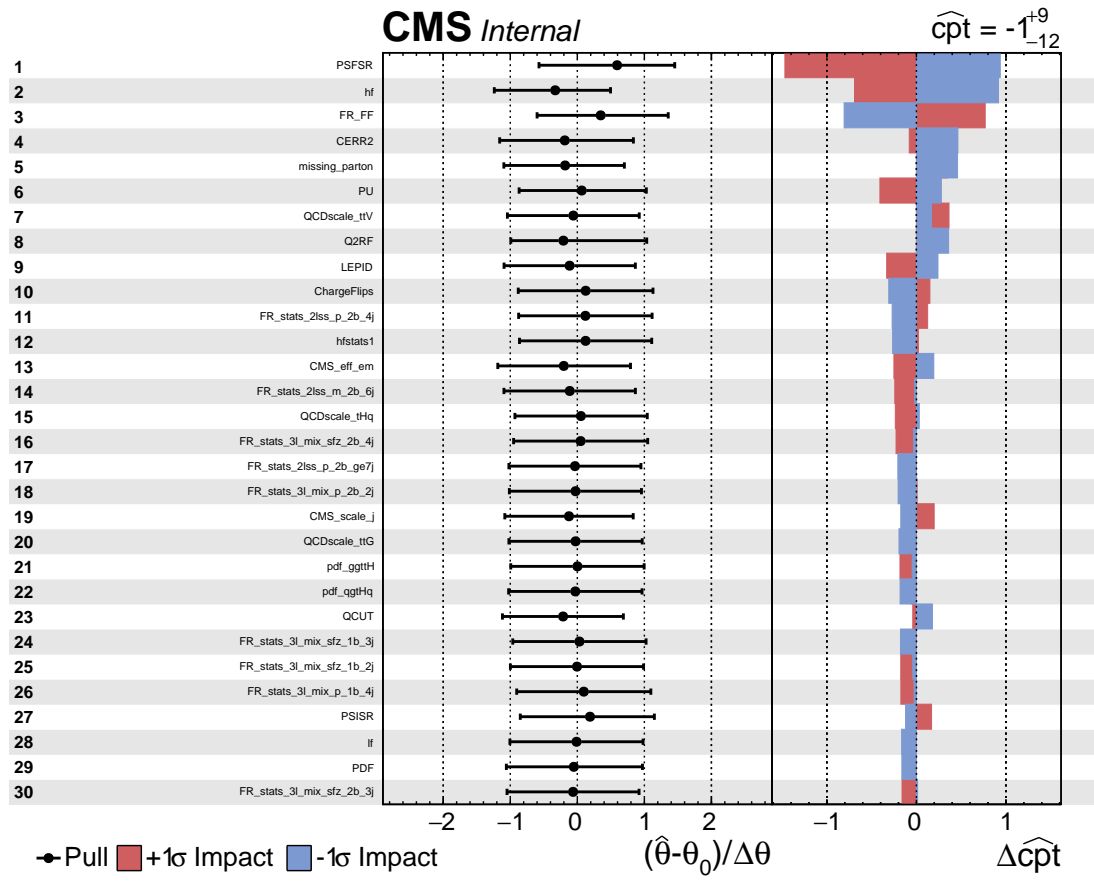


Figure 27: The impacts and pulls of the largest systematics for c_{pt} , ranked by impact.

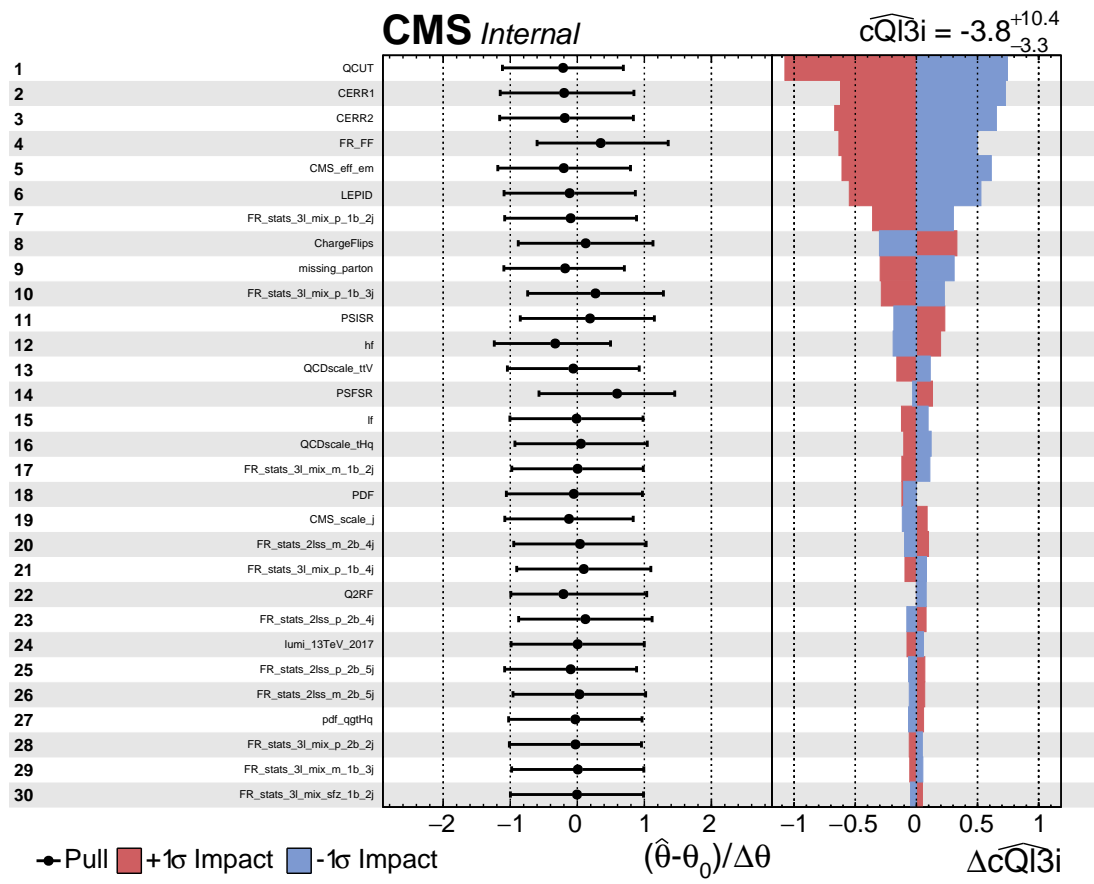


Figure 28: The impacts and pulls of the largest systematics for $c_{Q\ell}^{3(\ell)}$, ranked by impact.

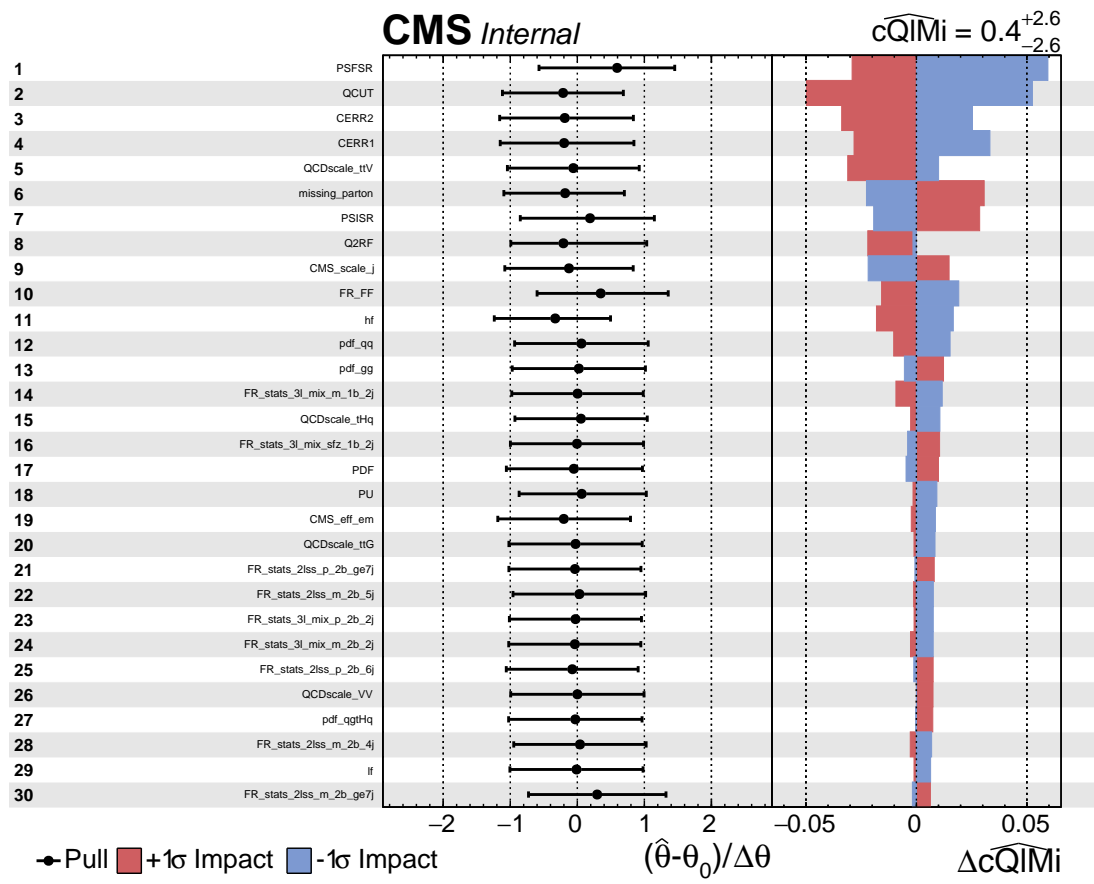


Figure 29: The impacts and pulls of the largest systematics for $c_{Q\ell}^{-(\ell)}$, ranked by impact.

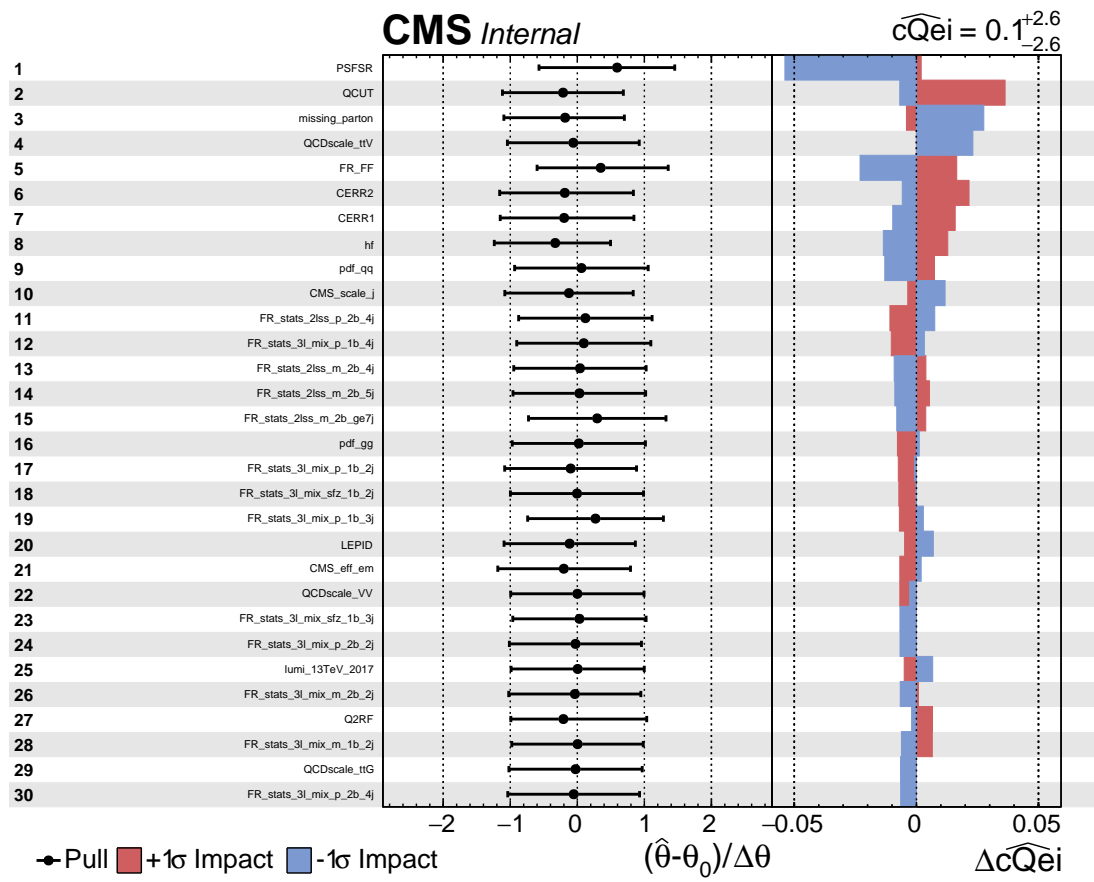


Figure 30: The impacts and pulls of the largest systematics for $c_{Qe}^{(\ell)}$, ranked by impact.

