SEARCH FOR CHARGED-LEPTON FLAVOR VIOLATION IN THE PRODUCTION AND DECAY OF TOP QUARKS AT $\sqrt{S} = 13$ TEV WITH THE CMS DETECTOR

JINGYAN LI
B.S. in Physics, Lanzhou University

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS
NORTHEASTERN UNIVERSITY

2023

Abstract

A search for charged-lepton flavor violation has been performed in the top quark sector through both top quark production and decay signal processes. The data were collected by the CMS experiment from proton-proton collisions at a center-of-mass energy of 13 TeV and correspond to an integrated luminosity of 138 fb^{-1}. The selected events are required to contain one opposite-sign electron-muon pair, a third charged lepton (electron or muon), at least one jet, and at most one jet associated with a bottom quark. The analysis utilizes boosted decision trees to separate background processes from a possible signal, exploiting differences in the kinematics of the final state particles. The data are found to be consistent with the standard model expectation. Upper limits at 95% confidence level are placed on the branching fractions involving up (charm) quarks, $t \to e\mu u$ ($t \to e\mu c$), of 0.032×10^{-6} (0.498×10^{-6}), 0.022×10^{-6} (0.369×10^{-6}), and 0.012×10^{-6} (0.216×10^{-6}) for tensor, vector, and scalar interactions, respectively.

Contents

List of Figures

List of Tables

List of Acronyms

AR Application Region

ATLAS A Toroidal LHC Apparatus

BDT Boosted Decision Tree

CL Confidence Level

CLFV Charged-Lepton Flavor Violation

CMS Compact Muon Solenoid

EFT Effective Field Theory

HLT High Level Trigger

HL-LHC High Luminosity-LHC

JES Jet Energy Scale

JER Jet Energy Resolution

LHC Large Hadron Collider

LO Leading Order

L1 Level-1

MC Monte-Carlo

ME Matrix-Element

MET Missing Transverse Momentum

MR Measurement Region

MVA Multivariate Analysis

NLO Next-to-Leading Order

NN Neural Network

NNLO Next-to-Next-to-Leading order

OSDF Opposite-Sign and Different-Flavor

OSSF Opposite-Sign and Same-Flavor

PD Primary Dataset

PDF Parton Distribution Function

PF Particle Flow

PS Parton Shower

PU Pile-Up

PV Primary Vertex

QCD Quantum Chromodynamics

ROC Receiver Operating Characteristic

SM Standard Model

SR Signal Region

SSSF Same-Sign and Same-Flavor

SV Secondary Vertex

VR Validation Region

WC Wilson Coefficient

Chapter 1

Introduction

Part I

Theoretical Framework

Chapter 2

Electroweak Standard Model

2.1 Gauge Theory

2.2 Higgs Mechanism

2.3 Flavor Sector

Chapter 3

Beyond the Standard Model

3.1 Overview of the BSM Theories

3.2 Leptoquark Model

Chapter 4

Effective Field Theory

4.1 Standard Model EFT

4.2 Low Energy EFT

Chapter 5

Event Simulation for Colliders

5.1 Perturbative Quantum Chromodynamics

5.2 Nonperturbative Quantum Chromodynamics

Part II

Experimental Apparatus

Chapter 6

The Large Hadron Collider

CHAPTER 7

The Compact Muon Solenoid Detector

7.1 Coordinate System Used in the CMS Detector

7.2 The Inner Tracking System

7.3 The Electromagnetic Calorimeter

7.4 The Hadronic Calorimeter

7.5 The Superconducting Magnet

7.6 The Muon System

7.7 The Trigger System

Chapter 8

Object Reconstruction in the CMS detector

8.1 Electron and Photon

8.2 Muon

8.3 Tau

8.4 Jets

8.5 Energy Sums

Chapter 9

Run-3 Operation of the CMS detector

9.1 The CMS Run Organisation

9.2 Central Shifting Crew

9.3 Detector On-call Expert

Chapter 10

The Phase-2 Upgrade of the CMS Detector

10.1 Overview of the Upgrade

10.2 Leve-1 Track Finder

10.3 Leve-1 Electron Trigger Algorithm

Part III

Search for Flavor-Violating eμtq Interactions

The Part III of this dissertation documents a physics analysis that was performed in 2020-2023 using data collected by the Compact Muon Solenoid (CMS) detector in 2016-2018. This analysis was made public in February 2023 [4] and was largely done by myself with the advice from Prof. Louise Skinnari and technical support provided centrally by the CMS Collaboration. The Part III is organized as follows. Chapter 11 gives brief overview of the existing CLFV searches performed at the A Toroidal LHC Apparatus (ATLAS) and CMS experiments that involves top quarks. Chapter 12 describes the datasets, simulated samples, and triggers used by this analysis. Object- and event-level selection criteria are described in Chapter 13 and Chapter 14, respectively. Treatments of the *nonprompt* backgrounds and signal extraction using BDT are described in Chapter 15 and Chapter 16 respectively. Finally, the systematic uncertainties that affect this analysis and statistical interpretation of the results are described in Chapter 17 and Chapter 18, respectively.

Chapter 11

Previous Searches

CLFV search involving top quarks is an active area of research at the Large Hadron Collider (LHC) experiments. So far, no significant excess over the SM predictions has been reported, and the observations from the ATLAS and CMS experiments are both interpreted using the framework of EFT. Two existing ATLAS analyses are described in Section 11.1 and one existing CMS analysis is described in Section 11.2.

11.1 ATLAS

The flavor-violating eμtq interactions were first studied by the ATLAS Collaboration [1] using data collected during 2015-2017 with an integrated luminosity of 79.8 fb^{-1}. In addition to three leptons, this analysis targets final states with two or more jets. Only the top quark decay signal mode is considered. Lorentz structures of dimension-6 operators are not probed separately. Discriminating variables, such as the p_T of the leptons, are combined into a BDT, which is used to interpret the observation. A representative Feynman diagram targeted by this analysis and the distributions of the BDT discriminator is shown in Figure 11.1.

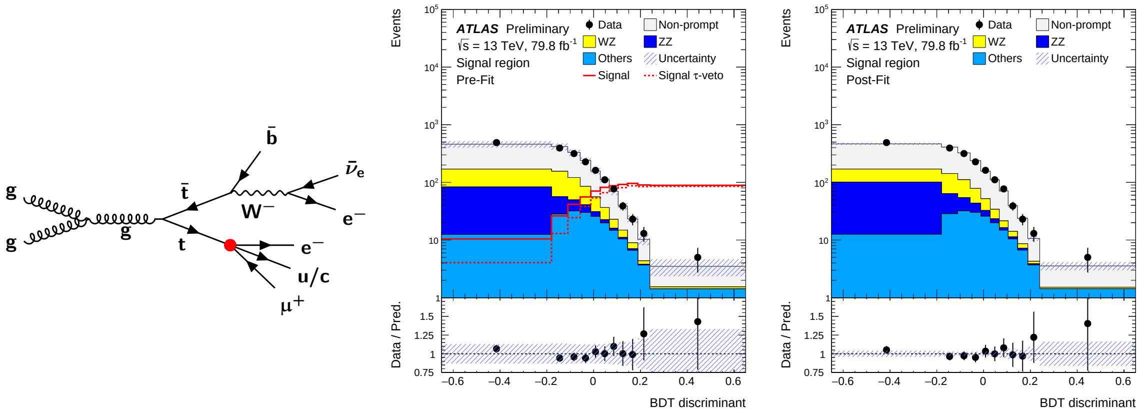


Figure 11.1: Representative Feynman diagram for the CLFV top quark decay processes that are targeted by [1] (left). The CLFV interaction vertex is shown as a solid red circle to indicate that it is not allowed in the SM. The middle (right) histogram shows the distribution of the pre-fit (post-fit) BDT discriminator targeting the CLFV top quark decay.

Data is found to be compatible with the SM predictions, and an upper limit on the branching fraction of $\mathcal{B}(\mathrm{t} \rightarrow \mathrm{e}\mu\mathrm{q}) < 6.6 \times 10^{-6}$ is set at 95% CL [5]. This result improves a previous bound established in an indirect search [6] by three orders of magnitude.

The ATLAS Collaboration also studied the $\mu\tau\mathrm{tq}$ interactions using data corresponds to 140 fb^{-1} [2]. This analysis targets final states with two same-sign muons, one hadronic tau, and one or more jets. Both top quark production and decay signals are considered in this analysis. Operators with different Lorentz structures are considered separately. Representative Feynman diagrams are shown in Figure 11.2.

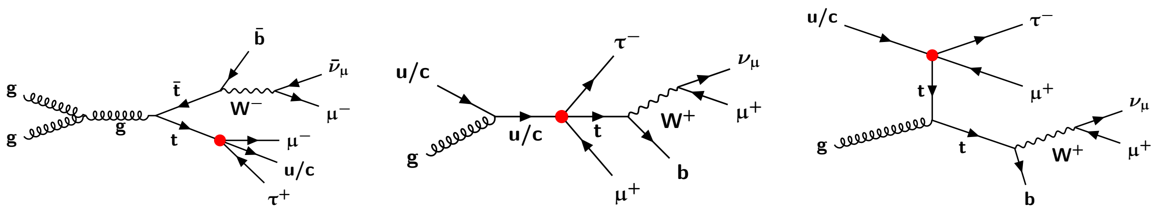


Figure 11.2: Representative Feynman diagrams for the signal processes that are targeted by [2]. Both top quark decay (left) and production (middle and right) CLFV processes are shown. The two muons in the final states are required to have the same electric charge.

Due to limited statistics, event yields of the SRs are directly used to interpret the observation, which is shown in Figure 11.3. An upper limit at 95% CL is placed on the branching fraction of $\mathcal{B}(\mathrm{t} \rightarrow \mu\tau\mathrm{q}) < 1.1 \times 10^{-6}$. The corresponding constraint on the WC improves the previous bound [7] by nearly a factor of 30.

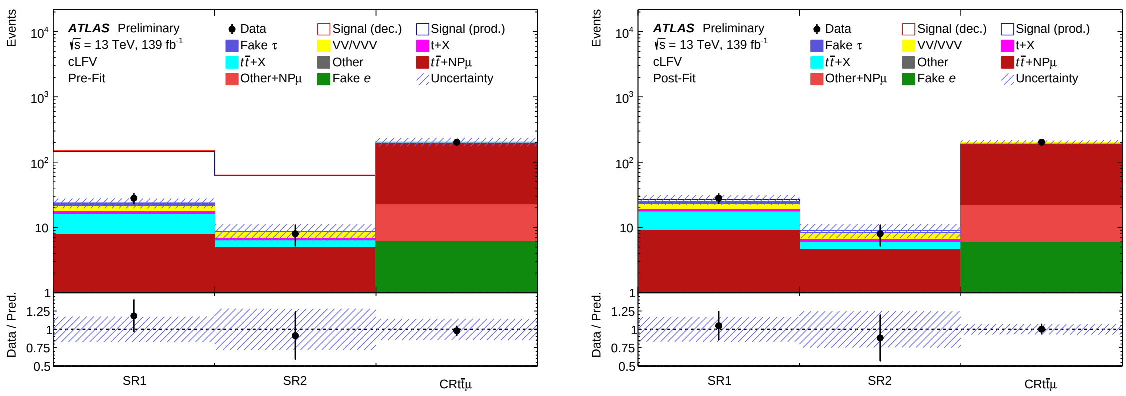


Figure 11.3: The left (right) histogram shows the pre-fit (post-fit) event yields of various regions studied by [2]. In these hisograms, "SR1" denotes the signal region with two or more jets while "SR2" denotes the signal region with exactly one jet. "CRt$\bar{\mathrm{t}}\mu$" denotes the control region of the t$\bar{\mathrm{t}}\mu$ background, where the μ is a *nonprompt* muon.

11.2 CMS

The CMS Collaboration followed up with a search for eμtq interactions using data corresponds to 138 fb^{-1} [3]. Unlike the previous ATLAS analysis [1], this CMS analysis targets final states with two leptons and a hadronically decaying top quark. Both top quark production and decay signals are considered in this analysis. Operators with different Lorentz structures are considered separately. Representative Feynman diagrams are shown in Figure 11.4.

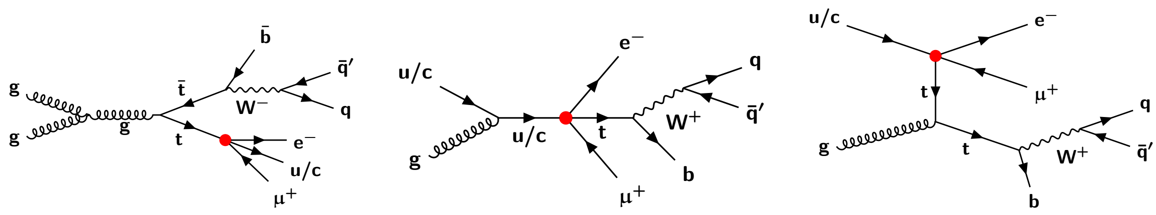


Figure 11.4: Representative Feynman diagrams for the signal processes that are targeted by [3]. Both top quark decay (left) and production (middle and right) CLFV processes are shown. The top quark that does not participate in the CLFV interaction is required to produce fully hadronic final states.

A BDT using multiple discriminating variables is trained to further enhance the sensitivity. Distributions of the BDT discriminator are shown in Figure 11.5. An upper limit at 95% CL is placed on the branching fraction of $\mathcal{B}(t \to \mu\tau q) < 7 \times 10^{-8}$, which improves the previous bound established by the ATLAS Collaboration [1] by two orders of magnitude.

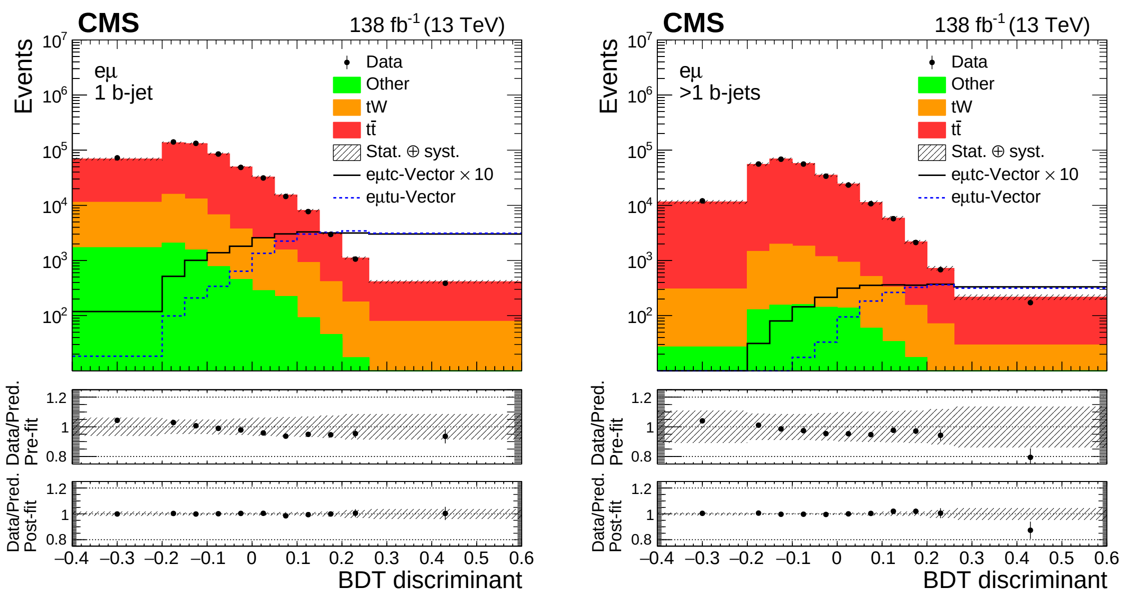


Figure 11.5: The left (right) histogram, taken from [3], shows the distribution of the BDT discriminator in regions with exactly (more than) one b-tagged jet. The middle (bottom) panel shows the ratio of data events and the pre-fit (post-fit) predictions.