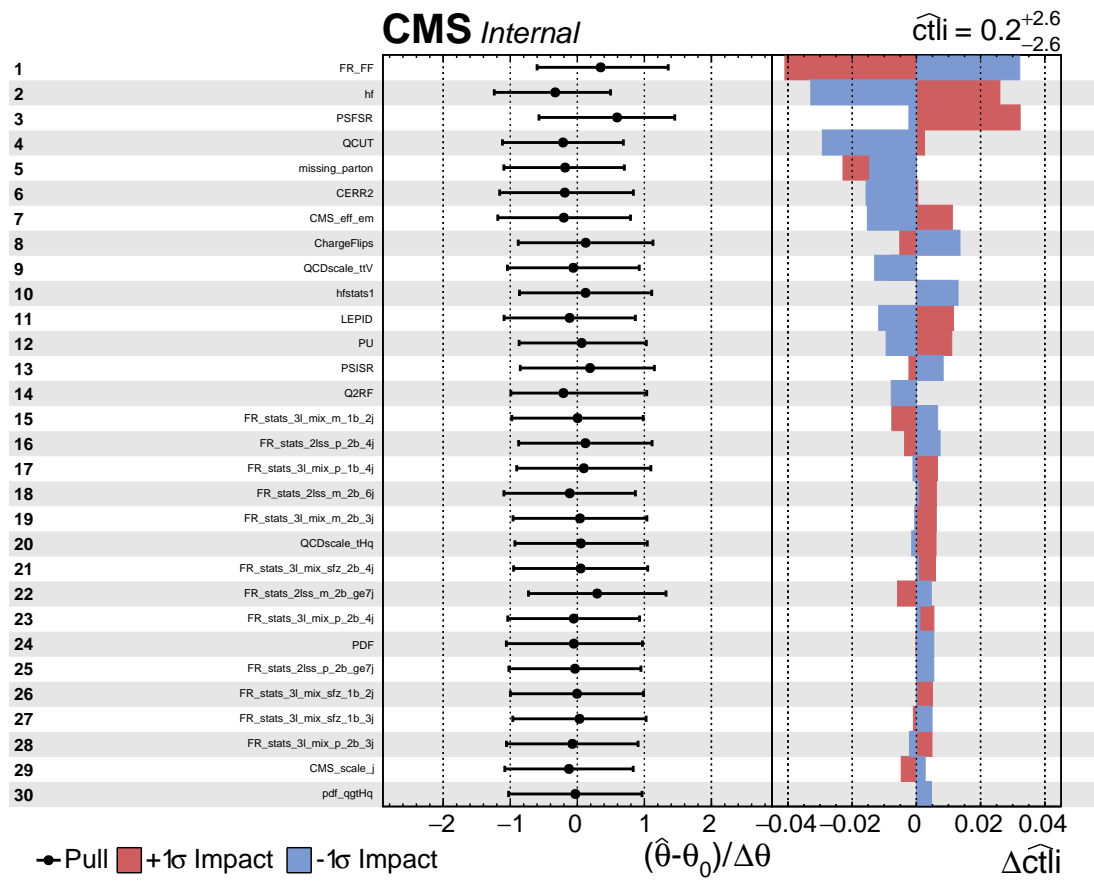


Figure 31: The impacts and pulls of the largest systematics for $c_{t\ell}^{(\ell)}$, ranked by impact.

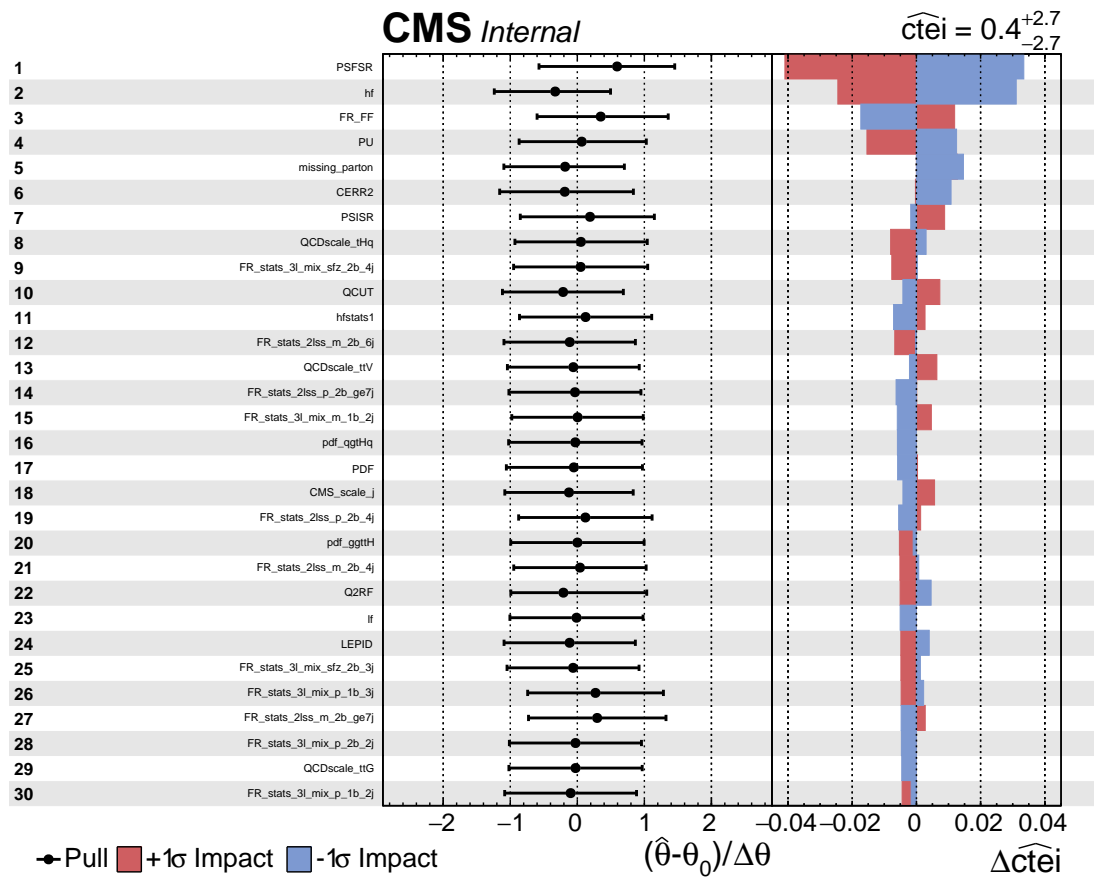


Figure 32: The impacts and pulls of the largest systematics for $c_{te}^{(\ell)}$, ranked by impact.

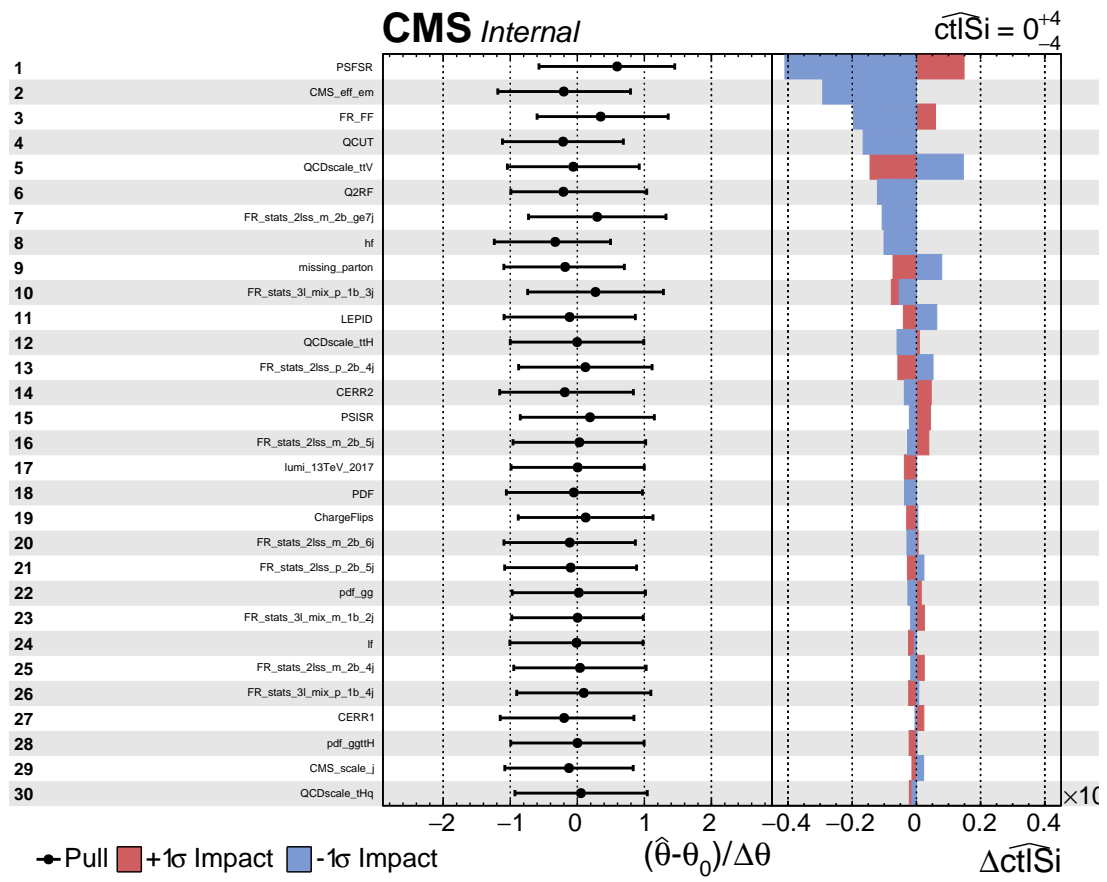


Figure 33: The impacts and pulls of the largest systematics for $c_t^{S(\ell)}$, ranked by impact.

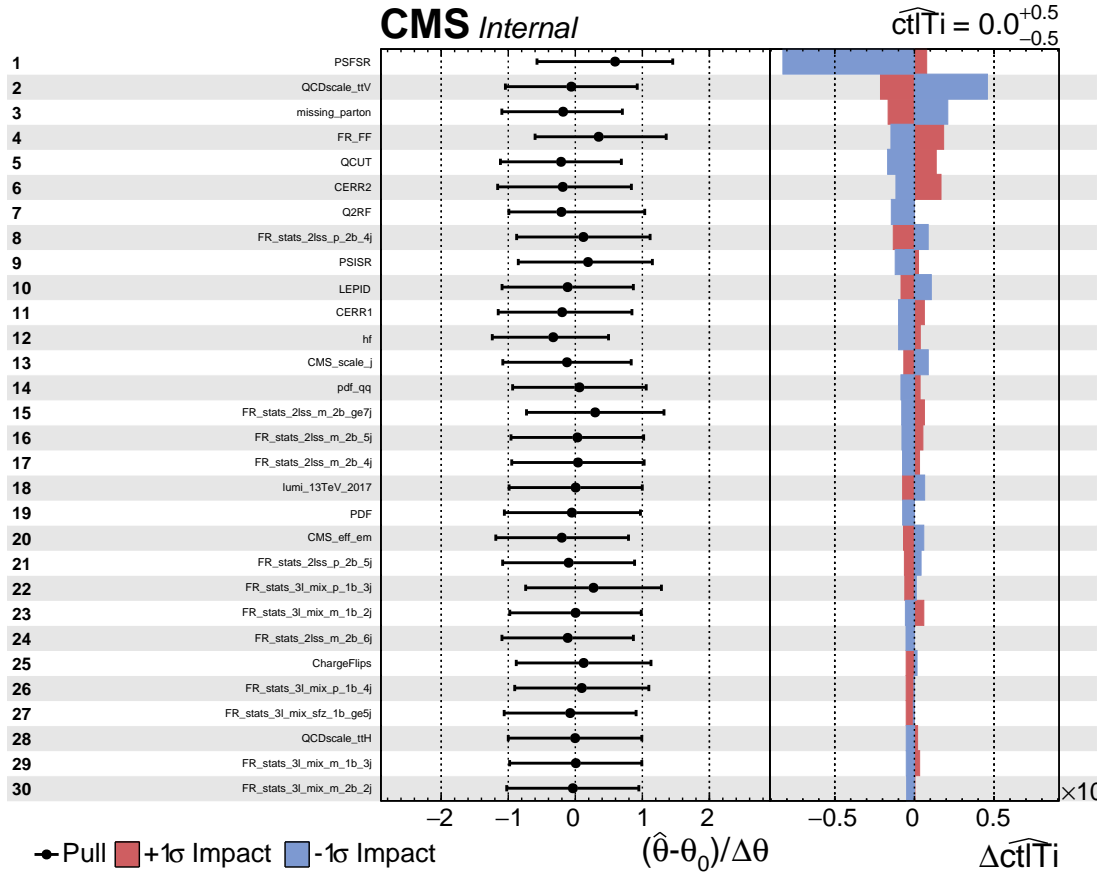


Figure 34: The impacts and pulls of the largest systematics for $c_t^{T(\ell)}$, ranked by impact.

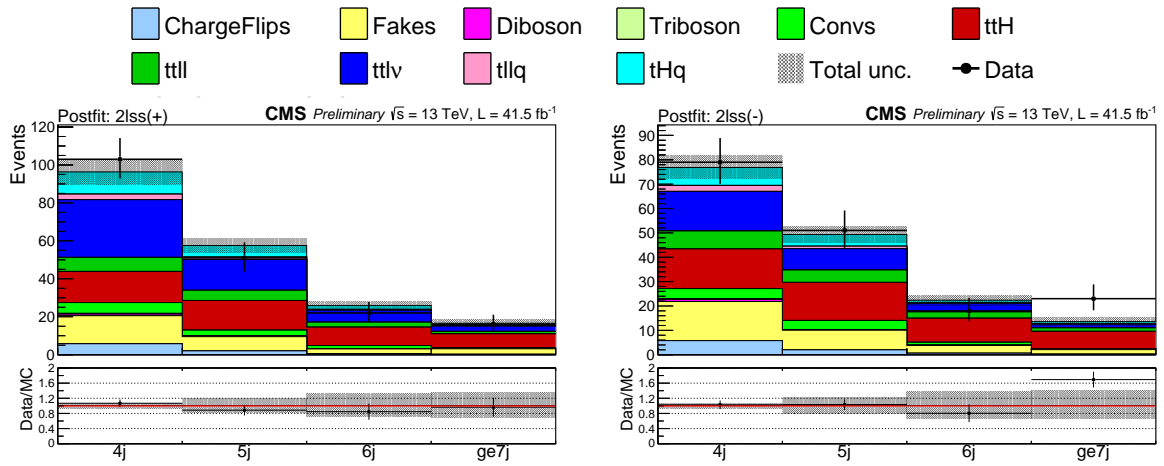


Figure 35: SM postfit yields in the 2lss categories. From left to right, the plots show events in the 2lss (“p”) and 2lss (“m”) categories.

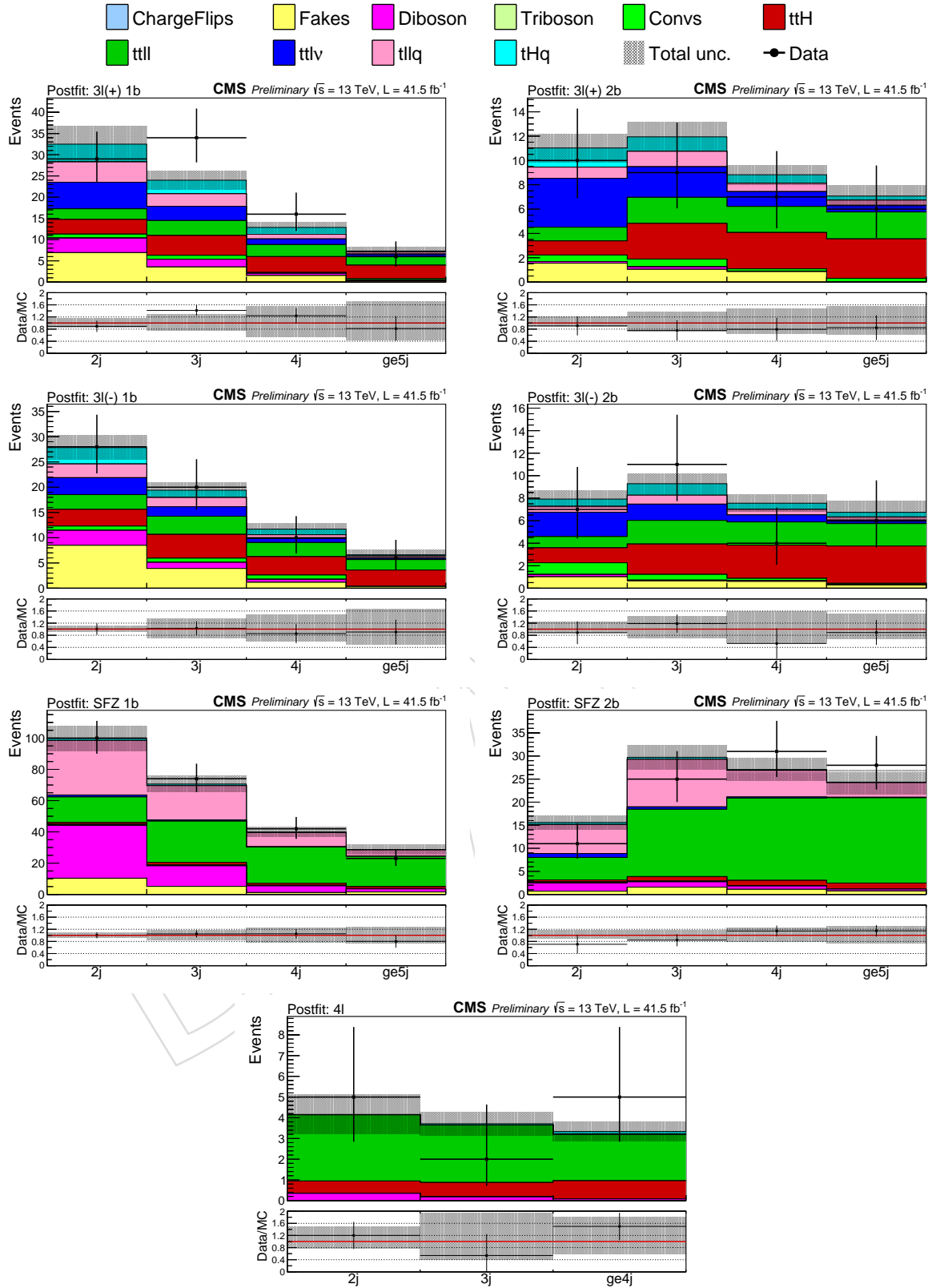


Figure 36: SM postfit yields in the 3ℓ and 4ℓ categories. Going from left to right and top to bottom on a per-row basis, the plots show events in the 3ℓ (1b “p”), 3ℓ ($\geq 2b$ “p”), 3ℓ (1b “m”), 3ℓ ($\geq 2b$ “m”), 3ℓ (SFOS Z, 1b), 3ℓ (SFOS Z, $\geq 2b$) and $\geq 4\ell$ categories, respectively.

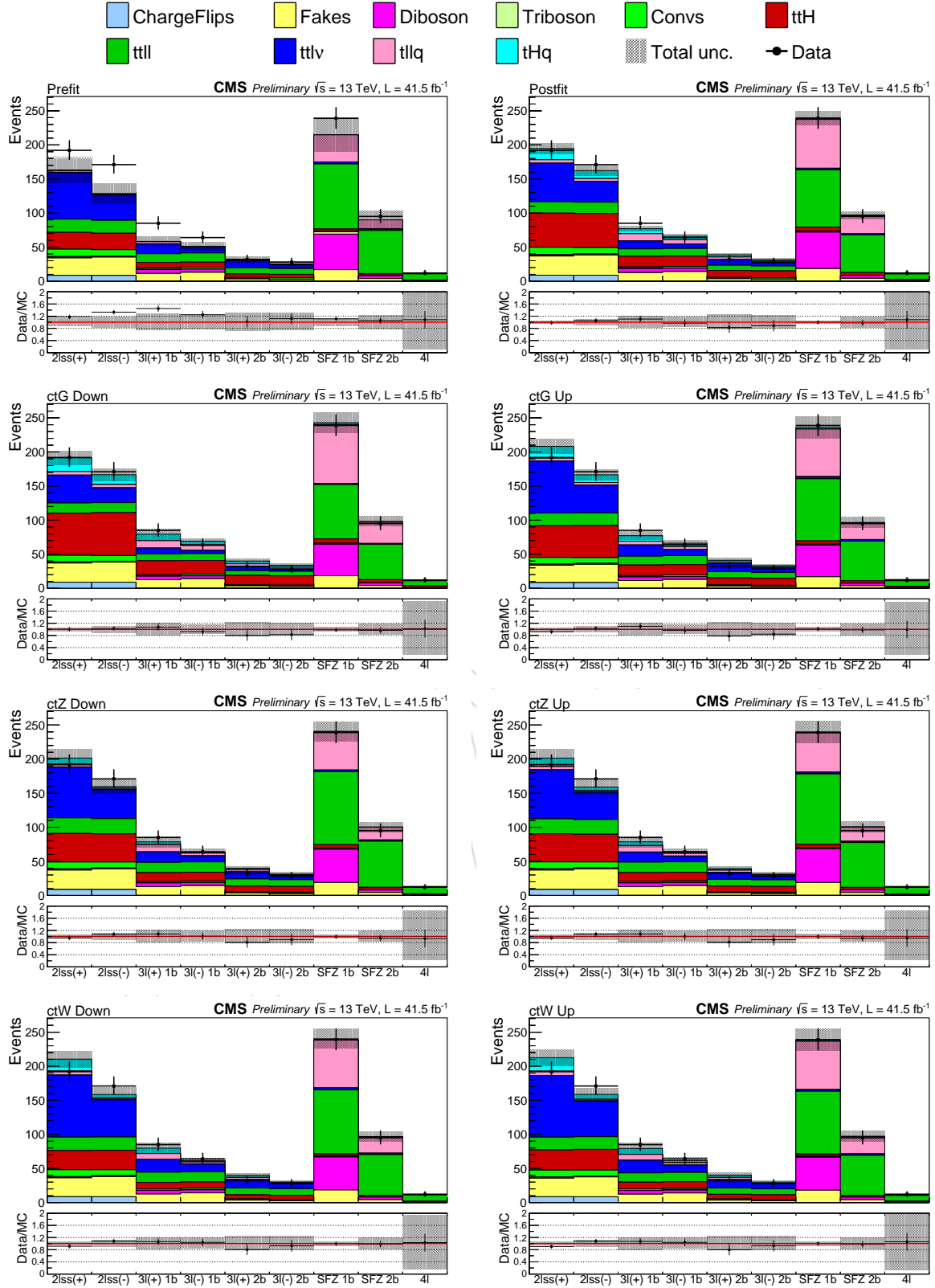


Figure 37: The first row shows the expected yields prefit (left) and postfit (right). From the second row on, the plots show the expected postfit yields, but with one WC set to its -2σ (left) and $+2\sigma$ (right) values. From top to bottom, the individual WCs fluctuated are: ctG, ctZ, and ctW.

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 $t\bar{t}H$ 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 $t\bar{t}H$ 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 $t\bar{t}H$ analysis is also applicable to our analysis, we employ the exact same set of lepton identification criteria, and remain synchronized with the $t\bar{t}H$ analysis on the object-ID level. We compared the yields of muons, electrons and jets selected by running our analysis framework over 56465 $t\bar{t}H$ MC events with those obtained by the $t\bar{t}H$ 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 38 and 39.

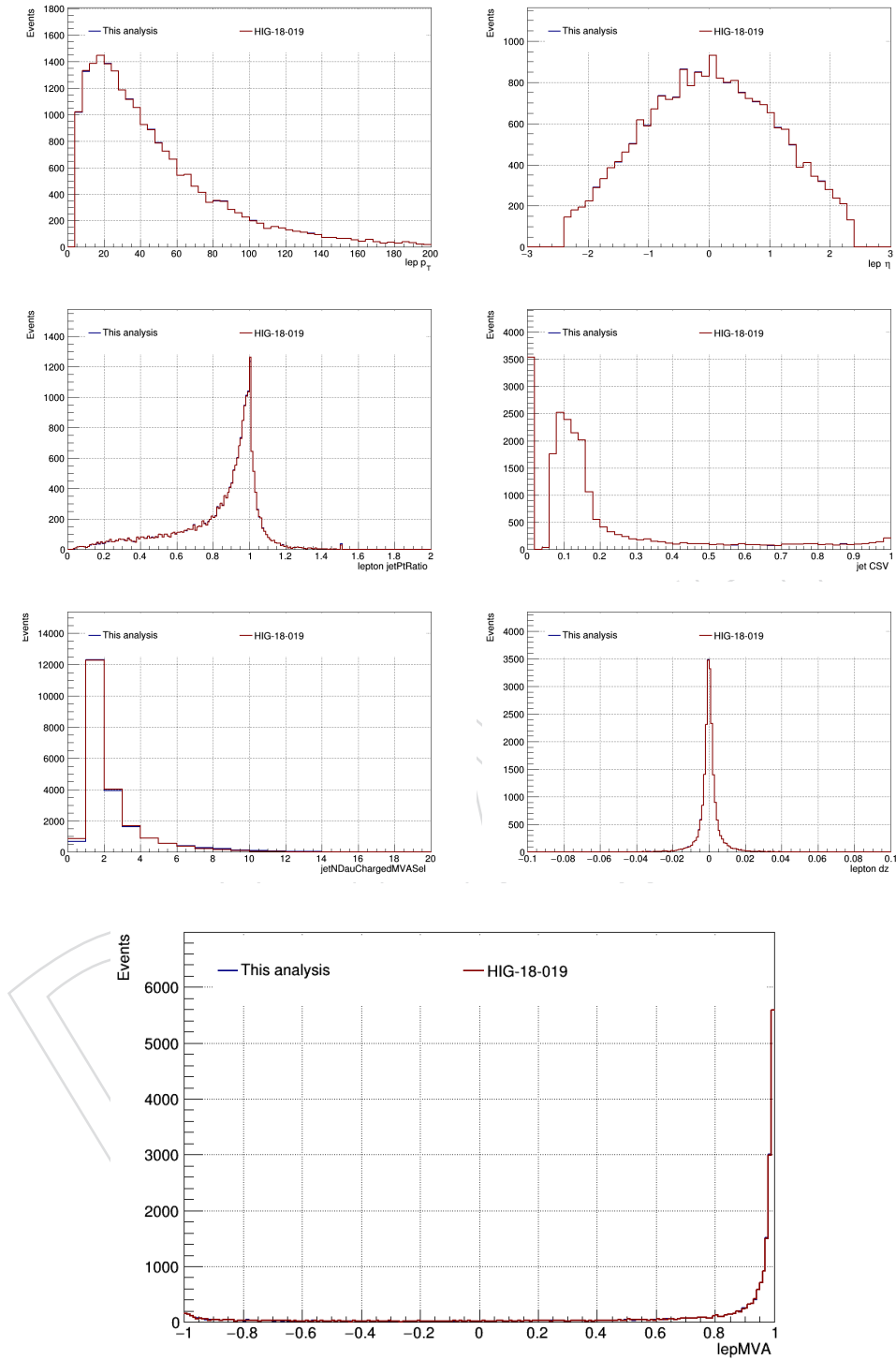


Figure 38: Comparison between distributions used to identify muons obtained from running two different analysis frameworks over the same set of $t\bar{t}H$ MC events. The plots are filled after applying the loose selection cuts listed in table 12, as well as the cross-cleaning between objects described 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).