Chapter 12

Datasets, Simulated Samples and Triggers

This analysis is based on data collected by the CMS experiment in 2016-2018 from pp collisions at a center-of-mass energy of 13 TeV corresponding to an integrated luminosity of 138 fb^{-1}. There were approximately 30 simultaneous pp collisions occurring per 25 ns. Based on online selection criteria, fully reconstructed collision data that contain high-level physics objects are divided into "Primary Datasets (PDs)". The PDs that make use of lepton information for selection include "DoubleEG", "DoubleMu", "MuonEG", "SingleElectron", and "SingleMuon" for the 2016 and 2017 data-taking era. In 2018, "SingleElectron" and "DoubleEG" are replaced by "EGamma". The names of these PDs reflect the selection criteria. In addition to these PDs, MC samples are also generated to model both signal and background processes, which are described in Section 12.1 and Section 12.2, respectively. To account for the different data-taking conditions across the years, all MC samples are generated separately for each year. HLT triggers are used to select events offline, which is described in Section 12.3.

12.1 Signal Samples

In this analysis, New Physics is described by Dimension-6 EFT operators,

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}}^{(4)} + \frac{1}{\Lambda^2}\sum_a C_a^{(6)} O_a^{(6)} + O\left(\frac{1}{\Lambda^4}\right). \tag{12.1}$$

Among many of the Dimension-6 operators in Warsaw basis [8], a total of 6 operators are considered, which are summarized in Table 12.1. To reduce the number of free parameters, the permutations of fermion flavors are combined. Taking eμtu vertex as an example, the WCs are parameterized in the following way:

$$\begin{aligned}
C_{\mathrm{lq}} &= C_{\mathrm{lq}}^{(1)1213} + C_{\mathrm{lq}}^{(1)2113} + C_{\mathrm{lq}}^{(1)1231} + C_{\mathrm{lq}}^{(1)1213}, & (12.2)\\
C_{\mathrm{lu}} &= C_{\mathrm{lu}}^{1213} + C_{\mathrm{lu}}^{2113} + C_{\mathrm{lu}}^{1231} + C_{\mathrm{lu}}^{1213}, & (12.3)\\
C_{\mathrm{eq}} &= C_{\mathrm{eq}}^{1213} + C_{\mathrm{eq}}^{2113} + C_{\mathrm{eq}}^{1231} + C_{\mathrm{eq}}^{1213}, & (12.4)\\
C_{\mathrm{eu}} &= C_{\mathrm{eu}}^{1213} + C_{\mathrm{eu}}^{2113} + C_{\mathrm{eu}}^{1231} + C_{\mathrm{eu}}^{1213}, & (12.5)
\end{aligned}$$

Table 12.1: Summary of relevant dimension-6 operators considered in this analysis. Here, ε is the two dimensional Levi-Civita symbol, γ^μ the gamma matrix, and $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$. The l and q denote left-handed doublets, whereas u and e denote right-handed singlets. The indices i and j are lepton flavor indices that run from 1 to 2 with $i \neq j$; m and n are quark flavor indices with the condition that one of them is 3 and the other one is 1 or 2.

Lorentz Structure	Operator		
vector	$O^{(1)ijmn}_{\mathrm{lq}}$	$=$	$(\bar{l}_i\gamma^\mu l_j)(\bar{q}_m\gamma_\mu q_n)$
	O^{ijmn}_{lu}	$=$	$(\bar{l}_i\gamma^\mu l_j)(\bar{u}_m\gamma_\mu u_n)$
	O^{ijmn}_{eq}	$=$	$(\bar{e}_i\gamma^\mu e_j)(\bar{q}_m\gamma_\mu q_n)$
	O^{ijmn}_{eu}	$=$	$(\bar{e}_i\gamma^\mu e_j)(\bar{u}_m\gamma_\mu u_n)$
scalar	$O^{(1)ijmn}_{\mathrm{lequ}}$	$=$	$(\bar{l}_i e_j)\ \varepsilon\ (\bar{q}_m u_n)$
tensor	$O^{(3)ijmn}_{\mathrm{lequ}}$	$=$	$(\bar{l}_i\sigma^{\mu\nu} e_j)\ \varepsilon\ (\bar{q}_m\sigma_{\mu\nu} u_n)$

$$C^{(1)}_{\mathrm{lequ}} = C^{(1)1213}_{\mathrm{lequ}} + C^{(1)2113}_{\mathrm{lequ}} + C^{(1)1231}_{\mathrm{lequ}} + C^{(1)1213}_{\mathrm{lequ}}, \tag{12.6}$$

$$C^{(3)}_{\mathrm{lequ}} = C^{(3)1213}_{\mathrm{lequ}} + C^{(3)2113}_{\mathrm{lequ}} + C^{(3)1231}_{\mathrm{lequ}} + C^{(3)1213}_{\mathrm{lequ}}. \tag{12.7}$$

Additionally, all vector-like operators are combined,

$$O^{\mathrm{vector}}_{\mathrm{e\mu tu}} = O_{\mathrm{lq}} + O_{\mathrm{lu}} + O_{\mathrm{eq}} + O_{\mathrm{eu}}, \tag{12.8}$$

$$O^{\mathrm{scalar}}_{\mathrm{e\mu tu}} = O^{(1)}_{\mathrm{lequ}} + \mathrm{h.c}, \tag{12.9}$$

$$O^{\mathrm{tensor}}_{\mathrm{e\mu tu}} = O^{(3)}_{\mathrm{lequ}} + \mathrm{h.c}, \tag{12.10}$$

which results in 6 independent WCs: $C^{\mathrm{vector}}_{\mathrm{e\mu tu}}$, $C^{\mathrm{scalar}}_{\mathrm{e\mu tu}}$, $C^{\mathrm{tensor}}_{\mathrm{e\mu tu}}$, $C^{\mathrm{vector}}_{\mathrm{e\mu tc}}$, $C^{\mathrm{scalar}}_{\mathrm{e\mu tc}}$, $C^{\mathrm{tensor}}_{\mathrm{e\mu tc}}$.

To generate signal MC samples, the effective Lagrangian described above is implemented using the SmeftFR v2 [9] model, and saved in the "UFO" format [10]. Additionally, all the WCs are set to 1 with $\Lambda = 1$ TeV in the UFO, which then interfaces with the FeynRules [11] package to calculate Feynman diagrams. The output of the FeynRules is used in ME event generator MadGraph5_aMC@NLO v2.4.2 [12] to generate events at Leading Order (LO).

In general, the calculations done by the ME event generators are model-agnostic assuming the same EFT configurations. In other words, models like SmeftFR or SMEFTsim [13] are expected to give the same or very similar results in terms of cross sections and four-momenta of

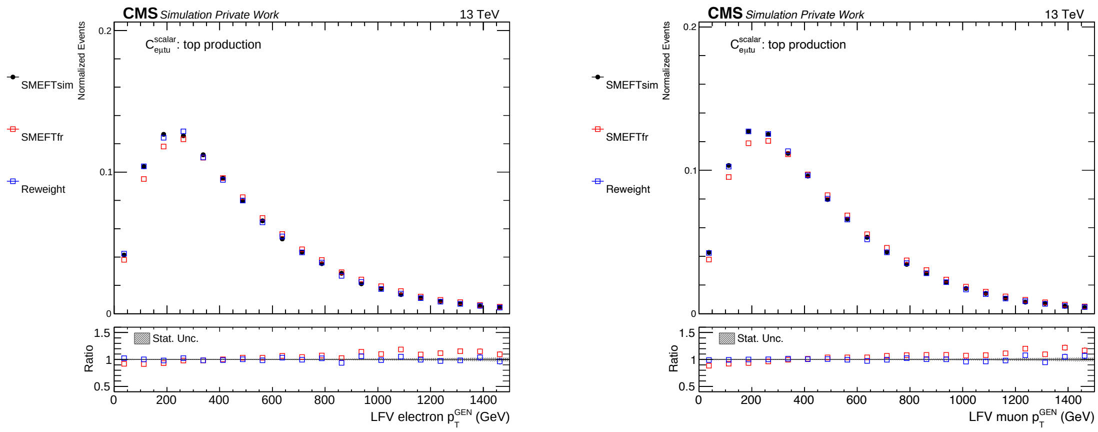


Figure 12.1: Comparison of kinematic distributions at ME-level produced by different models: LFV electron p_T (left), LFV muon p_T (right). The "SmeftFR" samples (shown in red curve) and "SMEFTsim" samples (shown in black curve) are statistically independent of each other. The "Reweight" (shown in blue curve) is produced by applying weights calculated by Equation 12.11 to "SmeftFR" samples.

final-state particles. Nevertheless, visible differences in kinematics have been observed and shown in Figure 12.1. Furthermore, the cross sections predicted by SmeftFR v2 also yield more than 20% difference relative to SMEFTsim due to a bug that was later fixed in SmeftFR v3. In light of these differences, the CMS and ATLAS Collaborations agreed to adopt the SMEFTsim model as the common standard. To quantify the impact of the choice of models on kinematics, the following ratio is calculated for each event i,

$$R^i_{\text{reweight}} = \frac{\omega^i_{\text{SMEFTsim}}}{\omega^i_{\text{SmeftFR}}}, \tag{12.11}$$

where ω^i_X is the per-event ME weight calculated by MadGraph5_aMC@NLO using model X. Since SMEFTsim was not used by CMS at the time when the signal samples were generated, R_{reweight} are used to "reweight" the original samples generated using SmeftFR.

Due to the significant differences in kinematic distributions between top decay and production signals, MC samples are generated separately for these processes. The cross sections for top production signals are taken directly from MadGraph5_aMC@NLO
with SMEFTsim UFO as input. The event generation for top decay signals at the ME-level takes two steps: (i) production of the SM $t\bar{t}$, and (ii) CLFV decay of one of the top quarks. Therefore, the $t\bar{t}$ cross-section at Next-to-Next-to-Leading order (NNLO) precision [14] is used to normalize the top decay signals,

$$\sigma_{\mathsf{CLFV}}^{\mathsf{Top\ Decay}} = 2 \times \sigma_{\mathsf{t\bar{t}}}^{\mathsf{NNLO}} \times \mathcal{B}(\mathsf{t} \to \mathsf{e\mu q}), \tag{12.12}$$

where q={u,c}, and $\mathcal{B}(\mathsf{t} \to \mathsf{e\mu q})$ [15] can be expressed as,

$$\mathcal{B}(\mathsf{t} \to \mathsf{e\mu q}) = \begin{matrix} \frac{|C_{\mathsf{e\mu tq}}^{\mathsf{vector}}|^2}{\Lambda^4} \frac{m_t^5}{384\pi^3\Gamma_t^{\mathsf{SM}}} \\ \frac{|C_{\mathsf{e\mu tq}}^{\mathsf{scalar}}|^2}{\Lambda^4} \frac{m_t^5}{3072\pi^3\Gamma_t^{\mathsf{SM}}} \\ \frac{|C_{\mathsf{e\mu tq}}^{\mathsf{tensor}}|^2}{\Lambda^4} \frac{m_t^5}{64\pi^3\Gamma_t^{\mathsf{SM}}} \end{matrix} \tag{12.13}$$

where m_t and Γ_t^{SM} are taken to be 172.5 GeV and 1.33 GeV in this analysis, respectively. The choice of u or c quark in final states does not affect the cross sections of the top decay signals. The cross sections for all signal MC samples are summarized in Table 12.2. These cross-sections are used as a baseline to define the signal strength μ, which is used to quantify the relative strength of the signals when their normalization change,

$$\mu(C/\Lambda^2) = \frac{\sigma_{\mathsf{CLFV}}(C/\Lambda^2)}{\sigma_{\mathsf{CLFV}}(1\mathsf{TeV}^{-2})} \propto (C/\Lambda^2)^2. \tag{12.14}$$

Table 12.2: Theoretical cross sections for top production and decay for each CLFV coupling, calculated at $C/\Lambda^2 = 1\ \mathsf{TeV}^{-2}$. Uncertainties related to PDF and QCD scale in ME calculation are given ($\sigma^{+\mathsf{scale}}_{-\mathsf{scale}} \pm$ PDF).

Lorentz Structure	Samples	XS (fb)
vector	top production via u quark	$460^{+81}_{-64} \pm 6$
	top production via c quark	$33^{+5}_{-4} \pm 6$
	top decay via u/c quark	$32^{+0.8}_{-1.1} \pm 1.3$
scalar	top production via u quark	$97^{+18}_{-14} \pm 1$
	top production via c quark	$6.3^{+0.9}_{-0.8} \pm 1.4$
	top decay via u/c quark	$4.0^{+0.1}_{-0.1} \pm 0.2$
tensor	top production via u quark	$2143^{+368}_{-293} \pm 31$
	top production via c quark	$164^{+22}_{-18} \pm 27$
	top decay via u/c quark	$187^{+5}_{-6} \pm 8$

Steps other than the ME calculation concerning signal MC generation follow the CMS standard, which is described in the following section.

12.2 Background Samples

Besides tZq, tHq, tHW, and tWZ processes, the Next-to-Leading Order (NLO) Parton Distribution Function (PDF) set from NNPDF3.0 [16] is used in 2016 to generate background MC samples. The NNLO PDF set from NNPDF3.1 [17] is used for tZq while the LO PDF set from NNPDF3.0 is used for tHq, tHW, and tWZ in 2016. In 2017 and 2018, the NNLO PDF set from NNPDF3.1 was used to generate all the samples.

The default choice of ME event generator is MadGraph5_aMC@NLO v2.4.2 (v2.2.2 for 2016), which is used to generate all but ZZ, $t\bar{t}$H, and $t\bar{t}$ samples. These three samples are generated with POWHEG v2 [18] instead. Samples with small contributions (tHq, tWZ, tHW, and low mass DY) are generated at LO while other samples are generated at NLO. Whenever possible and relevant, theoretical cross sections from high-order Quantum Chromodynamics (QCD) calculations are used. The references of these calculations are included in 12.3.

The PYTHIA v8.2 [19] is used to model parton shower and hadronization. The CUETP8M1 [20] was used in 2016 for underlying event tuning while the CP5 [21] was used in 2017 and 2018. The configurations of the MC samples are summarized in Table 12.3. The background processes are divided into two categories: (i) processes with three or more *prompt* leptons in the final states are classified as "*prompt* background", and (ii) other processes are classified as "*nonprompt* background". The *nonprompt* backgrounds in this analysis are modeled with a data-driven technique, which is discussed in Chapter 15. The MC samples listed in "*nonprompt*" category in Table 12.3 are therefore only used for validations.

12.3 Triggers

The target final states of this analysis contain three prompt leptons, which make lepton triggers the most optical choice to select events. To achieve maximum acceptance, a combination of single-lepton, di-lepton, and tri-lepton triggers are used. These triggers are summarized in Appendix A. Events in simulated samples are required to fire at least one of the triggers listed in Table A.1-A.3. Since multiple PDs are used to record data events and the orthogonality of these PDs is not guaranteed by the online selection criteria, the following trigger logic is implemented to remove the overlap between different PDs:

- Events in SingleMuon datasets are required to fire at least one of the triggers listed under "SingleMuon".
- Events in DoubleMuon datasets are required to fire at least one of the triggers listed under "DoubleMu". Events are removed if they also fire at least one of the triggers listed under "SingleMuon".
- Events in "MuonEG" datasets are required to fire at least one of the triggers listed under "MuonEG". Events are removed if they also fire at least one of the triggers listed under

Table 12.3: Summary of the configurations of the MC samples. DYM50 (DYM10to50) denotes a DY sample with a dilepton invariant mass greater than 50 GeV (between 10 and 50 GeV). V includes W and Z bosons. The cross-sections for samples without a citation are taken directly from their event generators.

Category	Process	Event Generator	Perturbative QCD	Tune	XS precision
prompt background	WZ	MadGraph	NLO	CUETP8M1(CP5)	NLO [22]
	ZZ	powheg	NLO	CUETP8M1(CP5)	NLO [22]
	VVV	MadGraph	NLO	CUETP8M1(CP5)	NLO
	$t\bar{t}$W, $t\bar{t}$Z	MadGraph	NLO	CUETP8M1(CP5)	NLO [23, 24]
	$t\bar{t}$H	powheg	NLO	CUETP8M1(CP5)	NLO [24]
	tZq	MadGraph	NLO	CP5	NLO
	tHq, tHW, tWZ	MadGraph	LO	CUETP8M1(CP5)	LO
nonprompt background	$t\bar{t}$	powheg	NLO	CUETP8M1(CP5)	NNLO [14]
	DYM50	MadGraph	NLO	CUETP8M1(CP5)	NNLO [25]
	DYM10to50	MadGraph	LO	CUETP8M1(CP5)	NLO [25]

"SingleMuon" or "DoubleMu".

- Events in Single Electron datasets are required to fire at least one of the triggers listed under "SingleElectron". Events are removed if they also fire at least one of the triggers listed under "SingleMuon", "DoubleMu", or "MuonEG".
- Events in DoubleEG datasets are required to fire at least one of the triggers listed under "DoubleEG". Events are removed if they also fire at least one of the triggers listed under "SingleMuon", "DoubleMu", "MuonEG", or "SingleElectron".
- Events in EGamma datasets are required to fire at least one of the triggers listed under "EGamma". Events are removed if they also fire at least one of the triggers listed under "SingleMuon", "DoubleMu", or "MuonEG".

Chapter 13

Object Selection

Objects described in Chapter 8, referred to as "candidates", are further selected with more stringent requirements to suppress the contributions from background processes while maintaining a high signal acceptance. In particular, prompt electron and muon candidates are identified through a custom-trained BDT classifier, which is discussed in Section 13.1. Two jet identification algorithms are deployed to select jet candidates originating from hard collisions, which is discussed in Section 13.2. Furthermore, jet candidates that originate from b quarks are identified with a Neural Network (NN) based algorithm, which is discussed in Section 13.3.

13.1 Lepton Selection

The target final states of this analysis feature exactly three leptons that originate either from decays of electroweak bosons or from the CLFV interaction, which in this analysis is a contact interaction that involves four fermions. These leptons, referred to as *prompt* leptons, typically appear to be isolated and not far away from the Primary Vertex (PV). In contrast, *nonprompt* leptons are leptons that originate from decays of hadrons, or photon conversions, or misidentified leptons. They often travel a noticeable distance away from the PV and appear to be less isolated due to nearby activities. Due to the high multiplicity of leptons in our selection, backgrounds with at least one *nonprompt* lepton outnumber any other SM processes that produce three or more *prompt* leptons. It is therefore crucial to exploit the differences between *nonprompt* and *prompt* leptons and bring the *nonprompt* background under control.

13.1.1 TOP LeptonMVA

The TOP LeptonMVA is an offline lepton identification algorithm that was originally developed for tZq analyses [26, 27]. It is based on Gradient BDT implemented using the TMVA package [28]. A total of 13 features are used as input to the BDT. They can be categorized into four groups: (i) positions and momenta of the lepton candidates, (ii) isolation variables, (iii) variables associated with the closest jet, and (iv) a quality variable that is specific to electron or muon candidate. The version of TOP LeptonMVA used by this analysis is the same as [27], where a detailed description of all input features can be found.

Prompt leptons from $t\bar{t}W$, $t\bar{t}Z$, and tZq samples are used as signals in the BDT training while *nonprompt* leptons from $t\bar{t}$ samples are used as backgrounds. The trained BDT outputs a single score for each lepton candidate ranging from -1 to 1 with -1 (1) being the most background- (signal-) like. the tight working point with a threshold of ($>$) 0.9 is chosen as the selection criteria for both electron candidates and muon candidates, which corresponds to a signal(background) efficiency of 90%(1%). The strategy is to trade a small percentage (<10%) of signal efficiency for several factors of background rejection.

13.1.2 Full Selection

In addition to the TOP LEPTONMVA requirement, a set of common selection criteria is applied to both electron and muon candidates. The minimum p_T requirement is 38GeV, 20 GeV, and 20 GeV for the leading, sub-leading, and trailing lepton in p_T, respectively. This requirement is driven by the p_T thresholds of the HLT triggers to avoid inefficiency at turn-on. Electron and muon candidates are required to be in the pseudorapidity range $|\eta|$ <2.4, which corresponds to the acceptance of CMS tracker and muon system in 2016-2018. The transverse (longitudinal) impact parameters with respect to the PV, denoted as d_{xy} (d_z), is required to be in the range $|d_{xy}| < 0.05$ cm ($|d_z| < 0.05$ cm). The significance of the 3-dimensional impact parameter, denoted as SIP_3, is defined as the 3-dimensional impact parameter divided by its uncertainty. It is required that $\mathrm{SIP}_3 < 8$. The three cuts on impact parameters are added due to the difference in distributions of these parameters between *prompt* and *nonprompt* leptons. Also, they are part of the pre-selection requirement in the BDT training.

Furthermore, all lepton candidates are required to be isolated. This is achieved by first defining a cone with a distance parameter of ΔR around each lepton candidate, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. Only particles within $\Delta R < R_{max}$ can contribute to the isolation variable, where R_{max} is referred to as the size of the cone. Secondly, a Particle Flow (PF) based isolation variable is defined as,

$$I^{rel}_{mini} = \frac{1}{p_T^\ell}\{\sum_{charged} p_T + \max(0, \sum_{neutral} p_T - \rho\mathcal{A}[\frac{\Delta R}{0.3}]^2)\}, \tag{13.1}$$

where p_T^ℓ is the p_T of the lepton candidate, the first term inside the curly braces is the scalar sum of all charged particles associated with the PV while the second term evaluates the contribution from neutral particles. This is done by first scalar-summing over p_T of all neutral particles associated to the PV. A correction term, known as effective area correction [29], is then subtracted. This term is used to mitigate the impact of Pile-Up (PU) interactions. The size of the cone scales with p_T^ℓ as,

$$R_{max} = \max(0.05, \min(0.2, \frac{10\mathrm{GeV}}{p_T^\ell})). \tag{13.2}$$

.

This type of isolation variable is known as "mini" isolation, which maximizes the signal efficiency at p_T^ℓ by reducing the cone size. It is required that lepton candidates to have $I_{\text{mini}}^{\text{rel}} < 0.12$.

For electron specifically, candidates are required to have a GSF track with one or less missing inner hits. Electron candidates reconstructed in the transition region between ECAL barrel and endcap (i.e. $1.44 < |\eta_{SC}| < 1.57$) are removed from consideration. For muon specifically, candidates are required to be PF muons and pass the medium working point discussed in Section 8.2.

Lepton candidates that pass all requirements stated above are referred to as "*tight*" leptons. Leptons selected with a separate set of criteria, known as "*loose*", is used in estimating the *nonprompt* background, which is discussed in Chapter 15. Unless explicitly stated, all lepton objects presented in this search are *tight* leptons.

The energy scale and resolution is calibrated for all electron candidates, as discussed in Section 8.1. The energy scale and resolution is also calibrated for muon candidates with $p_T < 200$ GeV, as discussed in Section 8.2. Per-object scale factors are applied to *tight* leptons in simulated events to correct for the differences in reconstruction, isolation, and identification between data and MC. These scale factors are obtained using dilepton events in the Z resonance window.

13.2 Jet Selection

Jet candidates are reconstructed from PF candidates using the anti-k_T algorithm described in Section 13.2 with a cone size of 0.4. Charged hadrons that are not associated to the PV are removed. Jet candidates are required to have a minimum p_T of 30 GeV and in the pseudorapidity range $|\eta| <2.4$, where b-tagging is still effective. It is further required that all jet candidates be isolated from *tight* leptons. A cone of the size 0.4 around each jet candidate is defined and candidates will be removed if any *tight* leptons are found within such a cone. This procedure is implemented to remove the overlap between leptons and jets.

The two primary sources of background are (i) detector noise, and (ii) jets from PU interactions. To suppress detector noise, a set of cut-based selections is applied to jet candidates. This algorithm utilizes information from PF candidates, including: (i) fraction of charged (neutral) hadrons energy, (ii) fraction of charged (neutral) EM energy, (iii) fraction of muon energy, and (iv) object multiplicity. The "tightLepVeto" working point is chosen to select jet candidates, which corresponds to 98-99% signal efficiency.

The second algorithm is designed to reject jet candidates that originate from PU interactions. This algorithm is based on a BDT that utilizes: (i) the trajectories of tracks associated to the jets, (ii) the topology of the jet shape, and (iii) object multiplicity. The *loose* working point is chosen to select jet candidates with $p_T < 50$ GeV, which corresponds to 99% signal efficiency. Applying this algorithm to jet candidates with $p_T > 50$ GeV is both ineffective and unnecessary as PU jets mostly reside in the low p_T spectrum. The overall effect of this algorithm on this analysis is small as PU jets constitute only a small fraction of all jet candidates in the phase space of this analysis.

As discussed in Section 8.4, the energy scale for all jet candidates (data and MC) are calibrated. One extra correction is applied to simulated jets to recreate the jet energy resolution as measured in the data.

13.3 Identification of b jets

The DeepJet algorithm [30] is used to identify jets that originate from b quark. The core strategy of this algorithm is to minimize information. This is achieved by removing entirely the selection of jet constituents, which limits the number of jet constituents considered. Additionally, an effort is made to use as many low-level features as possible, which further further deepens the feature space. Approximately a total of 650 input features are used, which can be categorized into four groups: (i) global variables, (ii) charged PF candidate features, (iii) neutral PF candidate features, (iv) and Secondary Vertex (SV) features associated with the jet. When compared to the existing DeepCSV algorithm [31], DeepJet algorithm delivers up to 20% improvement in signal efficiency while maintaining the same background efficiency.

The DeepJet algorithm outputs a score ranging between 0 and 1, with 0 (1) being the most background- (signal-) like. The medium working point is chosen to tag b jet candidate, which corresponds to 70%-80% signal efficiency. The shape of the DeepJet output distribution is corrected for the differences between data and MC in signal and background efficiencies. The per-event correction weight ω is defined as,

$$\omega = \prod_{i}^{N_{\mathrm{jets}}} \mathrm{SF}(p_{T_i}, \eta_i, F_i, D_i), \tag{13.3}$$

where SF is the ratio of efficiency in data to efficiency to MC parameterized as a function of p_{T}, η, (MC truth) flavor F, as well as DeepJet output D of each jet in the event. ω is applied to all MC events.

Additional corrections are applied to remove the normalization effect of ω before jet selection. These scale factors are measured using MC in $\mathrm{e}\mu\ell$ channel described in Chapter 14. The effect of these scale factors is shown in Figure 13.1.

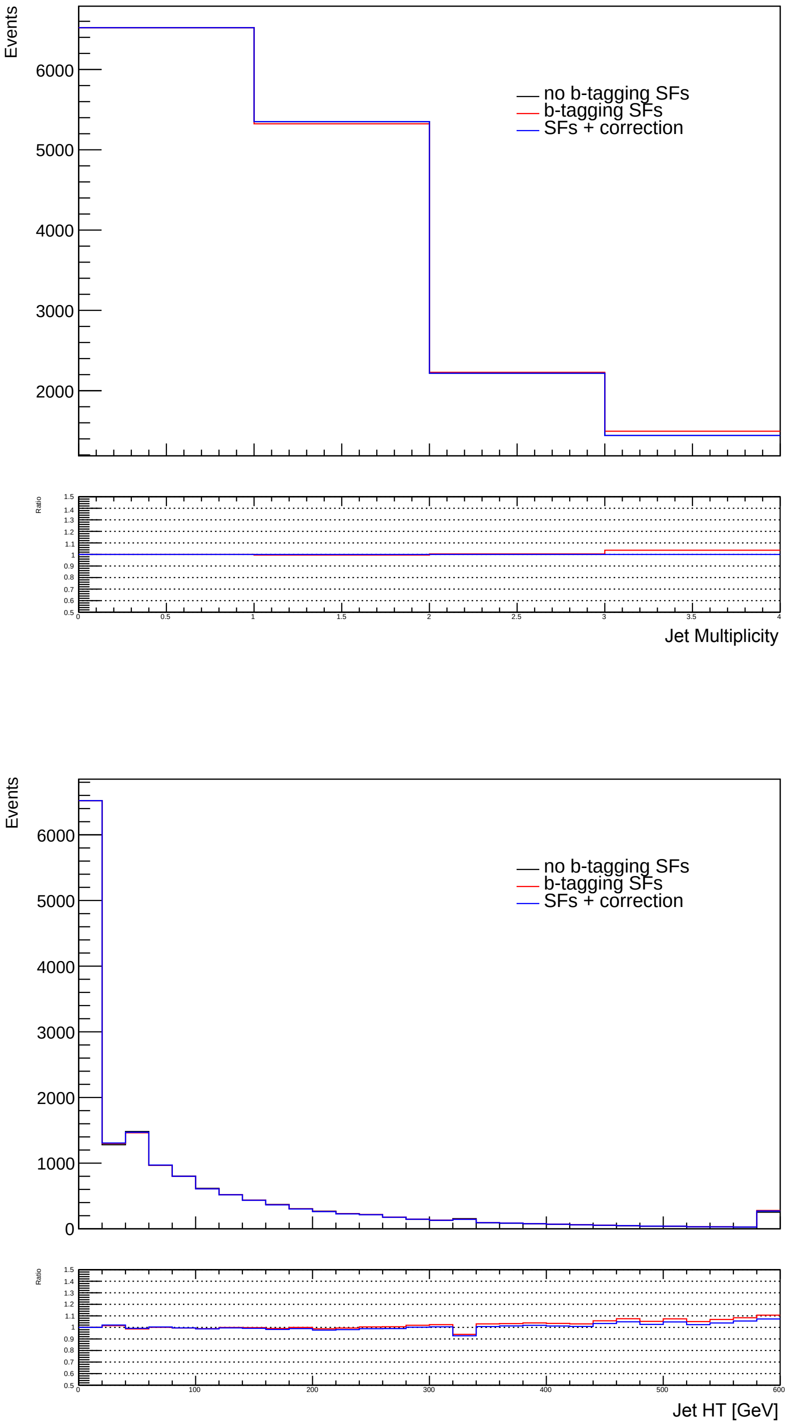


Figure 13.1: Simulated events in $e\mu\ell$ channel without additional requirements on jets. The top histogram shows distribution of jet multiplicity while the bottom histogram shows the distribution of H_{T}, which denotes the scalar sum of the p_{T} of all jets. Distributions without any jet-related scale factors are shown in black lines. Distributions with only b-tagging scale factors are shown in red lines. Distributions with b-tagging scale factors and corrections to remove normalization effects are shown in blue lines.