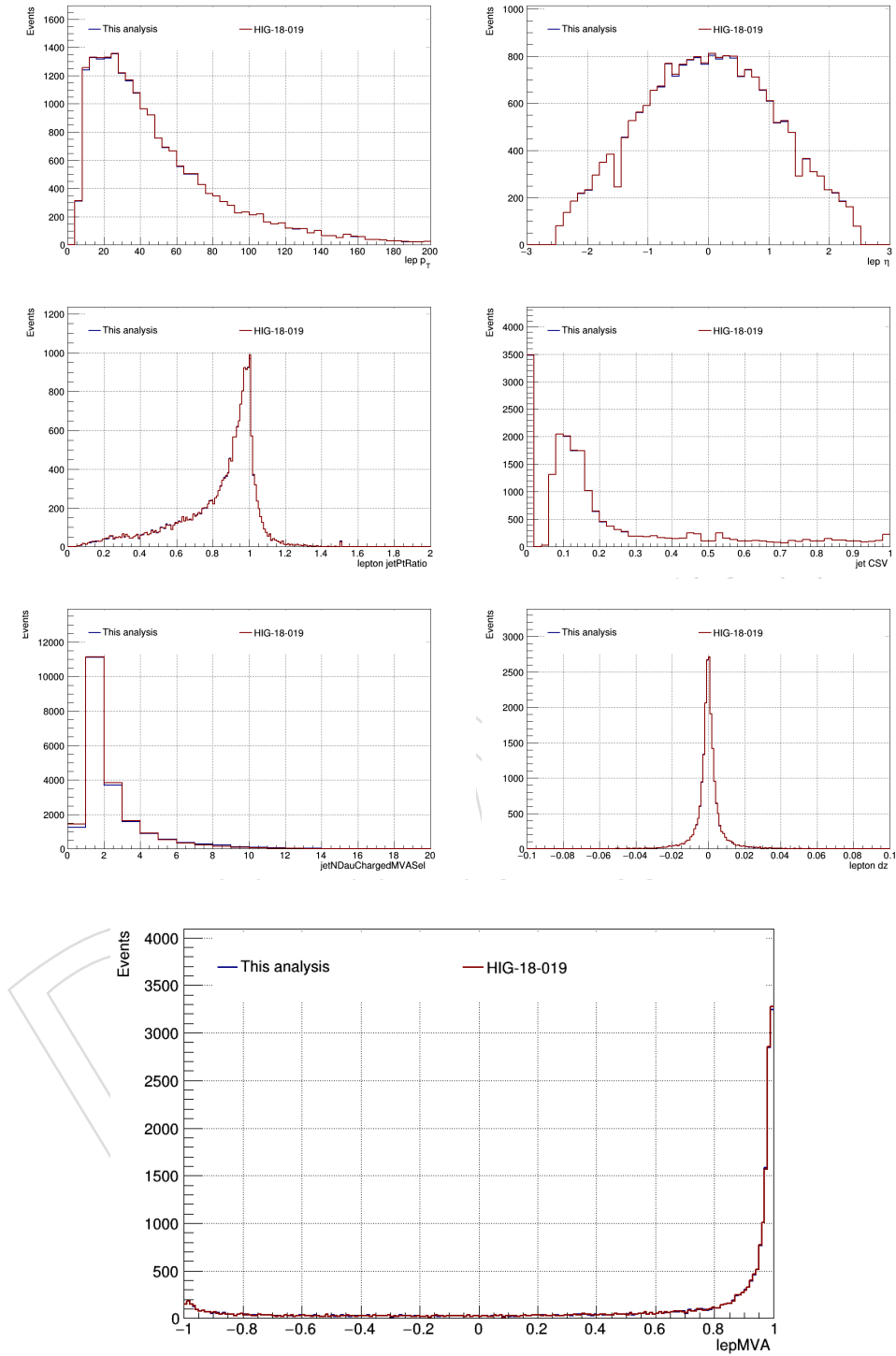


Figure 39: Comparison between distributions used to identify electrons obtained from running two different analysis frameworks over the same set of $t\bar{t}H$ MC events. The plots are filled after applying the loose selection cuts listed in table 11, as well as the cross-cleaning between objects described 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

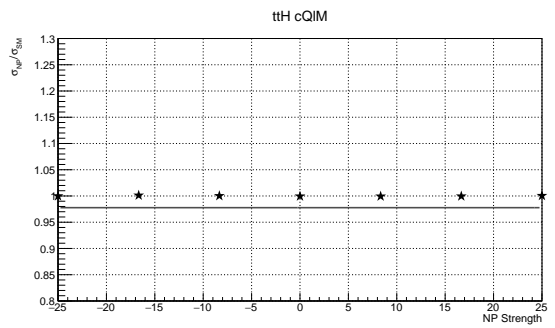
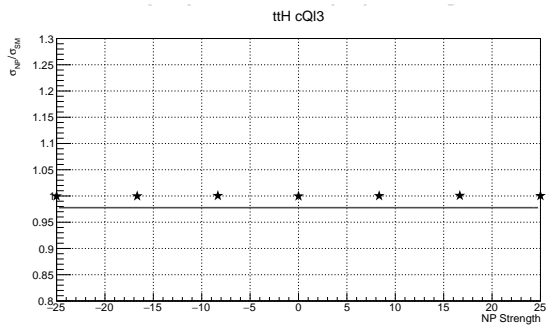
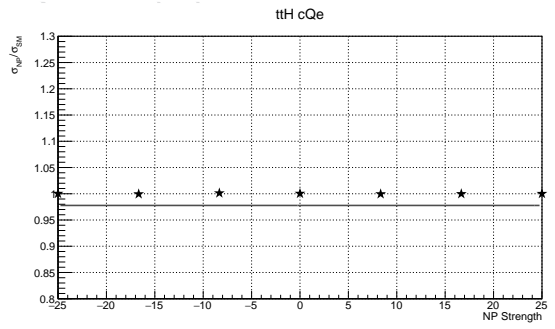
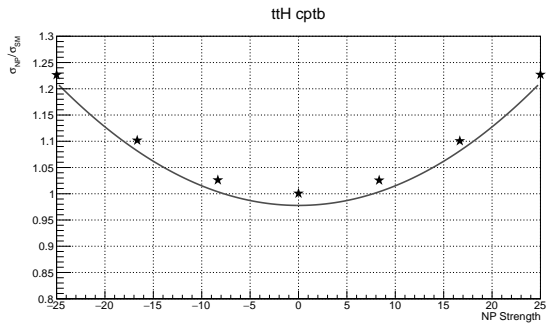
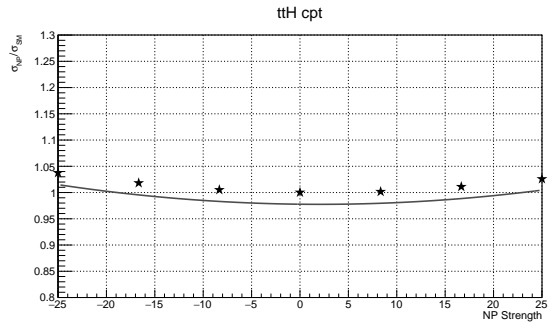
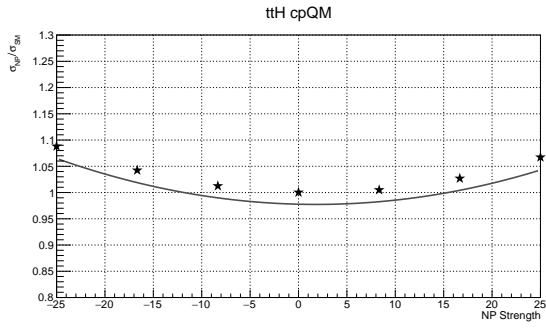
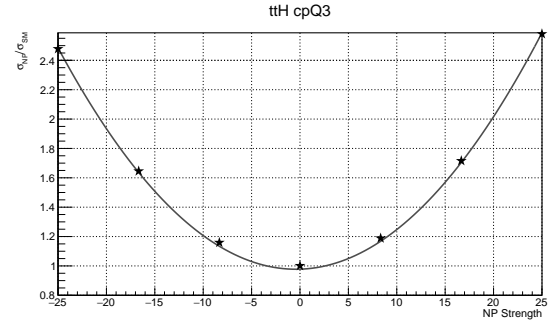
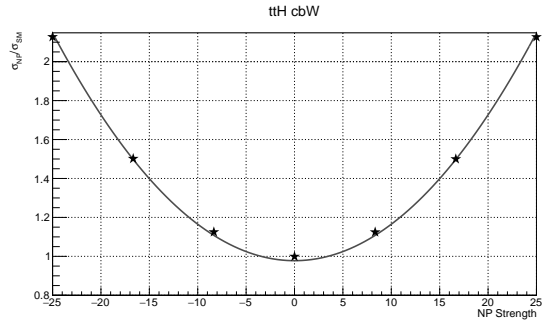
The MC signal samples used in this analysis make use of the MADGRAPH event-by-event reweighting feature. This involves starting with a set of hard events, which are generated for some specific choice of WC strengths. The sample is then rescaled on an event-by-event basis to represent a sample that was 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 40,41,42, and 43 show the comparison between the nominal sample reweighting and the dedicated reference samples.

The samples used to perform this check (Figures 40,41,42, and 43) do not include an extra parton in the final state. To verify that the reweighting procedure still works as expected when an extra parton is included, Figure 44 shows one example comparison between a nominal plus one parton sample reweighting and the dedicated plus one parton reference.

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 jet multiplicity 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 45. The plus one parton private samples are used for this comparison. Reasonably good agreement is seen between the privately-produced reweighted samples and the centrally-produced, non-reweighted samples.

In order to populate as much of the relevant phase space as possible, each signal sample was generated at a non-SM point. The starting point was randomly selected from a range of reasonable values in WC space. We tested several potential starting points and compared the quadratic parameterizations of each. While all showed largely consistent quadratic fits, some starting points resulted in a better distribution of event weights when the samples were reweighted to the SM. For each sample, we chose a starting point that resulted in a reasonable distribution of weights at the SM. This resulted in the $t\bar{t}H$ and $t\bar{t}lq$ samples being generated at the same point, the $t\bar{t}l\bar{l}$ and tHq samples being generated at the same point, and the $t\bar{t}l\nu$ sample being generated at a different point from the other samples. The summary of starting points for all full production samples is shown in Table 25.



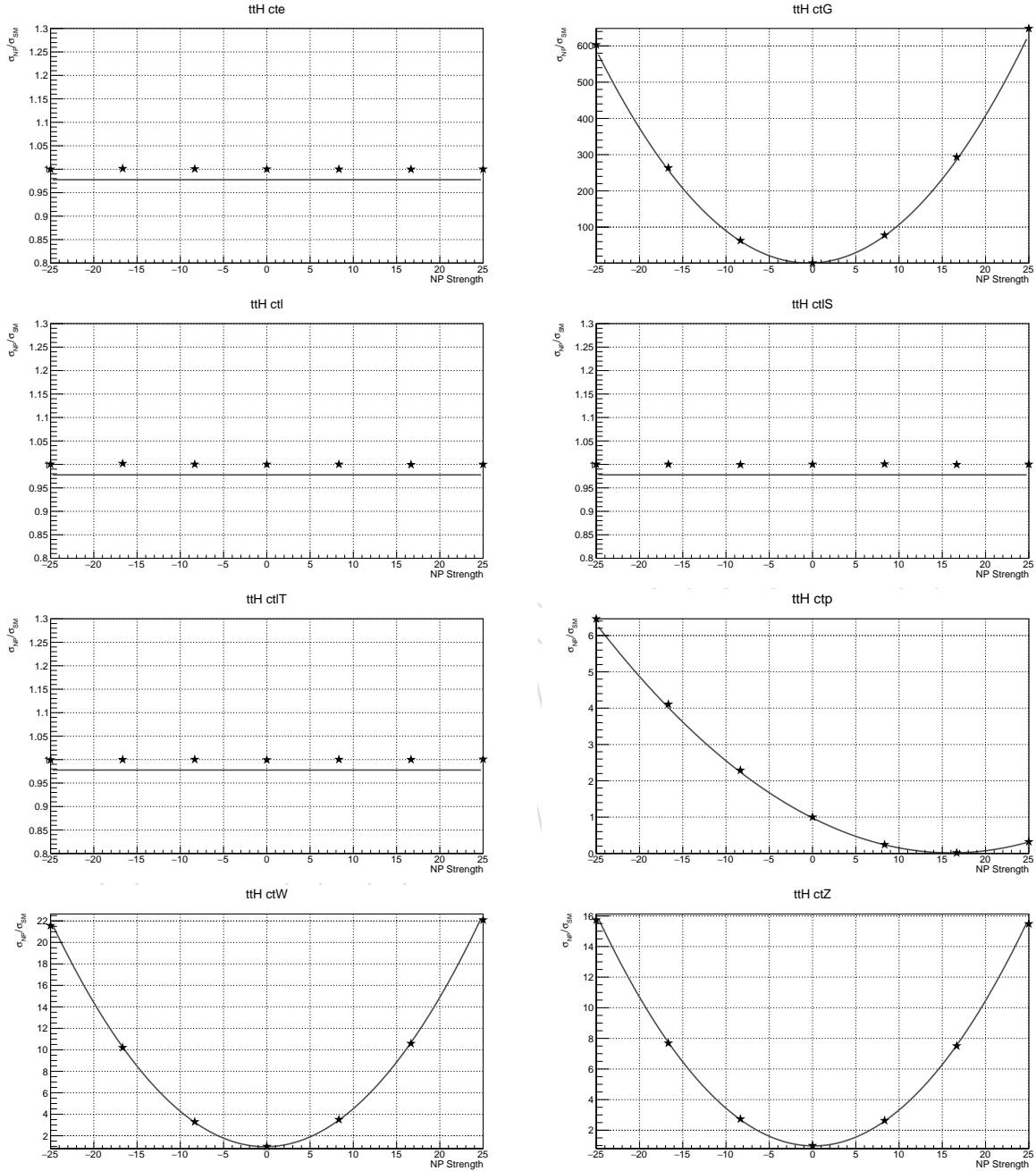
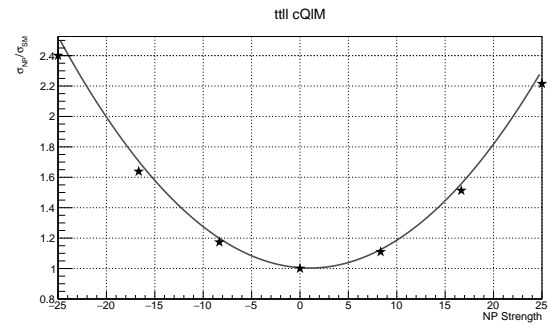
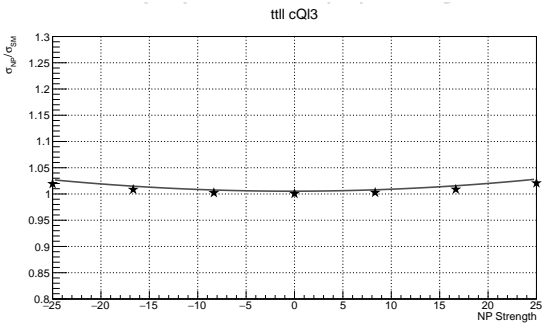
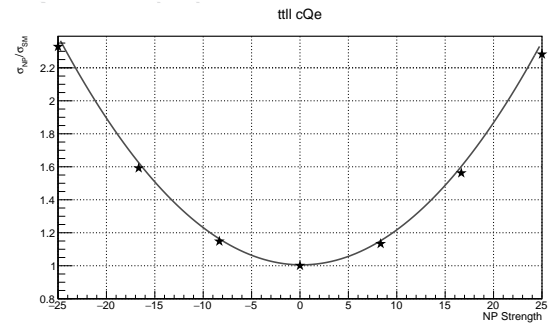
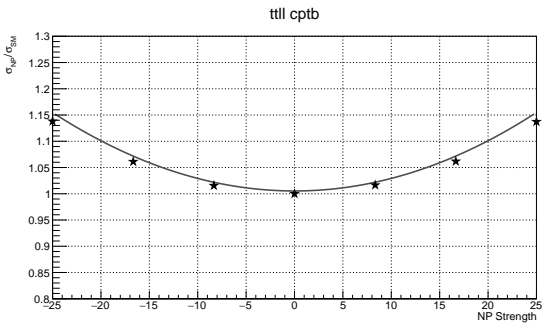
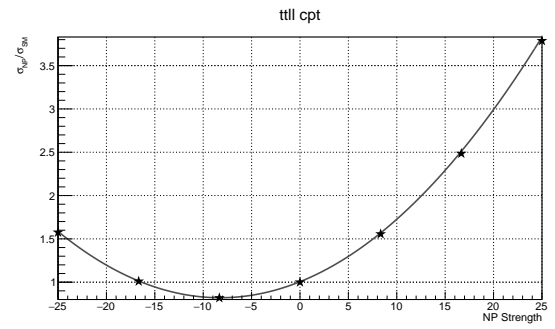
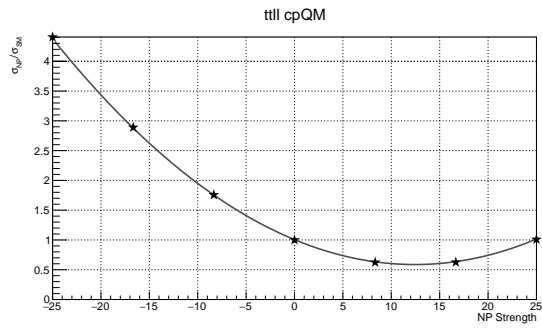
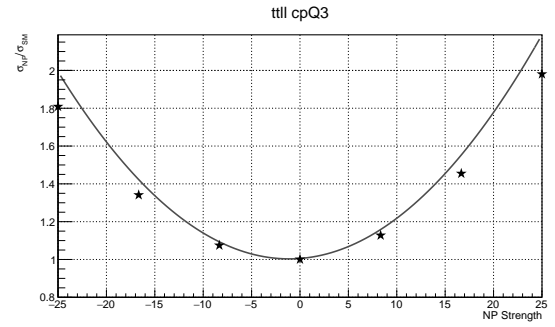
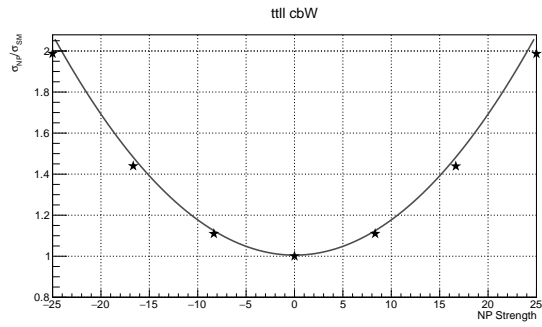


Figure 40: Comparison of the WC parameterization fit for $t\bar{t}H$. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



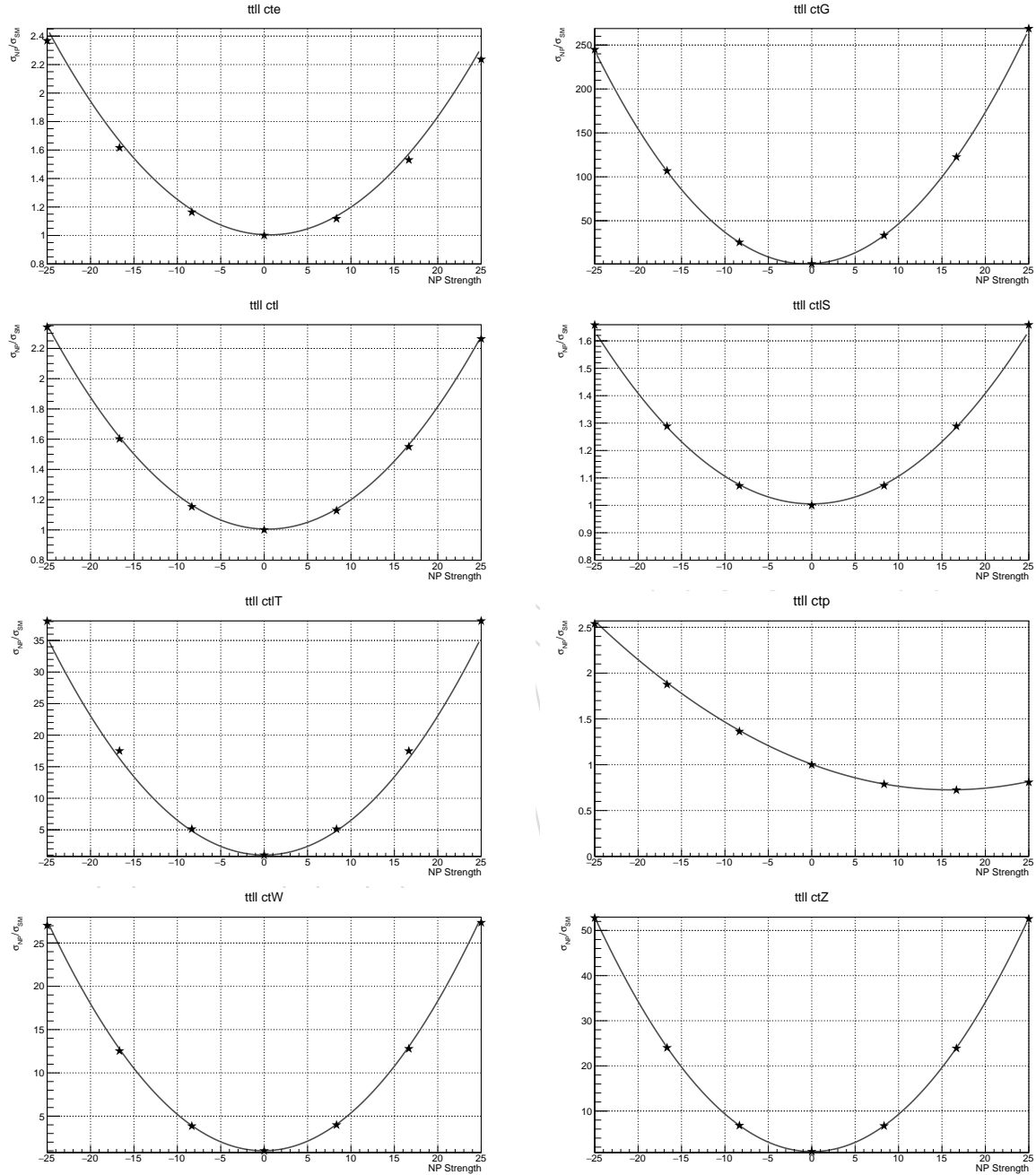
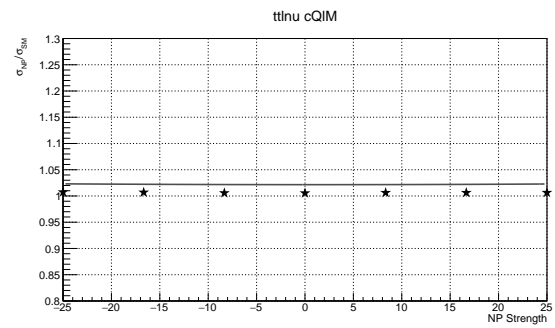
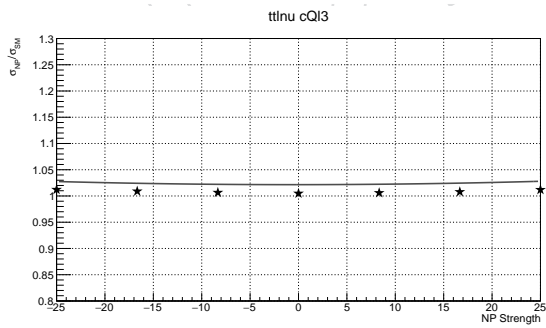
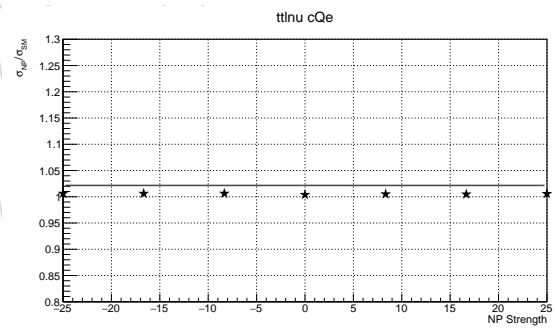
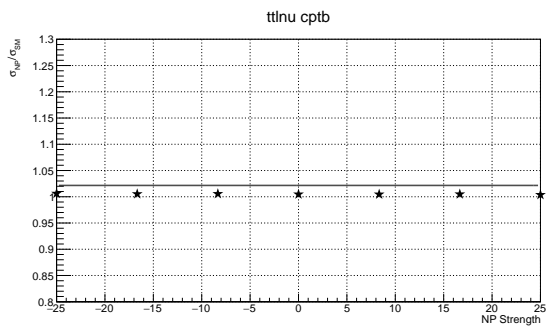
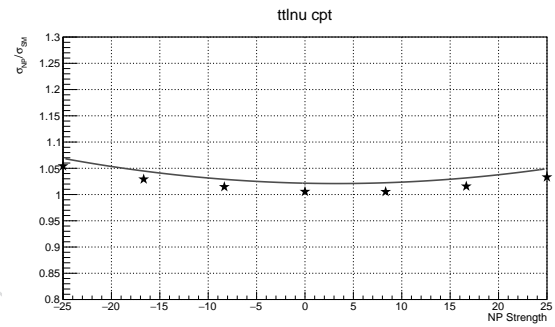
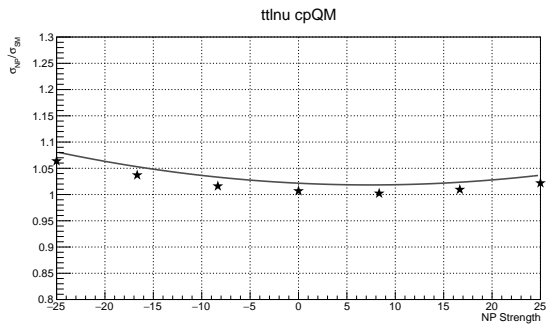
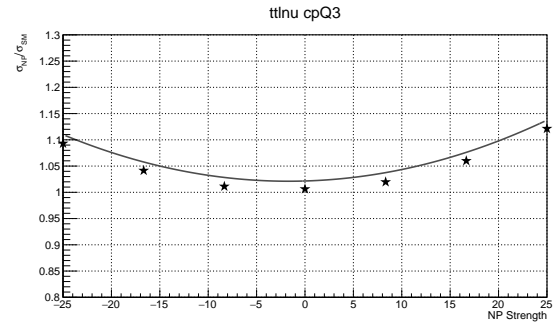
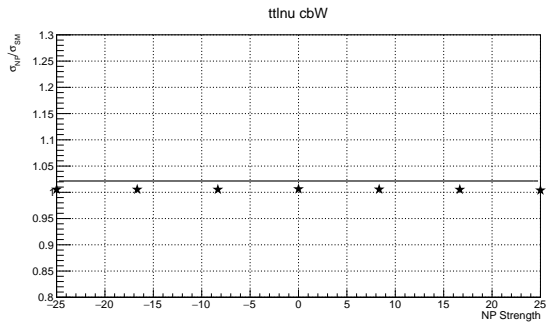