Chapter 14

Event Selection

Events are required to contain exactly three *tight* leptons described in Section 13.1. Furthermore, events are selected with HLT triggers discussed in Section 12.3. Events with different lepton flavor composites are further categorized into three exclusive channels: eee, $\mu\mu\mu$, $e\mu\ell$. In all three channels, the sum of the electric charges of the selected leptons is required to be 1 or -1. The leading leptons in all selected events are required to be matched with trigger objects within $\Delta R < 0.2$. Within each channel, different regions are defined to further understand signal and background.

$e\mu\ell$ is the channel where close to 100% of the simulated signal events reside. This channel is divided into signal-enriched SRs and signal-depleted VRs, which are discussed in Section 14.1 and Section 14.2, respectively. Due to the lack of different flavors, the eee and $\mu\mu\mu$ channels are signal-depleted by definition. Therefore, events found in these two channels are only used to study background processes, which are discussed in Section 14.2. The kinematic reconstruction of heavy particles, such as the top quark, is described in Section 14.3.

14.1 Signal Region

The core feature of the signal events is the presence of the "LFV $e\mu$" pair, which consists of a pair of Opposite-Sign and Different-Flavor (OSDF) leptons. It is guaranteed that there is at least one OSDF pair in all events residing in $e\mu\ell$ channel due to the requirement on electric charges. The OSDF pair is immediately labeled as the LFV $e\mu$ pair if it is only possible to reconstruct one OSDF pair. In events where a pair of Same-Sign and Same-Flavor (SSSF) leptons are present, a kinematic reconstruction is used to determine which one of two leptons form the LFV $e\mu$ pair with the third lepton, which is detailed in Section 14.3. Leptons that form the LFV $e\mu$ pair are referred to as the LFV electron or muon as it is assumed that they originate from the CLFV interaction. Based on the event topology of the signal process, further selection criteria are applied to define the SR. These selection criteria help achieve an optimal signal-to-background ratio by removing the majority of the background events present in $e\mu\ell$ channel.

At the tree-level, signal events are expected to contain one or two jets, which motivates a requirement of at least one jet in SR. Furthermore, it is required that there is no more than one b-tagged jet to suppress the contribution from $t\bar{t}$ events. Another prominent background is Drell-Yan production that features an OSSF pair. To suppress Drell-Yan processes in SR, events that contain an OSSF lepton pair with an invariant mass between 50 GeV and 106 GeV are removed. The lower bound of this veto is lower than the typical value (e.g. 75 GeV) because the mass range between 50 GeV and 75 GeV has very few signal events and is dominated by *nonprompt* background from photon conversion. Additionally, a modest threshold of 20 GeV is applied to MET due to the presence of neutrinos in the signal events.

Distributions of the LFV eμ mass and the Z boson mass are shown in Figure 14.1. All backgrounds in Figure 14.1 are estimated using MC simulation even though the strategy is to use a data-driven method to estimate the *nonprompt* background. This serves as a preliminary check to understand the components of different backgrounds in SR. Distributions of more variables in SR are included in Appendix B.

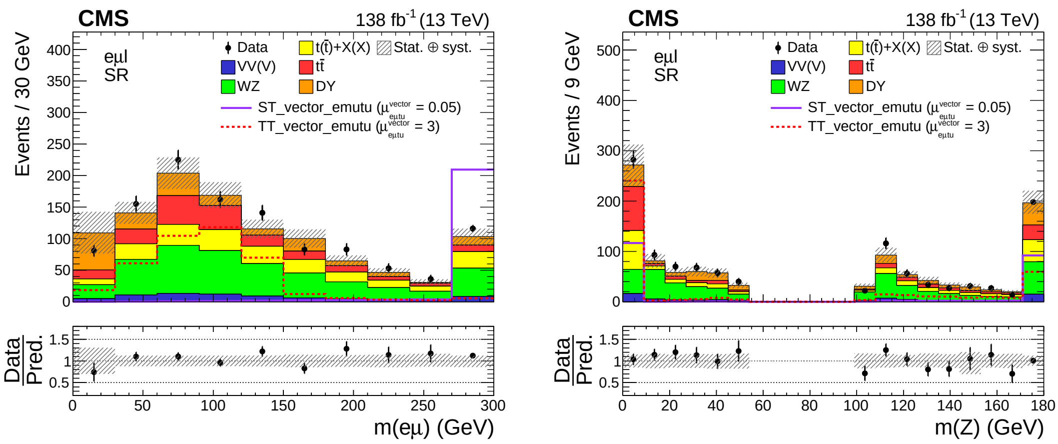


Figure 14.1: Distributions of the LFV eμ mass (left) and the Z boson mass (right) in SR. The data are shown as filled points and the SM background predictions as histograms. The VV(V) background includes ZZ and triboson production, while the $t\bar{t}$ + X(X) component includes $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, tZq, and smaller backgrounds containing one or two top quarks plus a boson or quark. All backgrounds are estimated using MC simulation. The hatched bands indicate statistical and systematic uncertainties for the SM background predictions. The normalization of the signal processes is chosen arbitrarily for improved visualization. The last bin of both histograms includes the overflow events.

Using the LFV eμ mass, the SR is further divided into two subsets to create top production and decay enriched regions:

- SR1, m(eμ) $<$ 150 GeV: top quark decay enriched.
- SR2, m(eμ) $>$ 150 GeV: top quark production enriched.

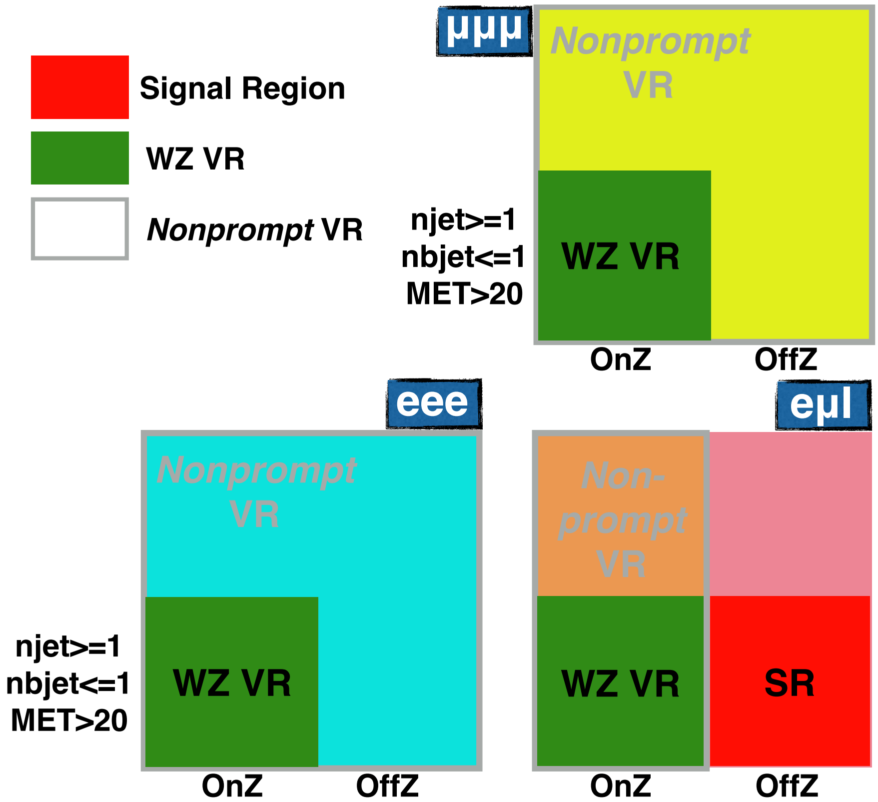


Figure 14.2: Illustration of selection criteria used to define different regions. "OnZ" means the presence of at least one OSSF pair with an invariant mass between 50 GeV and 106 GeV. Events are labeled as "OffZ" when they fail "OnZ" criteria.

14.2 Validation Region

There are two types of signal-depleted VR defined across three channels: *nonprompt* VR and WZ VR. The purpose of these two types of VR is only limited to the validation of the background modeling as neither of them enters the final fit. It is expected that the *nonprompt* VR has a significant fraction of *nonprompt* background while WZ production is responsible for most of the backgrounds in the WZ VR. Distributions of leading lepton p_T and leading lepton η in the WZ control region can be found in Figure 14.3. The *nonprompt* VRs are further discussed in Chapter 15.

Selection criteria used to define different regions are illustrated in Figure 14.2 and are summarized in Table 14.1.

14.3 Kinematic Reconstruction

As mentioned, the LFV eμ pair is assumed to be the product of the CLFV interaction, while the third lepton, referred to as the standalone lepton, is assumed to originate from the leptonically

Table 14.1: Summary of the selection criteria used to define different event regions. "OnZ" means the presence of at least one OSSF pair with an invariant mass between 50 GeV and 106 GeV. Events are labeled as "OffZ" when they fail "OnZ" criteria.

Channel	Region	OnZ	OffZ	MET > 20 GeV	njet>=1	nbjet<=1
eee	VR	-	-	-	-	-
	WZ VR	✓	-	✓	✓	✓
$e\mu\ell$	SR	-	✓	✓	✓	✓
	Nonprompt VR	✓	-	-	-	-
	WZ VR	✓	-	✓	✓	✓
$\mu\mu\mu$	Nonprompt VR	-	-	-	-	-
	WZ VR	-	✓	✓	✓	✓

decaying top quark. To distinguish this top quark ($t \to \ell\nu b$) from the top quark that decays via the CLFV interaction ($t \to e\mu q$), the former is referred to as the SM top quark while the latter is referred to as the LFV top quark.

Jet with the highest b-tagging score, regardless of whether or not it crosses the medium working point threshold, is assumed to originate from bottom quark decay. Therefore, it is combined with MET to build the SM top quark. The x and y components of MET are taken as measurements of neutrino p_x and p_y. The z component of neutrino momentum is calculated by imposing the constraint that the invariant mass of the combined object (standalone lepton + neutrino) must be equal to W boson mass. If there is no real solution, the real part of the complex solution is taken. If there is more than one real solution, the solution that is the closest to the p_z of the standalone lepton is taken. In events where there is more than one candidate of standalone lepton (i.e. the presence of the SSSF pair), the lepton that gives a top mass that is the closest to the SM top quark mass ($m_t = 172.5$ GeV) is taken as the standalone lepton.

Once the standalone lepton has been determined, the remaining two leptons are labeled as the LFV $e\mu$ pair and are combined with each selected jet to reconstruct the LFV top quark candidates. Jet with the highest b-tagging score is excluded from this reconstruction since it is assumed to be from the decay of the SM top quark. Out of all the LFV top quark candidates, the candidate that gives a top mass that is the closest to the SM top quark mass is taken. The LFV top quark mass is set to 0 in events where there are less than two jets.

Z boson candidate is reconstructed using the OSSF pair, which is not guaranteed to be present in the $e\mu\ell$ channel. The Z boson mass (m_Z) is set to 0 in events where the OSSF is absent. Z boson candidate is the only heavy particle reconstructed in the eee and $\mu\mu\mu$ channels. Since there are always two ways to form the OSSF pair, the OSSF pair with an invariant mass that is closer to the Z boson mass ($m_Z = 91.2$ GeV) is taken.

Jets with high b-tagging scores are combined with leptons to form so-called "$m_{b\ell}$" systems. The first $m_{b\ell}$ system takes the jet with the highest b-tagging score and combines it with each *tight* lepton in events. Out of the three $m_{b\ell}$ system candidates, the one with the lowest $m_{b\ell}$ is taken, and the two constitutes are excluded from the consideration of the second $m_{b\ell}$ system. If additional jets exist, the second $m_{b\ell}$ system takes the jet with the highest b-tagging score and combines it with two of the remaining leptons separately. Out of the two candidates, the one with the lowest $m_{b\ell}$ is taken. $m_{b\ell}$ is set to 0 if no additional jet exists after the formation of the first $m_{b\ell}$ system.