Figure 41: Comparison of the WC parameterization fit for ttll. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



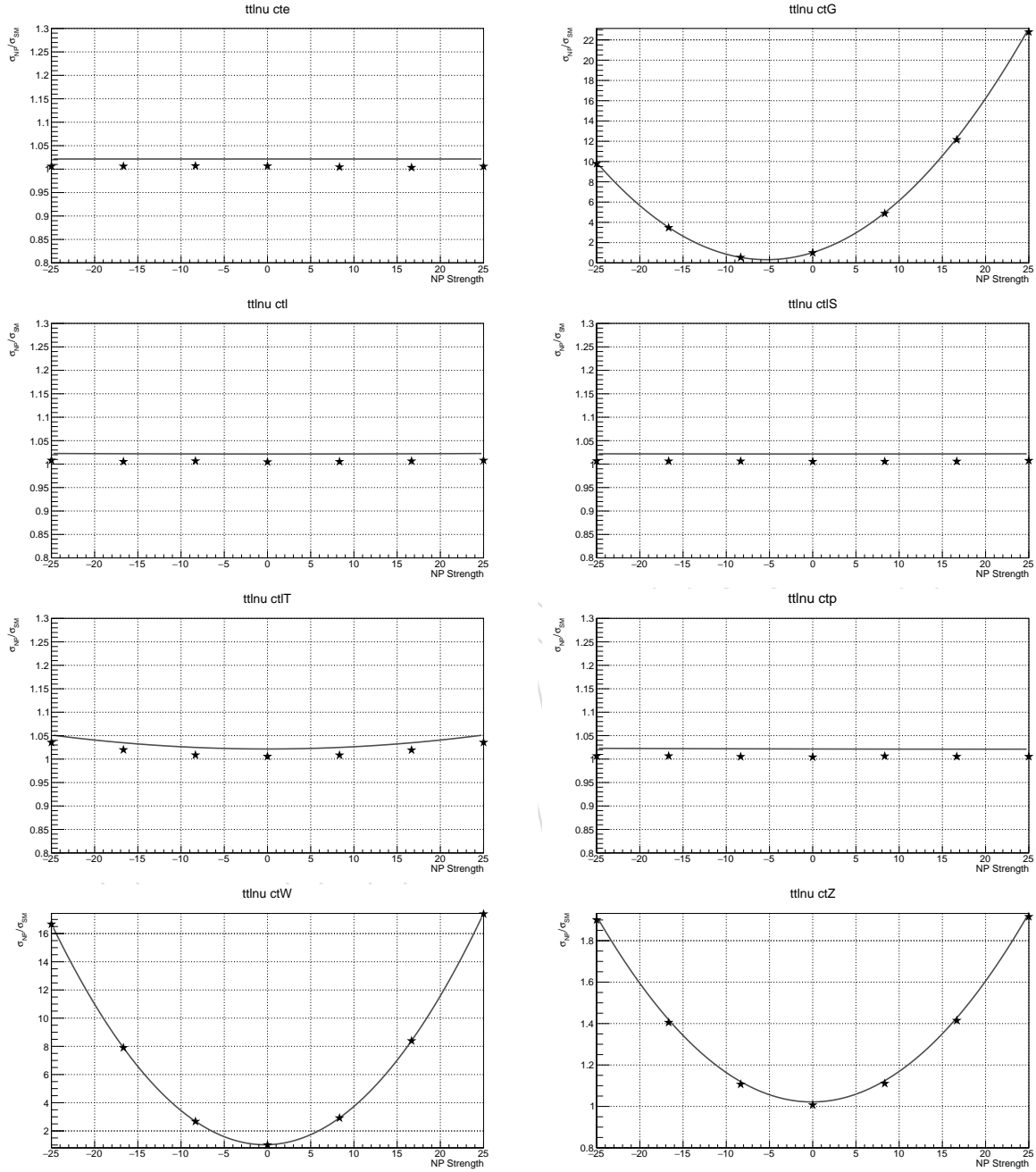
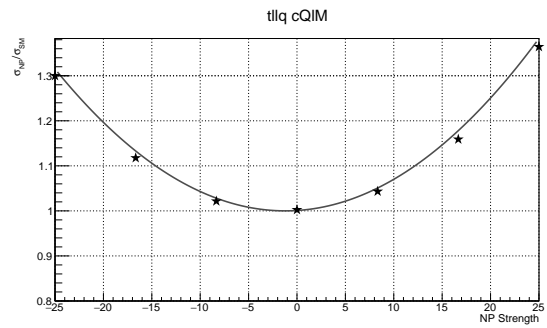
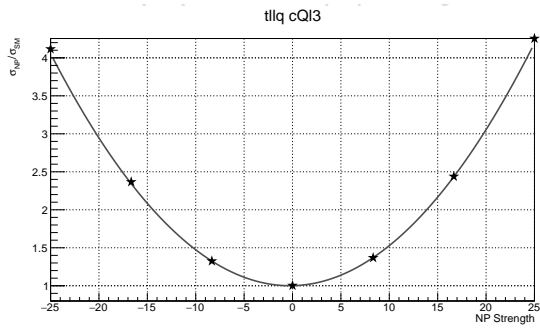
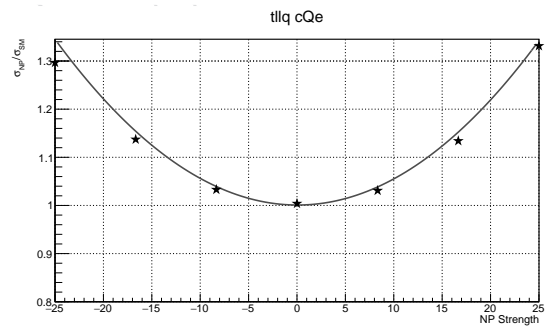
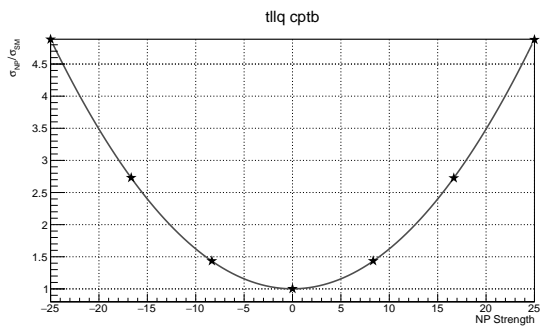
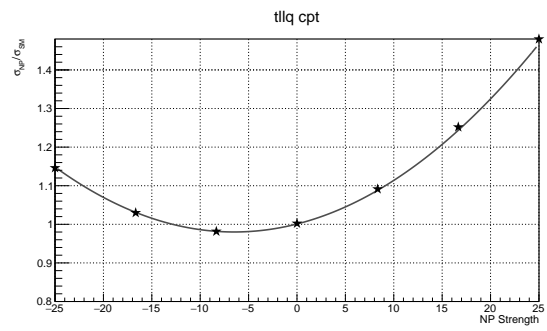
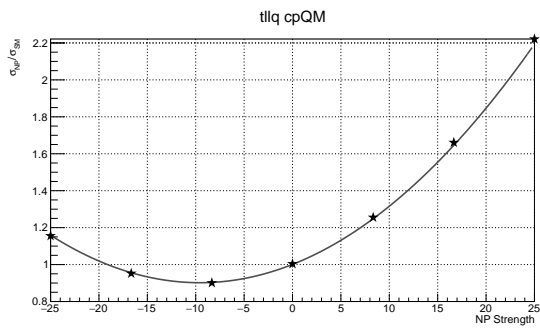
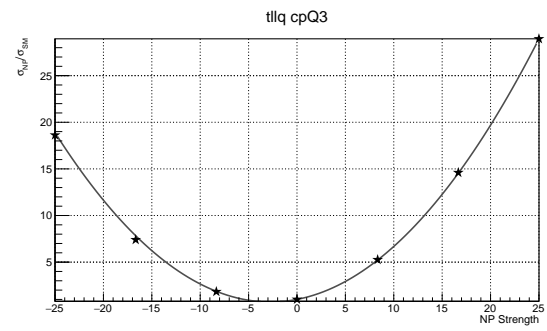
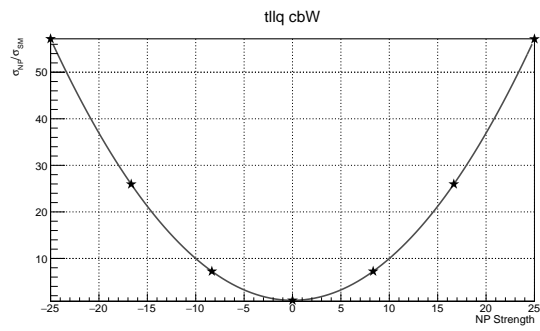


Figure 42: Comparison of the WC parameterization fit for $tt\bar{\nu}$. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.



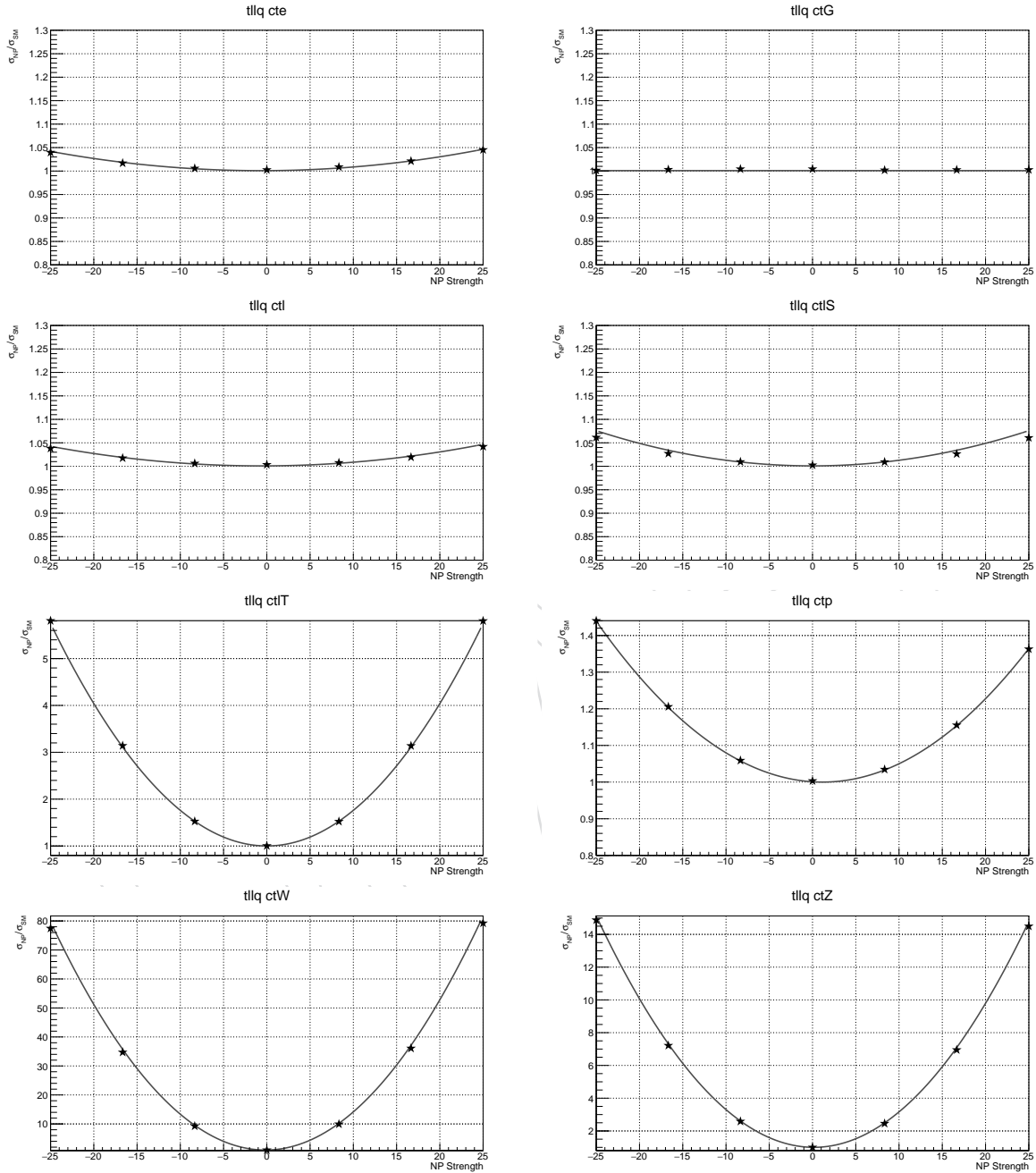


Figure 43: Comparison of the WC parameterization fit for tllq. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The same fit is used in each plot.

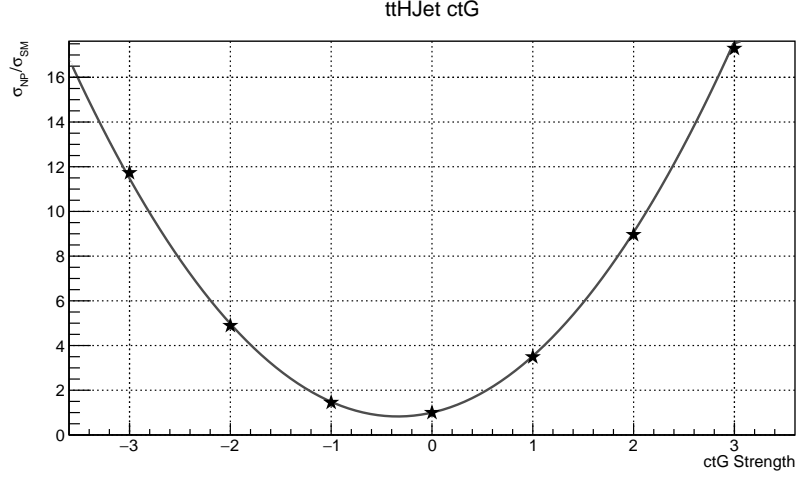


Figure 44: Comparison of the WC parameterization fit for the $t\bar{t}H$ sample that includes an extra parton. The black stars are reference points calculated from separate MG samples generated at those specific points in the WC phase space. The plot is normalized such that the NP cross section is equal to the SM cross section when the value of ctG is set to zero.

WC	$t\bar{t}H$ start pt	$t\bar{t}l\nu$ start pt	$t\bar{t}ll$ start pt	$t\bar{t}lq$ start pt	tHq start pt
ctW	-8.30	-3.82	-5.02	-8.30	-5.02
ctp	64.33	51.50	32.91	64.33	32.91
cpQM	45.88	23.00	-8.06	45.88	-8.06
ctei	24.32	8.938	6.003	24.32	6.003
ctli	24.43	-7.00	10.16	24.43	10.16
cQei	23.75	8.968	4.804	23.75	4.804
ctZ	-6.09	5.727	-3.86	-6.09	-3.86
cQlMi	23.95	6.952	-7.15	23.95	-7.15
cQl3i	21.54	9.243	-8.33	21.54	-8.33
ctG	-3.60	2.430	1.606	-3.60	1.606
ctlTi	21.80	2.116	2.824	21.80	2.824
cbW	49.59	-7.37	-3.82	49.59	-3.82
cpQ3	-51.1	-14.4	-13.2	-51.1	-13.2
cptb	136.1	-21.8	14.49	136.1	14.49
cpt	-43.5	-20.3	-32.6	-43.5	-32.6
ctlSi	-20.0	-9.99	-7.06	-20.0	-7.06

Table 25: Summary of starting points for all full production samples

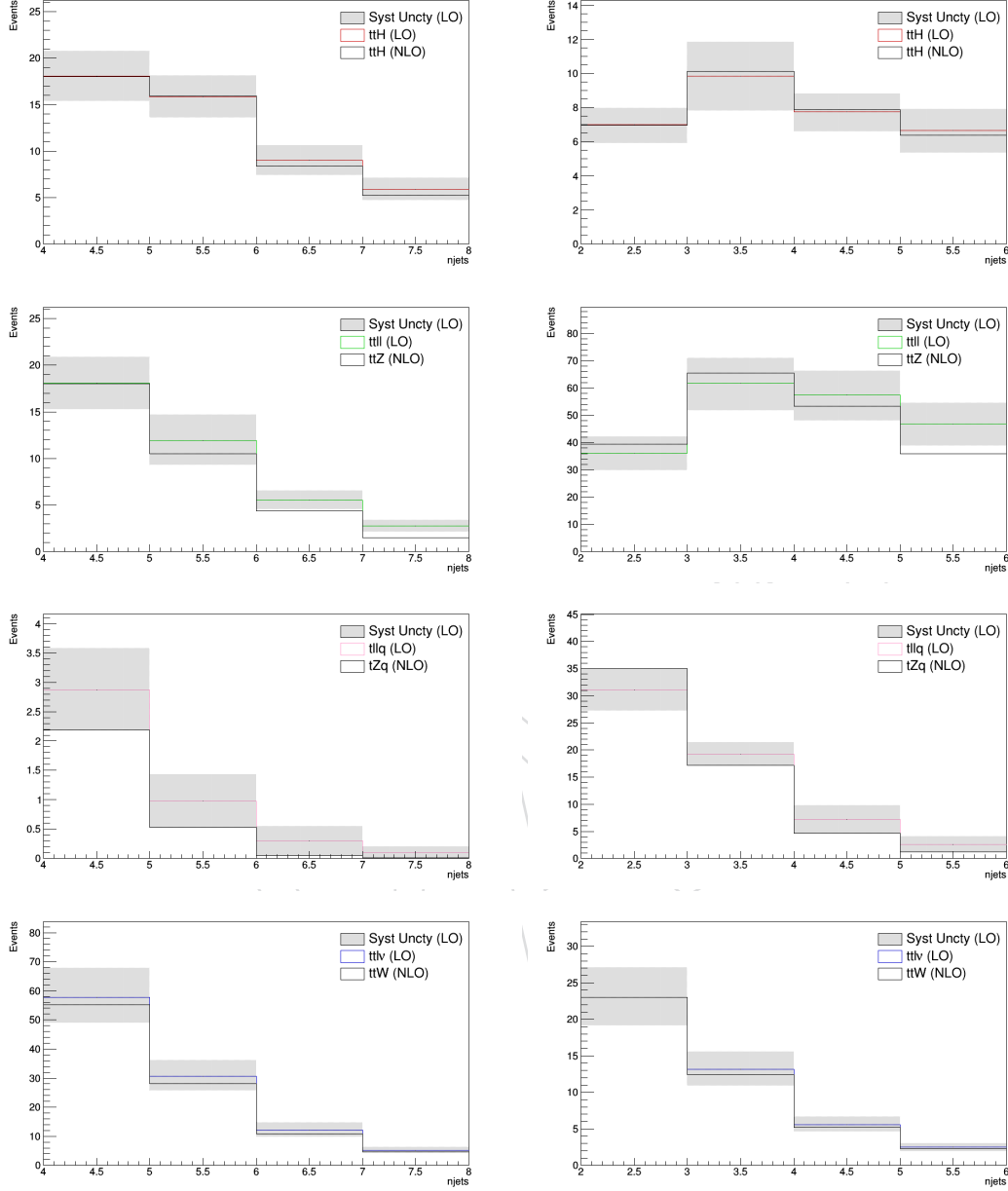


Figure 45: Comparison between centrally-produced NLO MC samples and LO (plus one parton) privately-produced MC samples with EFT weights, showing jet multiplicity distributions in the signal regions. Events are scaled to the SM expectation; in the case of the LO samples, a k-factor is used to adjust (uniformly and simultaneously across all categories) for the difference between the NLO and LO inclusive SM cross sections. Left column: SM expected yields in combined $2\ell_{ss}$ categories. Right column: SM expected yields in combined 3ℓ categories. The error band centered on the LO sample prediction shows the combined systematic uncertainty on the yield. Top to bottom, the rows show the LO(NLO) SM expectation for $t\bar{t}H$ ($t\bar{t}H$), $t\bar{t}l\bar{l}$ ($t\bar{t}Z$), $t\bar{t}lq$ ($t\bar{t}Zq$) and $t\bar{t}l\nu$ ($t\bar{t}W$), respectively. Differences between LO and NLO are expected, but generally fall within the uncertainty of the yield prediction.

C Systematics Checks

A potential concern when estimating the uncertainty contribution coming from certain nuisance parameters is whether or not the estimation of the size of that uncertainty is dependent on the choice of the EFT parameters used to generate the events in the MC. This is a concern for this analysis since EFT reweighting is used in order to search a very large region of phase space, but the weights used to estimate the uncertainty of certain systematics are generated with respect to the original EFT phase space starting point used to generate the events.

In order to estimate the impact of choosing a particular EFT starting point, an additional set of MC events is produced for each of the signal samples considered in this analysis. These samples have much lower statistics $O(200k)$ and do not include any EFT weights, so no EFT reweighting can be done. The samples are otherwise processed identically as the larger statistics samples used in the analysis proper. Three sets of starting points are chosen: The SM point, where all WCs are set to zero; EFT lower, where all WCs are set to their -2σ limits; EFT Upper, where all WCs are set to their $+2\sigma$ limits.

The systematics considered here are the uncertainties on the renormalization/factorization scale (Q2RF), as well as the the PDF shape uncertainties. Fig. 46 shows a summary of the differences for these uncertainties in each of the signal processes and averaged over all bins in the analysis when using events produced from different EFT phase spaces. Fig. 47 shows the uncertainties for a particular lepton category and broken down by njet bins. Overall, the choice of EFT starting point does not greatly affect the size of the uncertainties and also does not seem to have a bias tied to any given starting point. It is therefore reasonable to assume that these systematics do not have a strong dependence on the choice of WCs used to generate the template variations and that the weights from the main large statistics samples are safe to be used throughout the analysis.