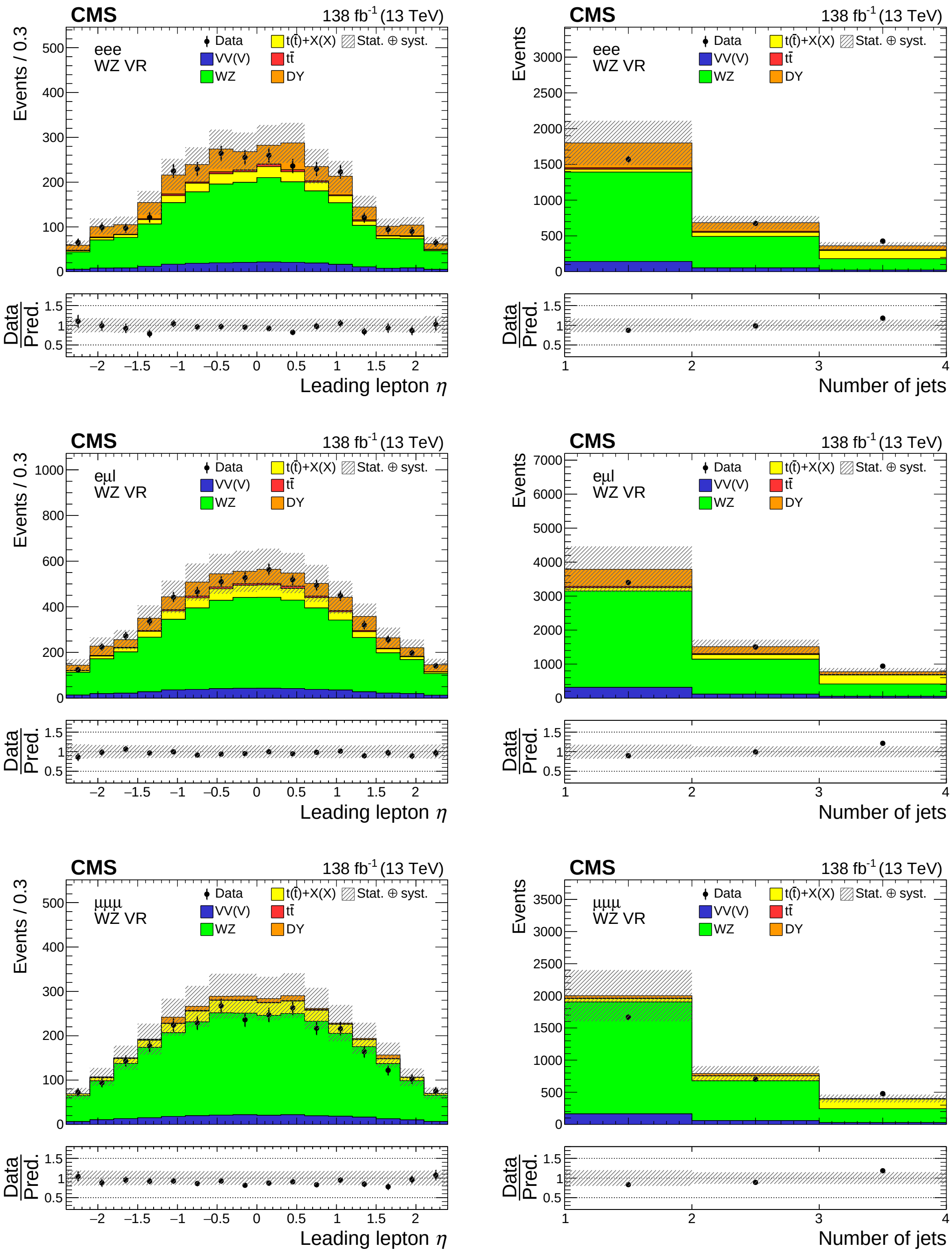


Figure 14.3: Distributions of the leading lepton η (left column) and the jet multiplicity (right column) in the WZ VRs. Events in the eee, e$\mu\ell$, and $\mu\mu\mu$ WZ VRs are shown in the upper, middle, and lower row, respectively. The data are shown as filled points and the background predictions as histograms. All backgrounds are estimated with MC simulation. The hatched bands indicate statistical and systematic uncertainties for the background predictions. The last bin of the right column histograms includes the overflow events.

Chapter 15

Nonprompt Background Estimation

In this analysis, the term *prompt* leptons refers to leptons that originate from the CLFV vertex, the Drell-Yan process, or an electroweak boson decay, including leptons from τ decays if the τ lepton originates from the latter two processes. *Nonprompt* leptons refer to leptons that originate from hadron decays and photon conversions, as well as particles misidentified as leptons. *Nonprompt* leptons are suppressed through isolation requirements and a Multivariate Analysis (MVA)-based identification specifically trained to reject them.

Nonprompt backgrounds are defined to be backgrounds with at least one *nonprompt* lepton passing the *tight* selection criteria, in this case generally dominated by Drell-Yan and $t\bar{t}$ production. An accurate estimation of *nonprompt* backgrounds is difficult to achieve through MC modeling. Therefore, a data-driven technique called the "*matrix method*" [32] is used to estimate the *nonprompt* backgrounds.

A brief description of the *matrix method* in its simplest form is given in Section 15.1 followed by its generalization and implementation in Section 15.2. This method is validated using three VRs and is described in Section 14.2. Lastly, the *nonprompt* estimation in the SR is presented in Section 15.4.

15.1 The Matrix Method

The *matrix method* is a data-driven technique used to estimate the fraction of *nonprompt* leptons that pass a given lepton selection, referred to as "*tight*". The *tight* selection usually incorporates tight lepton identification and isolation requirements and corresponds to the full lepton selection used in an analysis. The *loose* selection is obtained by loosening the *tight* selection. The *loose* selection is used as a baseline such that any *loose* leptons fall into one of the two exclusive categories: *tight* or *not tight*. The *matrix method* deals with *prompt* and *nonprompt* leptons separately. As a result, *prompt* and *nonprompt* efficiencies are introduced, as illustrated in Figure 15.1.

In a simplified scenario with only one lepton in the final state, the *prompt* efficiency r measures the probability of *prompt* leptons passing *tight* selection. It is treated as an observable that can be obtained through measurement,

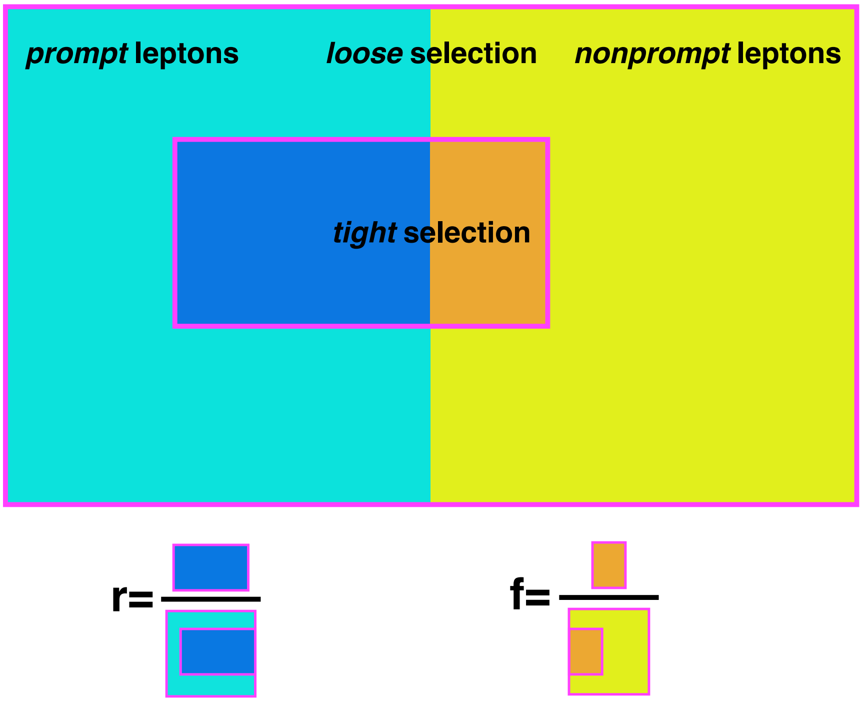


Figure 15.1: Illustration of the *prompt* efficiency r and the *nonprompt* efficiency f. The *loose* selection is typically a subset of the *tight* selection, which guarantees both r and f to be greater than 0 and smaller 1.

$$r = \frac{n_P^T}{n_P^T + n_P^{\overline{T}}}, \tag{15.1}$$

in which $n_P^T/n_P^{\overline{T}}$ denotes the number of events with a *prompt* lepton that is *tight*/*not tight*.

Similarly, *nonprompt* efficiency f can be expressed as,

$$f = \frac{n_N^T}{n_N^T + n_N^{\overline{T}}}, \tag{15.2}$$

in which $n_N^T/n_N^{\overline{T}}$ denotes the number of events with a *nonprompt* lepton that is *tight*/*not tight*.

The measurement of r/f is often performed in dedicated control regions, where high purity of *prompt*/*nonprompt* leptons is expected. These regions are referred to as the MR. It is assumed that r/f is a universal property of *prompt*/*nonprompt* leptons that is independent of physics processes. Therefore, r/f extracted from MR can be used to estimate the contamination of *nonprompt* leptons in a different region (e.g. SR) even though these two regions are orthogonal to each other.

In this simplified scenario, the total number of events in the region of interest (e.g. SR/VR) with a *tight*/*not tight* lepton can be expressed in a system of equations,

$$
\begin{aligned}
N^{T} &= N_P^{T} + N_N^{T} \\
N^{\overline{T}} &= N_P^{\overline{T}} + N_N^{\overline{T}},
\end{aligned}
\tag{15.3}
$$

in which the capital letter "N" is used to indicate that these numbers are referring to events in a region that is different from MR. $N_P^{\overline{T}}/N_N^{\overline{T}}$ can be expressed in terms of r/f and N_P^{T}/N_N^{T} according to Equation 15.1/15.2 and the assumption that r/f remains the same across different regions,

$$
\begin{aligned}
N^{T} &= r\frac{N_P^{T}}{r} + f\frac{N_N^{T}}{f} \\
N^{\overline{T}} &= (1-r)\frac{N_P^{T}}{r} + (1-f)\frac{N_N^{T}}{f}.
\end{aligned}
\tag{15.4}
$$

Equation 15.4 can also be expressed in the form of a matrix,

$$
\begin{pmatrix} N^{T} \\ N^{\overline{T}} \end{pmatrix} = \begin{pmatrix} r & f \\ 1-r & 1-f \end{pmatrix} \begin{pmatrix} N_P^{T}/r \\ N_N^{T}/f \end{pmatrix}. \tag{15.5}
$$

Regions that correspond to the two numbers that appear in the right-hand side vector of Equation 15.5 are referred to as the "Application Regions (ARs)", which can be constructed using experimental data. The estimation of *nonprompt* background, denoted by N_N^{T}, can be obtained by a simple matrix inversion.

15.2 Generialization and Implementation of the Matrix Method

The description in the previous section deals with a scenario where only one lepton is studied. This analysis uses a generalized version of the *matrix method*, where all three *tight* leptons are considered to be possibly *nonprompt*. Equation 15.5 is generalized as,

$$
\begin{pmatrix} N^{TTT} \\ N^{TT\bar{T}} \\ N^{T\bar{T}T} \\ N^{T\bar{T}\bar{T}} \\ N^{\bar{T}TT} \\ N^{\bar{T}T\bar{T}} \\ N^{\bar{T}\bar{T}T} \\ N^{\bar{T}\bar{T}\bar{T}} \end{pmatrix}
=
\begin{pmatrix}
r_1r_2r_3 & r_1r_2f_3 & r_1f_2r_3 & r_1f_2f_3 & f_1r_2r_3 & f_1r_2f_3 & f_1f_2r_3 & f_1f_2f_3 \\
r_1r_2(1-r_3) & r_1r_2(1-f_3) & r_1f_2(1-r_3) & r_1f_2(1-f_3) & f_1r_2(1-r_3) & f_1r_2(1-f_3) & f_1f_2(1-r_3) & f_1f_2(1-f_3) \\
r_1(1-r_2)r_3 & r_1(1-r_2)f_3 & r_1(1-f_2)r_3 & r_1(1-f_2)f_3 & f_1(1-r_2)r_3 & f_1(1-r_2)f_3 & f_1(1-f_2)r_3 & f_1(1-f_2)f_3 \\
r_1(1-r_2)(1-r_3) & r_1(1-r_2)(1-f_3) & r_1(1-f_2)(1-r_3) & r_1(1-f_2)(1-f_3) & f_1(1-r_2)(1-r_3) & f_1(1-r_2)(1-f_3) & f_1(1-f_2)(1-r_3) & f_1(1-f_2)(1-f_3) \\
(1-r_1)r_2r_3 & (1-r_1)r_2f_3 & (1-r_1)f_2r_3 & (1-r_1)f_2f_3 & (1-f_1)r_2r_3 & (1-f_1)r_2f_3 & (1-f_1)f_2r_3 & (1-f_1)f_2f_3 \\
(1-r_1)r_2(1-r_3) & (1-r_1)r_2(1-f_3) & (1-r_1)f_2(1-r_3) & (1-r_1)f_2(1-f_30 & (1-f_1)r_2(1-r_3) & (1-f_1)r_2(1-f_3) & (1-f_1)f_2(1-r_3) & (1-f_1)f_2(1-f_3) \\
(1-r_1)(1-r_2)r_3 & (1-r_1)(1-r_2)f_3 & (1-r_1)(1-f_2)r_3 & (1-r_1)(1-f_2)f_3 & (1-f_1)(1-r_2)r_3 & (1-f_1)(1-r_2)f_3 & (1-f_1)(1-f_2)r_3 & (1-f_1)(1-f_2)f_3 \\
(1-r_1)(1-r_2)(1-r_3) & (1-r_1)(1-r_2)(1-f_3) & (1-r_1)(1-f_2)(1-r_3) & (1-r_1)(1-f_2)(1-f_3) & (1-f_1)(1-r_2)(1-r_3) & (1-f_1)(1-r_2)(1-f_3) & (1-f_1)(1-f_2)(1-r_3) & (1-f_1)(1-f_2)(1-f_3)
\end{pmatrix}
\begin{pmatrix} N^{TTT}_{PPP}/r_1r_2r_3 \\ N^{TTT}_{PPN}/r_1r_2f_3 \\ N^{TTT}_{PNP}/r_1f_2r_3 \\ N^{TTT}_{PNN}/r_1f_2f_3 \\ N^{TTT}_{NPP}/f_1r_2r_3 \\ N^{TTT}_{NPN}/f_1r_2f_3 \\ N^{TTT}_{NNP}/f_1f_2r_3 \\ N^{TTT}_{NNN}/f_1f_2f_3 \end{pmatrix}.
\tag{15.6}
$$

Except for the first number, all other numbers that appear in the right-hand side vector correspond to events with at least one *nonprompt* lepton that passes *tight* selection criteria. Therefore, the overall *nonprompt* background is expresses as,

$$
N^{TTT}_{Nonprompt} = N^{TTT}_{PPN} + N^{TTT}_{PNP} + N^{TTT}_{PNN} + N^{TTT}_{NPP} + N^{TTT}_{NPN} + N^{TTT}_{NNP} + N^{TTT}_{NNN}, \tag{15.7}
$$

which can be obtained by first constructing 8 ARs to form the lefthand side vector. Secondly, the 8 $\times$ 8 matrix is constructed and inverted. Lastly, the righthand side vector can be obtained by multiplying the lefthand side vector by the inverted matrix.

Only two PDs "SingleElectron" and "SingleMuon" are used in the construction of MR in 2016 and 2017 while "SingleElectron" is replaced with "EGamma" in 2018. In addition to PDs, the measurements of r/f also utilize the $t\bar{t}$ sample and all MC samples listed under the "*prompt* background" category in Table 12.3. Depending on the flavor of the leading lepton in MC, events are selected with either a single-electron or a single-muon trigger, which is summarized in Table 15.1. Data events are selected with the same HLT triggers as well but events in "SingleMuon" ("SingleElectron" or "EGamma") PD are only accepted if the leading lepton is a muon (electron).

Both r and f are parameterized in bins of lepton p_{T}, $|\eta|$, and jet multiplicity. The bin range is optimized to retain sufficient statistics for each bin:

- Electron p_{T} bin range: {20.0, 24.6, 28.8, 33.0, 37.2, 41.4, 46.1, 52.1, 59.3, 68.3, 82.7, 110.6} GeV,
- Muon p_{T} bin range: {20.0, 23.8, 27.7, 31.3, 35.0, 38.9, 42.8, 45.6, 50.7, 59.5, 72.9, 94.3} GeV,
- $|\eta|$ bin range: {0, 0.8, 1.6, 2.4},

Table 15.1: Summary of the HLT triggers used in the measurement of r and f. These are unprescaled single-lepton triggers with the lowest p_T threshold. The threshold of the electron trigger is higher in the 2017 and 2017 datasets due to increased instantaneous luminosity in those two years.

Channel	Path	Dataset	2016	2017	2018
Electron	HLT_Ele27_WPTight_Gsf	Data & MC	✓	-	-
	HLT_Ele35_WPTight_Gsf	Data & MC	-	✓	✓
Muon	HLT_IsoMu27	Data & MC	✓	✓	✓

- Jet multiplicity: {0 jets, $\geq$ 1 jet}.

The jet multiplicity bin is a proxy for variation in the composition of physics processes. In addition to requiring at least one jet, the MR corresponds to the second jet multiplicity bin requires no more than one b-tagged jet as this is also required in the SR.

The *nonprompt* efficiency is measured in same-sign dilepton regions, in which the leading lepton in p_T, used as a *tag*, is required to be matched with trigger objects within $\Delta R < 0.2$. The sub-leading lepton is required to pass the *loose* selection and is taken as the *probe*. Events that have two same-sign electrons with an invariant mass between 76 and 106 GeV are removed from MR to suppress the backgrounds that originate from charge misidentification. No such requirement has been introduced to the muon MR due to its negligible rate of charge misidentification.

The contribution from *prompt* backgrounds, estimated from MC simulation, is subtracted from the data. A representative composition of backgrounds in MR is shown in Figure 15.2.

The fake efficiency f is calculated as:

$$f = \frac{n_{data}^{tag+tight} - n_{MC(prompt)}^{tag+tight}}{n_{data}^{tag+loose} - n_{MC(prompt)}^{tag+loose}}, \tag{15.8}$$

where the numerator is selected with one *tag* and one *tight* lepton while the denominator is selected with one *tag* and one *loose* lepton. The selection criteria for *tag*, *loose*, and *tight* lepton is summarised in Table 15.3.

The measured *nonprompt* efficiency f exhibits a dependency on flavor composition, as is shown in Figure 15.3. This dependency is treated as a source of the systematic uncertainties of the *nonprompt* estimation and is further discussed in Section 17.2.

The *prompt* efficiency r is measured in simulated $t\bar{t}$ events in opposite-sign dilepton regions. The same lepton selection listed in Table 15.2 is used to perform the *Tag-and-Probe*. The

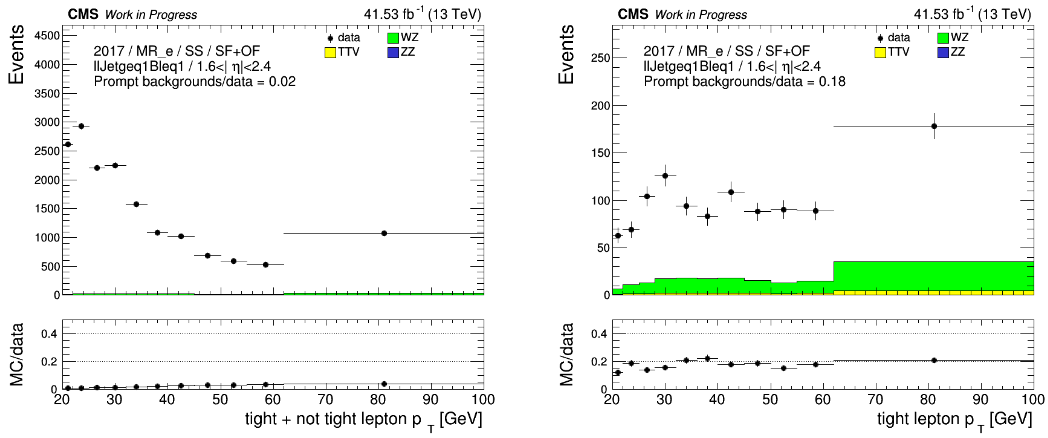


Figure 15.2: Distribution of lepton p_T in a representative electron *nonprompt* efficiency MR. In this particular example, both ee and μe flavor composites are considered. At least one jet and at most one b-tagged jet are required (the second jet multiplicity bin). *Probe* electron is required to have $1.6 < |\eta| < 2.4$ (the third η bin). Contamination from *prompt* backgrounds are estimated with MC simulation, and are shown as histograms. The data are shown as filled points. From left to right: *loose* (i.e. *tight + not tight*) electron p_T, *tight* electron p_T.

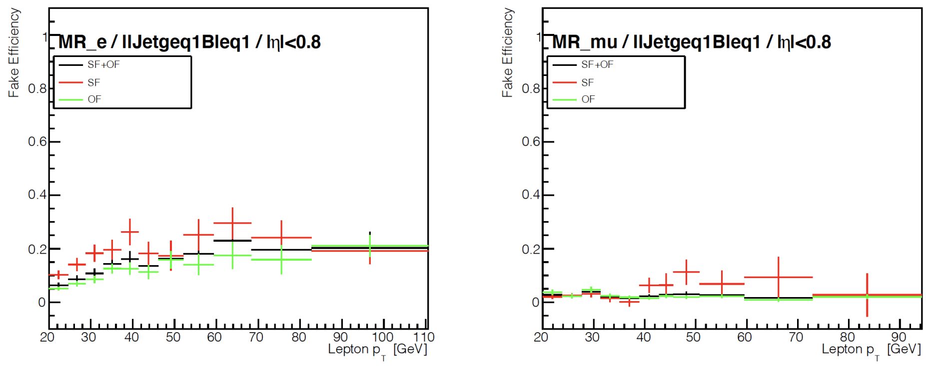


Figure 15.3: Representative *nonprompt* electron efficiency measured in data events. From left to right: electron f, muon f. Events with a same-flavor lepton pair are shown in red points while events selected with a different-flavor lepton pair are shown in green points. Events with a same-flavor or different-flavor lepton pair are shown in black points. These plots correspond to the first $|\eta|$ bin ($|\eta|$ <0.8) and the second jet multiplicity bin. Events selected Error bars displayed in these plots include statistical uncertainty only.

leading lepton in p_T is used as a *tag* while the oppositely charged sub-leading lepton is taken as a *probe*. The variation of r between different flavor compositions is negligible, as is shown in Figure 15.4. Therefore, only eμ events are used to measure *prompt* efficiency in order to minimize the contamination of *nonprompt* leptons.

Table 15.2: Summary of the lepton selections needed for the measurement of r and f. Please note: (i) the minimum p_T cut for *tag* electron in the 2016 dataset is reduced to 30 GeV to adjust for the trigger threshold, and (ii) the *tight* selection here is the same as the *tight* lepton selection described in Section 13.1.

Lepton	Selection	*loose*	*tag*	*tight*[ii]
Electron	p_T	$>$ 20 GeV	$>$ 38 GeV[i]	$>$ 20 GeV
	$I_{\text{mini}}^{\text{rel}}$	$<$0.4	$<$0.1	$<$0.12
	TOP LeptonMVA	$>$-0.9	$>$0.95	$>$0.9
	Match with trigger objects	-	✓	-
Muon	p_T	$>$ 20 GeV	$>$ 30 GeV	$>$ 20 GeV
	$I_{\text{mini}}^{\text{rel}}$	$<$0.4	$<$0.1	$<$0.12
	Cut-based ID	-	Medium WP	Medium WP
	TOP LeptonMVA	$>$0.5	$>$0.9	$>$0.9
	Match with trigger objects	-	✓	-

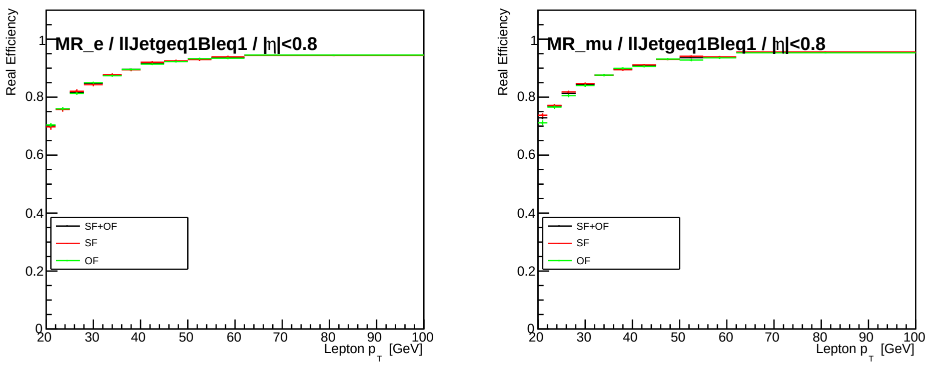


Figure 15.4: Representative *prompt* efficiency measured in simulated $t\bar{t}$ events. From left to right: electron r, muon r. Events with a same-flavor lepton pair are shown in red points while events selected with a different-flavor lepton pair are shown in green points. Events with a same-flavor or different-flavor lepton pair are shown in black points. These plots correspond to the first $|\eta|$ bin ($|\eta|$ <0.8) and the second jet multiplicity bin. Error bars displayed in these plots include statistical uncertainty only.

The selection criteria for various MRs is summarised in Table 15.3.

Table 15.3: Summary of the selection criteria applied to the measurement regions of r and f. "OffZ" means events containing two same-sign electrons with an in variant mass between 76 and 106 GeV are removed. C_i denotes the electric charge of the selected lepton.

Observable	jet bin	# of selected leptons	lepton flavor composite	$\lvert\sum_i C_i\rvert$	OffZ	njet	nbjet
f	0 jet	2	any	2	same-sign ee	$=0$	$=0$
	1 or more jet	2	any	2	same-sign ee	≥ 1	≤ 1
r	0 jet	2	eμ only	0	-	$=0$	$=0$
	1 or more jet	2	eμ only	0	-	≥ 1	≤ 1

15.3 Validation of the Matrix Method

The performance of the *matrix method* is validated using three regions that are tangential to the SR, referred to as VRs. In these VRs, *prompt* backgrounds are estimated using MC simulation while *nonprompt* background is estimated with the *matrix method*. A summary of the selections applied to these VRs is given in Chapter 14.

Distribution of the leading lepton η and jet multiplicity are shown in Figure 15.5. Good agreement between data and background estimate has been observed in all three VRs.

15.4 Nonprompt Estimate in SR

The *matrix method* is used to estimate *nonprompt* background in the SR. Distributions of the LFV eμ mass and the Z boson mass are shown in Figure 15.6. When compared to the background estimate from pure MC simulation (Figure 14.1), the updated background template is smoother with lower statistical uncertainties.

The number of expected events from various kinds of backgrounds is shown in Table 15.4.

Table 15.4: Expected background contributions and the number of events observed in data collected during 2016–2018. The statistical and systematic uncertainties are added in quadrature. The category "other backgrounds" includes smaller background contributions containing one or two top quarks plus a boson or quark. The CLFV signal, generated with $C^{vector}_{e\mu tu}/\Lambda^2 = 1\text{TeV}^{-2}$, is also listed for reference. The signal yields include contributions from both top production and decay modes.

Process	m(eμ)<150 GeV	m(eμ)>150 GeV
Nonprompt	351 ± 92	146 ± 38
WZ	275 ± 64	145 ± 35
ZZ	33.2 ± 6.5	13.1 ± 2.6
VVV	17.0 ± 8.5	12.0 ± 6.0
$t\bar{t}W$	47.6 ± 10.0	40.0 ± 9.1
$t\bar{t}Z$	39.1 ± 7.9	25.8 ± 5.4
$t\bar{t}H$	28.2 ± 4.5	10.0 ± 1.6
tZq	5.5 ± 1.1	2.5 ± 0.5
Other	7.3 ± 3.7	4.5 ± 2.3
Total expected	805 ± 123	398 ± 57
Data	783	378
CLFV	207 ± 15	4440 ± 215

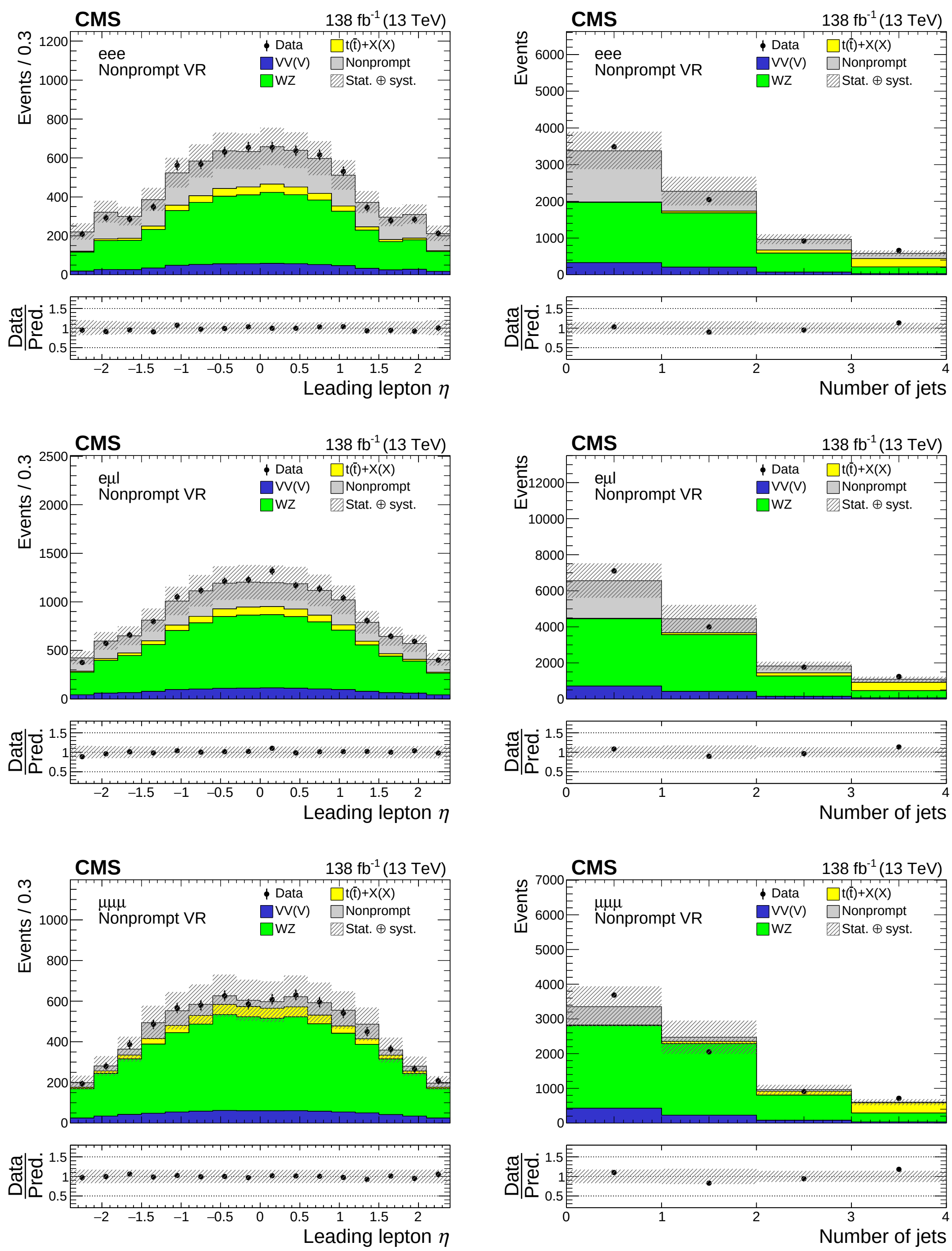


Figure 15.5: Distributions of the leading lepton η (left column) and the jet multiplicity (right column) in the *nonprompt* VRs. Events in the eee, eμℓ, and μμμ *nonprompt* VRs are shown in the upper, middle, and lower row, respectively. The data are shown as filled points and the SM background predictions as histograms. The *nonprompt* background is estimated using control samples in data, while other backgrounds are estimated using MC simulation. The hatched bands indicate statistical and systematic uncertainties for the SM background predictions. The last bin of the right histogram includes the overflow events.