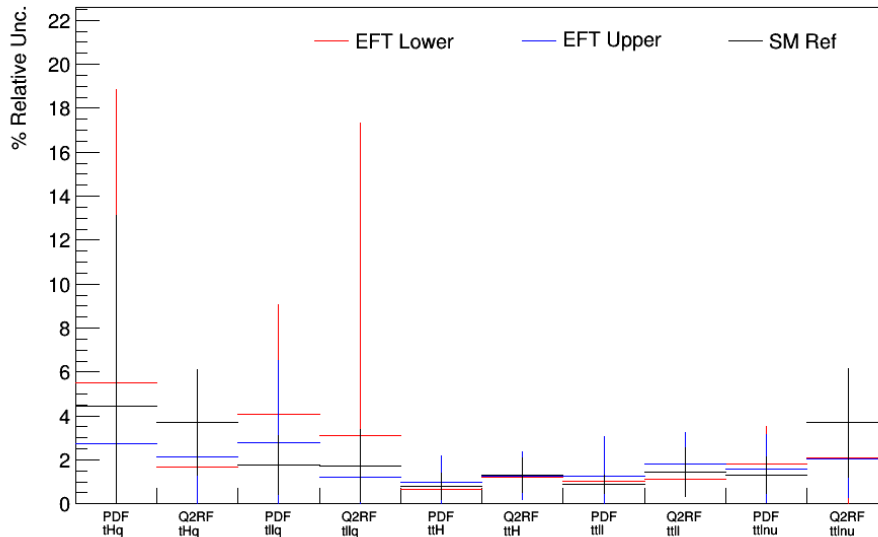


Figure 46: The percent relative uncertainty in the various signal process for the renormalization and factorization scale, as well as the PDF shape systematics. The relative uncertainties shown here come from taking the simple mean over all category bins for a given process in the analysis. The error bars shown here correspond to the RMS of values used to calculate the mean.

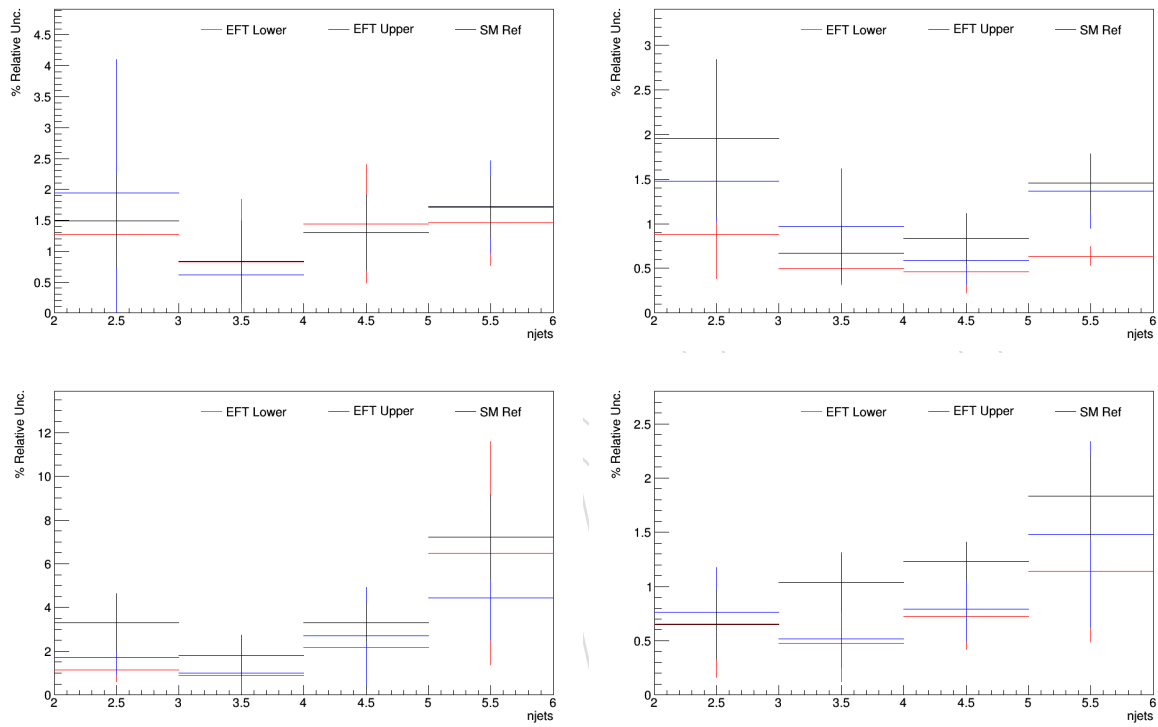


Figure 47: The percent relative uncertainty in the various signal process for the renormalization and factorization scale (Q2RF), in the 3ℓ lepton category and broken down by n_{jets} . From top to bottom and left to right the plots are for $t\bar{t}H$, $t\bar{t}l\bar{l}$, $t\bar{t}l\nu$, and $t\bar{t}lq$. The relative uncertainties shown here come from taking the simple mean over all 3ℓ categories for a given process in the analysis. The error bars shown here correspond to the RMS of values used to calculate the mean.

D 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 analyses of the LHC data. We have chosen the most recent results of a global EFT fit for doing a comparison of the obtained limits [64]. In Ref. [64], 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 48, we have compare the limits obtained in this analysis to the limits obtained in [64]. There are no direct bounds on two leptons - two quarks operators.

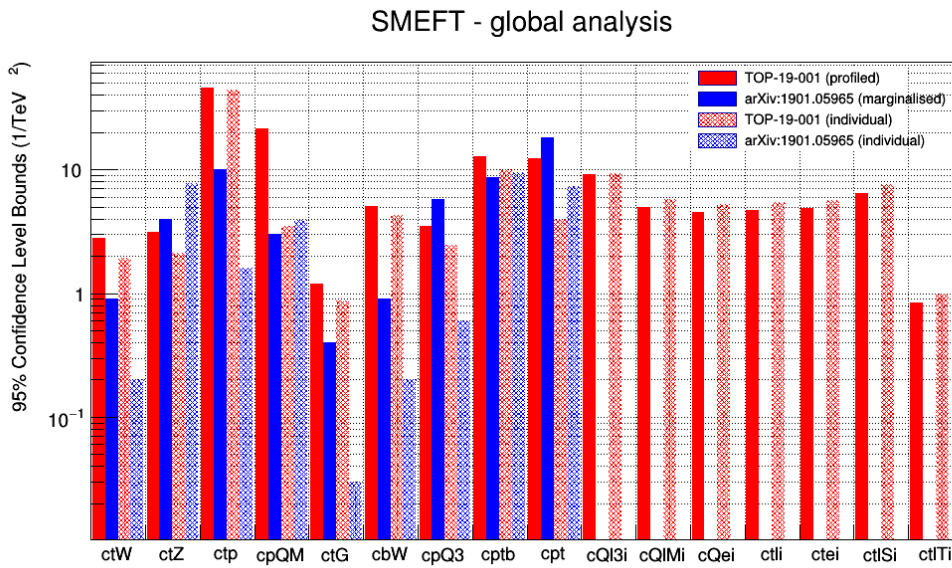


Figure 48: Limits obtained in this analysis (red) are compared to the limits obtained in Ref. [64] (blue). Filled histograms show the limits when other couplings are float and hatched histograms show the limits with other couplings are set to zero.

One should taken into account the following points before comparing our results to the results in Ref. [64].

- In Ref. [64], 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. [64], results are observed while our results are expected
- In Ref. [64], effects of EFT operators on backgrounds are ignored.
- In Ref. [64], correlation between various measurements (e.g. $t\bar{t}Z$, tZq , ...) are ignored.

In addition, some of the EFT operators considered in this analysis are probed in other channels by the CMS collaboration [65, 66]. In Figure 49, expected limits obtained in this analysis for each operator at a time (individual) are compared to other CMS analyses.

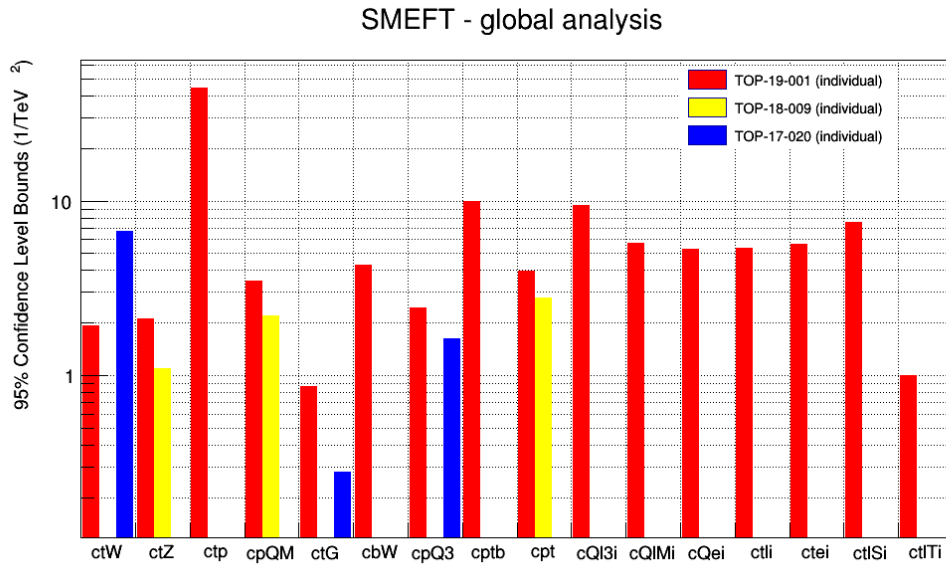


Figure 49: Limits obtained in this analysis (red) are compared to the limits obtained in other CMS analyses [65, 66].

E Additional plots requested during review

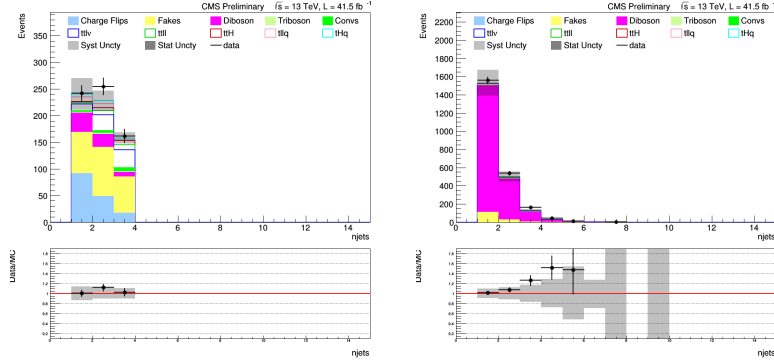


Figure 50: We looked at various methods of extending the control region (CR) plots of Figs. 4 and 5. Adding the 3-jet bin to the $2\ell_{\text{ss}}$ control region was investigated, but was found to have high signal contamination, so this was not adopted (left plot). Methods of forming a non-prompt CR 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 b tag 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] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, “Dimension-Six Terms in the Standard Model Lagrangian”, *JHEP* **10** (2010) 085, doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.
- [6] 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.
- [7] C. Collaboration, “PdmV 2017 Analysis Recipes”.
<https://twiki.cern.ch/twiki/bin/view/CMS/PdmV2017Analysis>.
- [8] 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.
- [9] 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.
- [10] J. Kim, ““Modified dim6TopEFT Model” [software], commit 05a585de765583b0194a622caf4ab476d3eb60ed”.
- [11] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [12] 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.
- [13] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [14] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.
- [15] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J.* **C53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [16] 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.

- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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).
- [28] 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.
- [29] 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.
- [30] EGamma POG Collaboration.
<https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2#Recommended.M>

- [31] A. Hocker et al., “TMVA - toolkit for multivariate data analysis”, *PoS ACAT* (2007) 040, arXiv:physics/0703039.
- [32] 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.
- [33] Muon POG Collaboration.
https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Medium_Muon.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] C. Botta, M. Peruzzi, and G. Petrucciani.
<https://indico.cern.ch/event/446448/contributions/1952877/attachments/1161380/1672225/lepaware>
- [39] 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.
- [40] 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.
- [41] JetMET POG. <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2017>.
- [42] 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.
- [43] 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.
- [44] A. Tiko.
<https://indico.cern.ch/event/606232/contributions/2444209/attachments/1398539/2133074/chargefli>
- [45] .
- [46] .
- [47] T. Boccali et al.
<https://indico.cern.ch/event/695872/contributions/2877123/attachments/1593469/2522749/pileup-p>
- [48] 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).
- [49] BTV POG Collaboration. <https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods>.

- [50] L. T. Group.
<https://twiki.cern.ch/twiki/bin/view/CMS/L1ECALPrefiringWeightRecipe>.
- [51] 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).
- [52] B. Stieger et al., “Search for ttH in multilepton final states at 13 TeV”, *CMS Analysis Note* **2015/321** (2016).
- [53] C. Mueller et al., “Search for ttH in multilepton final states with 2016 data”, *CMS Analysis Note* **2016/211** (2016).
- [54] C. H. project, ““Combine Harvester” [software], commit 4adba2b7fcf993ba1304fbddb66abae197549880”,..
- [55] H.-C. project, ““HiggsAnalysis-CombineLimit” [software], commit a1dc956452f5b8e055de43bfb093e5243a1e99d9”,..
- [56] L. Moneta et al., “The RooStats Project”, *PoS ACAT2010* (2010) 057, doi:10.22323/1.093.0057, arXiv:1009.1003.
- [57] Muon POG Collaboration.
<https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017>.
- [58] EGamma POG Collaboration.
https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale.
- [59] JetMET POG.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections>.
- [60] 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).
- [61] 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.
- [62] 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.
- [63] 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.
- [64] N. P. Hartland et al., “A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector”, arXiv:1901.05965.
- [65] 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.
- [66] 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.