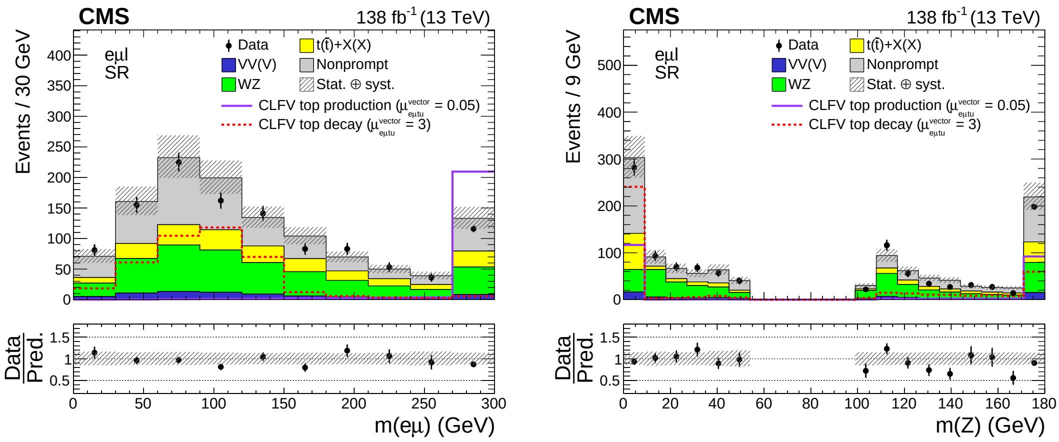


Figure 15.6: Distributions of the LFV eμ mass (left) and the Z boson mass (right) in SR. The data are shown as filled points and the SM background predictions as histograms. The *nonprompt* background is estimated using control samples in data, while other backgrounds are estimated using MC simulation. The hatched bands indicate statistical and systematic uncertainties for the SM background predictions. The normalization of the signal processes is chosen arbitrarily for improved visualization. The last bin of both histograms includes the overflow events.

Chapter 16

Signal Extraction with Boosted Decision Trees

A MVA is performed in SR to further separate the LFV signals from the backgrounds, and enhance the sensitivity of this analysis. More specifically, a dozen of discriminating variables, referred to as "features", are selected and combined by a gradient-BDT, which is implemented using the XGBoost package [33]. There are several reasons why BDT is chosen: (i) the goal of the MVA is to achieve maximum separation between signals and backgrounds using a small number of already well-separated kinematic variables, instead of exploring some complicated structures hidden in event topology, (ii) under such a goal, the potential performance gain from a more sophisticated algorithm like a NN is limited, (iii) a BDT-based algorithm is straightforward to implement and consumes only a moderate amount of computational resources, and (iv) the interpretability of a BDT-based algorithm is excellent.

The top production and decay signals are longer distinguished by the BDT. They are combined into a single signal class, just like all backgrounds are combined into a single background class. The training of the BDT depends entirely on MC samples that are statistically orthogonal to the samples used in the actual background estimation. More details on the configurations of the BDT are described in Section 16.1. The input features are described in Section 16.2. The output of the BDT is presented in Section 16.3.

16.1 BDT Configuration

The LFV eμ mass of the top decay signal is bounded by the SM top quark mass, as is shown in Figure 14.1. On the contrary, the LFV eμ mass of the top production signal has no such restriction and often reaches TeV level. Therefore, a 150 GeV threshold is used to divide the SR into two SRs enriched in different signal modes. The MVA strategy is to combine two signal modes within each SR and train binary BDTs separately for each SR. In other words, only two signal datasets and two background datasets are needed.

Other aspects of the signal MC samples, such as the Lorentz structure and the flavor of the up-type quark involved in LFV interaction, are shown to have a relatively small impact on the kinematics of final state particles, as is shown in Figure 16.1. Therefore, they are not distinguished by the BDT. The selection criteria used to define SR, described in Section 14.1, are used to preselect events before the construction of both signal and background datasets.

The construction of the signal datasets takes a few steps. Firstly, the cross-sections of all top production signal samples, regardless of the Lorentz structure or the quark flavor, are set to the cross-section of the vector-like top production signal with an eμtu vertex, which is shown in Table 12.2. This is done to remove potential bias towards the signals with higher cross sections. Similarly, the cross-sections of all top decay signal samples are set to the cross-section of the vector-like top decay signal. For each sample, a normalization weight is calculated and is used to replace the original normalization component of the MC weights. These updated MC weights are eventually passed on to the BDT to weight each signal event. Secondly, all top production and decay signal samples are combined into a single dataset, which is then subdivided into two datasets using a 150 GeV threshold on LFV eμ. The last step is to adjust the overall normalization (i.e. sum of the MC weights) of each of the two signal datasets to match the overall normalization of the corresponding background dataset.

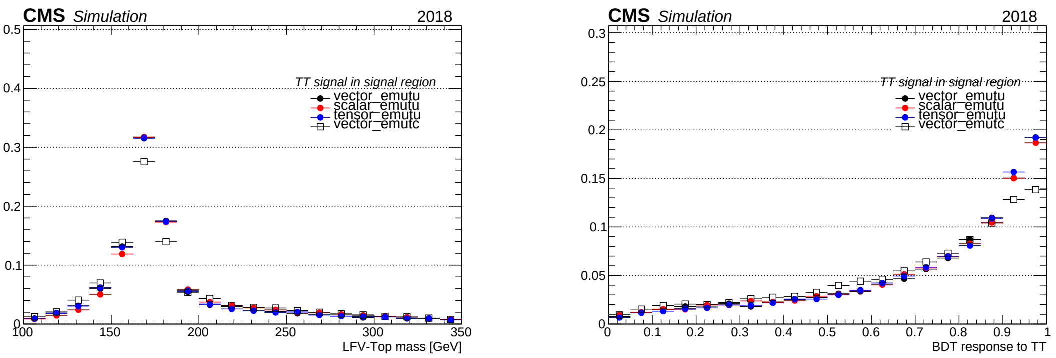
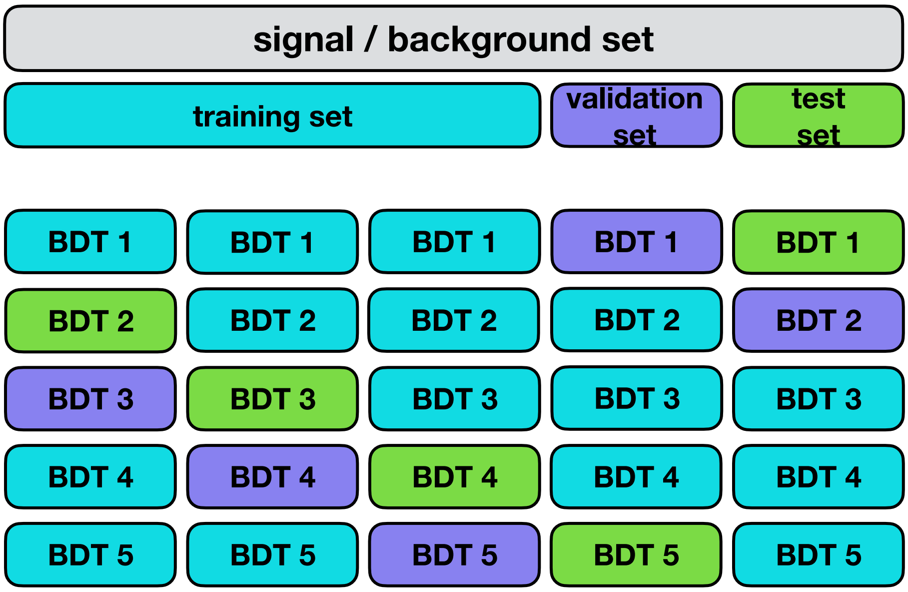


Figure 16.1: Normalized distributions of the simulated top quark decay signals in SR1 using the 2018 dataset. From left to right: LFV top mass, BDT output. In the legend of these histograms, "vector", "scalar", and "tensor" denote the Lorentz structures of EFT operators, and "emutu(c)" denote the eμtu(c) interaction vertex.

The *prompt* background dataset is constructed by combining all MC samples listed under the "prompt" category listed in Table 12.3. Cross sections referenced Table 12.3 are directly used to normalize the *prompt* backgrounds. The construction of the *nonpeompt* background dataset is different since the *nonprompt* backgrounds are modeled with the *matrix method*, which is itself constructed from 8 ARs. Therefore, 8 ARs are constructed to collect simulated $t\bar{t}$ and Drell-Yan events. These events are used to form the *nonprompt* dataset. Each event in the *nonprompt* dataset is then "weighted" using the output of the *matrix method*. Finally, the *nonprompt* dataset is combined with the *prompt* dataset and then divided into two datasets using a 150 GeV threshold on LFV eμ mass.

A technique known as the "*k*-fold cross validation" is used to minimize the loss of statistics when partitioning datasets into training, validation, and test sets. For each targeted SR, the corresponding signal/background set is divided evenly into five subsets. Three out of the five subsets are used in the training while a fourth subset is used as a validation set. The fifth set is used to test the performance of the trained BDT. A second BDT is trained using a different combination of subsets to form training, validation, and test sets. This process is repeated five times until a unique test set no longer exists, which is illustrated in Figure 16.2. This technique ensures that the test set is always statistically independent of the process of parameters tuning, which serves as the basis for the bias-free evaluation after training: when evaluating each event using the trained model, it is always possible to pick one of the five BDTs where this particular event was not included in the training or validation.

Figure 16.2: Illustration of a 5-fold cross-validation. In this setup, five BDTs are trained/tested using the same dataset arranged in different configurations. Each of the bottom five rows represents the configuration of a BDT.



The same set of hyperparameters is used for all BDTs, which are optimized using a randomized grid search algorithm. The number of estimators is set to 1000 with a max depth of 5. The standard loss function implemented in [33] is used as the evaluation metric. The performance of the BDT is visualized using a metric known as "ROC curve", which is shown in Figure 16.3. In general, the BDTs trained in SR2 (i.e. $m(e\mu) > 150$ GeV) are much more performant due to the high p_T objects in the final states.

16.2 BDT Features

The discriminating variables used as input in training are referred to as "features" in this analysis. A total of 14 features are used for BDT trained in SR1 and SR2. The names and descriptions of these 14 features are summarized in Table 16.1. Many of these features are derived from reconstructed heavy objects which are described in Section 14.3.

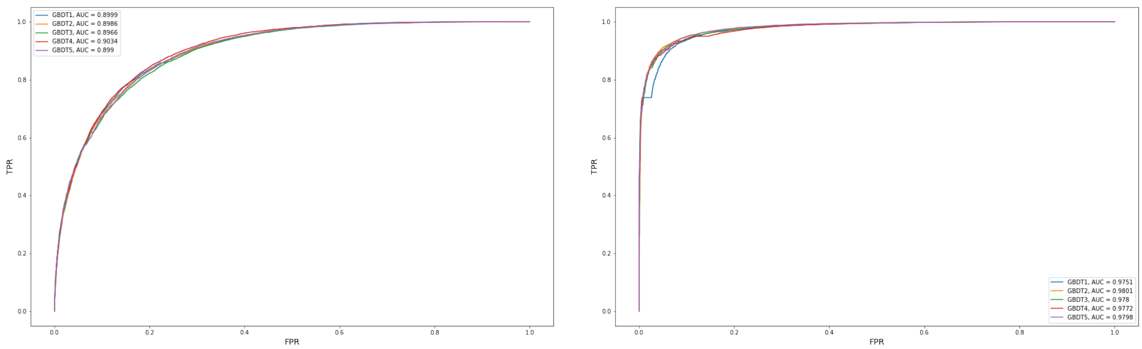


Figure 16.3: ROC curves extracted using the test sets specified in the 5-fold cross-validation. The left (right) figure shows the ROC curves of the BDTs trained in SR1 (SR2). The area under the ROC curves are showed in the legends.

Four additional features are added to the BDT trained in SR1. The "MVA_JeDr" and "MVA_JeDr" variables are defined by using the jet that forms the LFV top quark candidate and calculating the opening angle between this jet and the LFV leptons. It is expected that this angle is smaller in the LFV decay mode than LFV production mode. Two additional features are added to the BDT trained in SR2. A description of how the standalone lepton is determined can be found in Section 14.3.

Distributions of selected features are shown in Figure 16.4-16.8. Distributions of the full list of features can be found in Appendix B. The relative importance of these features is extracted using the "gain" method and is shown in Figure 16.9. The correlations between different features are shown in Figure 16.10-16.11.

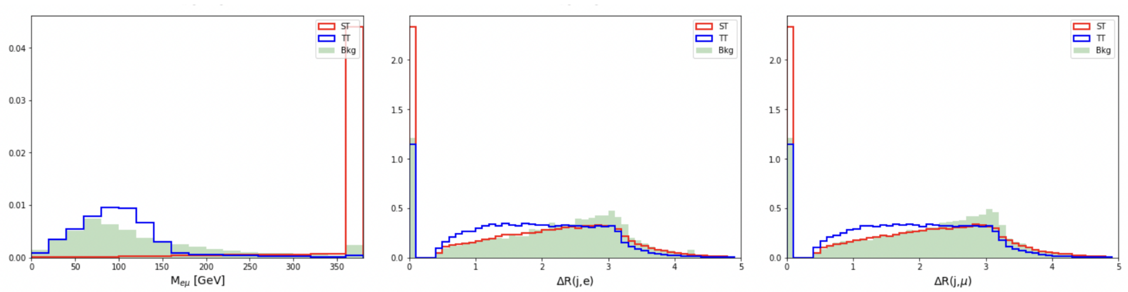


Figure 16.4: Normalized distribution of various features in SR. From left to right: MVA_M$_{e\mu}$, MVA_JeDr, MVA_JeDr.

16.3 BDT Output

The output of the BDTs in SRs are shown in Figure 16.12. The *nonprompt* background is estimated with the matrix method. Other backgrounds are estimated with MC simulation.

Table 16.1: Features shared by BDTs trained in both SR1 and SR2.

Name	Description
MVA_Memu	invariant mass of the LFV-eμ pair
MVA_LFVePt	p_T of the LFV electron
MVA_LFVmuPt	p_T of the LFV muon
MVA_LFVTopmass	invariant mass of the LFV top quark candidate
MVA_Zmass	invariant mass of Z boson candidate
MVA_Jet2Btag	b-tagging score of the jet with the second highest b-tagging score
MVA_Mbl2	invariant mass of the second $m_{b\ell}$ system
MVA_njet	number of jets
MVA_nbjet	number of b-tagged jets
MVA_tM	transverse mass of the W boson candidate (from the SM top quark)
MVA_llDr	ΔR between LFV electron and LFV muon
MVA_SSee_Zmass	invariant mass of a Same-Sign di-electron pair
MVA_Topmass	invariant mass of the SM top quark candidate
MVA_Met	missing transverse momentum (MET)

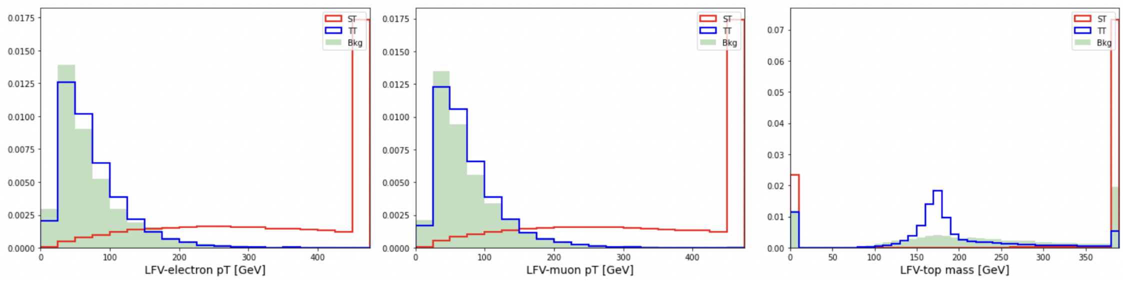


Figure 16.5: Normalized distribution of additional features in SR. From to left to right: MVA_LFVePt, MVA_LFVmuPt, MVA_LFVTopmass.

Table 16.2: Features only used by BDT trained in SR1

Name	Description
MVA_Ht	scalar sum of the p_{T} of all jets and leptons
MVA_Mbl1	invariant mass of the second $\mathrm{m}_{\mathrm{b}\ell}$ system
MVA_JeDr	ΔR between LFV electron and a light jet (non b jet)
MVA_JmuDr	ΔR between LFV muon and a light jet (non b jet)

Table 16.3: Features only used by BDT trained in SR2

Name	Description
MVA_BaPt	p_{T} of the standalone lepton
MVA_JetHt	scalar sum of the p_{T} of all jets

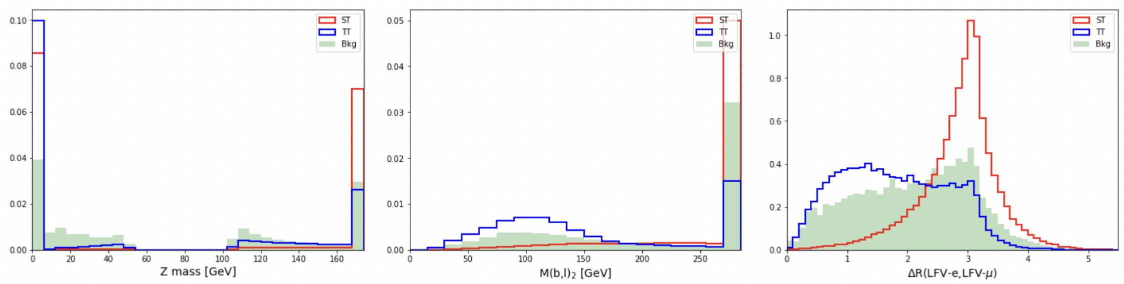


Figure 16.6: Normalized distribution of additional features in SR. From left to right: MVA_Zmass, MVA_Mbl2, MVA_llDr.

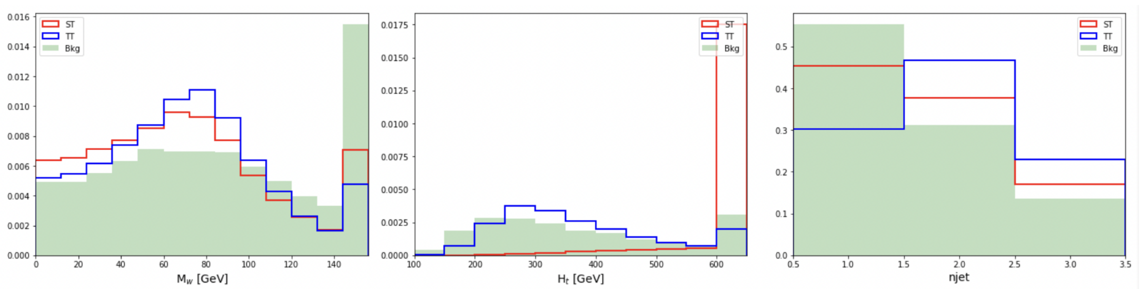


Figure 16.7: Normalized distribution of additional features in SR. From left to right: MVA_tM, MVA_Ht, MVA_njet.

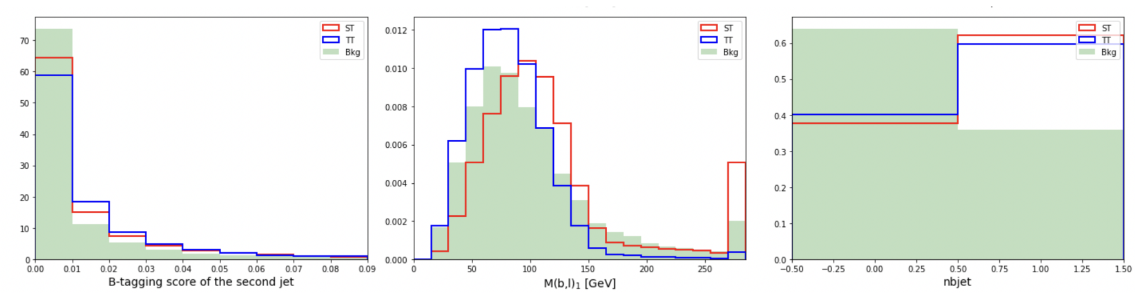


Figure 16.8: Normalized distribution of additional features in SR. From left to right: MVA_Jet2Btag, MVA_Mbl1, MVA_nbjet.

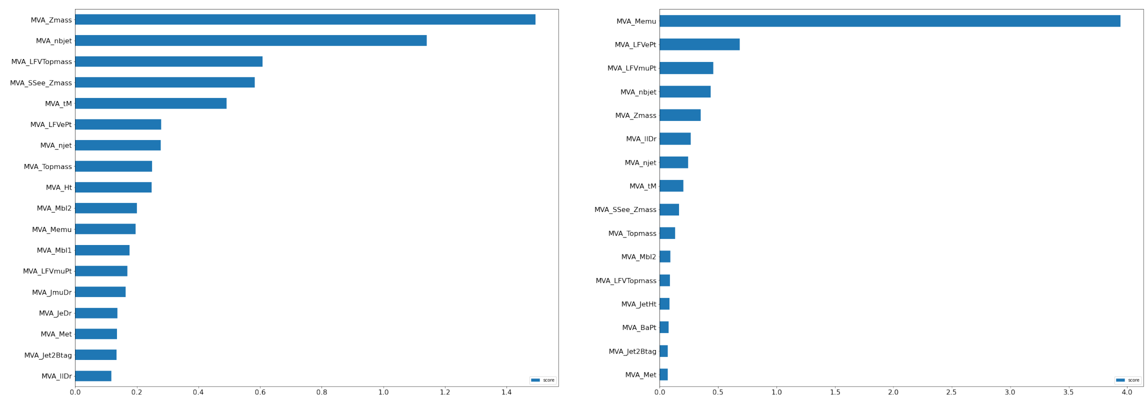


Figure 16.9: List of features ranked by their relative importance. From left to right: BDT targeting TT (SR1), BDT targeting ST (SR2)

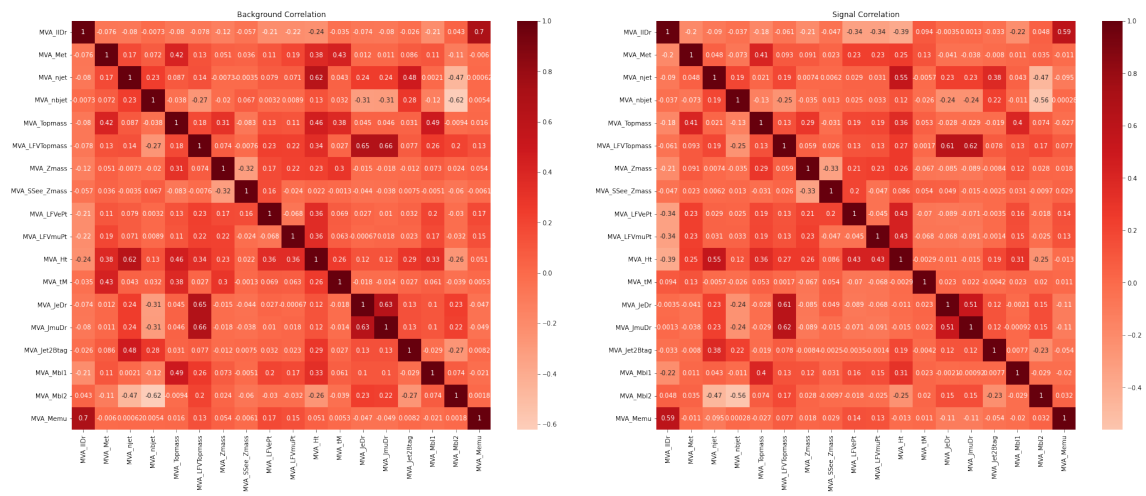


Figure 16.10: Correlation matrices (SR1), from left to right: background correlation, signal correlation.

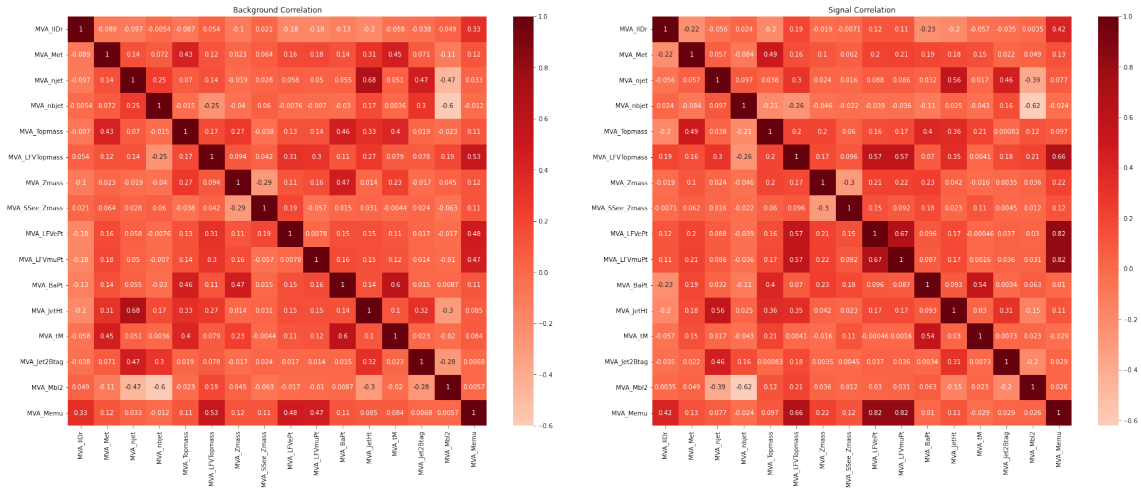


Figure 16.11: Correlation matrices (SR2), from left to right: background correlation, signal correlation.

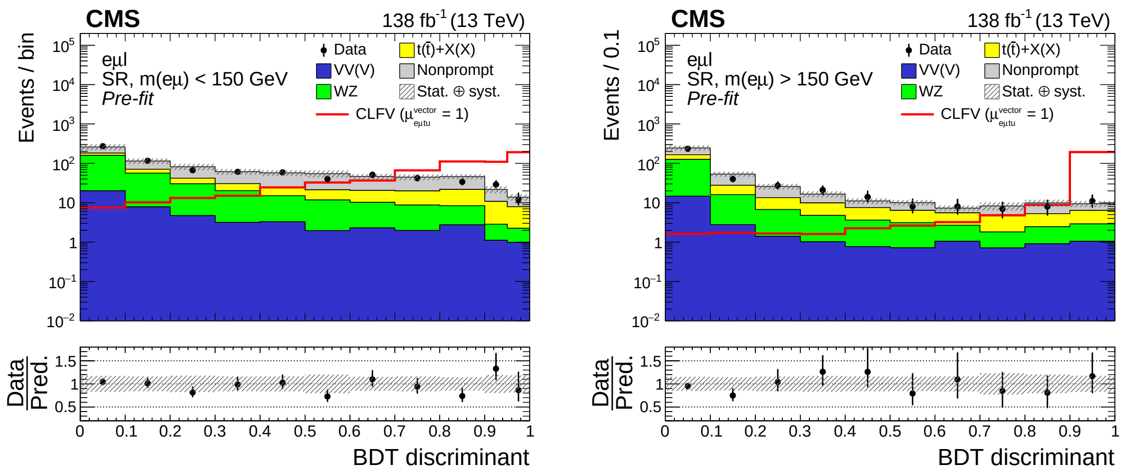


Figure 16.12: Distributions of the BDT discriminator targeting the CLFV top quark decay (left) and production (right) signal. Contributions from the two signal modes (production and decay) are combined within each SR and are shown as the solid red line. The pre-fit signal strength ($\mu^{\text{vector}}_{\text{e}\mu\text{tu}} = 1$), corresponding to $C^{\text{vector}}_{\text{e}\mu\text{tu}}/\Lambda^2 = 1\ \text{TeV}^{-2}$, is used to normalise the signal cross sections. The hatched bands indicate statistical and systematic uncertainties for the background predictions.

Chapter 17

Systematic Uncertainties

Different sources of systematic uncertainty contribute to the estimation of background events and modeling of the signal. The chapter is organized as follows. Theoretical uncertainties concerning signals and major backgrounds are discussed in Section 17.1. Uncertainties concerning the *nonprompt* background are discussed in Section 17.2. Uncertainties concerning the modeling of the diboson processes are described in Section 17.3. Finally, other systematic uncertainties are discussed in Section 17.4.

17.1 Theoretical Uncertainties

Variations on theoretical cross sections for *prompt* backgrounds are introduced to cover the uncertainties in perturbative QCD calculations. A 6% normalization uncertainty is assigned to WZ and ZZ processes [22]. A 15% normalization uncertainty is assigned to $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$ processes [23, 24]. A 20% normalization uncertainty is assigned to tZq process, which is a conservative estimate taken from the MC generator. A conservative 50 % normalization uncertainty is assigned to other smaller *prompt* backgrounds. All normalization uncertainties are considered uncorrelated between different processes but correlated across the years.

Uncertainties associated with the PDF are evaluated by using 100 replicas of the NNPDF sets [16, 17]. The procedure described in [34] is followed. Firstly, the sum of the generator weights of each replica is normalized to the nominal sum of the generator weights. This is done before any event selection to ensure no additional normalization effect is introduced. After the previous step, the bin-by-bin variations of the BDT templates are obtained by calculating the bin-by-bin difference of the BDT templates when switching from nominal PDF to each PDF replica. Finally, PDF uncertainty for each bin is assigned by taking the root mean square value of the 100 variations of the corresponding bin. This uncertainty is treated as uncorrelated between different processes but correlated across the years. This uncertainty is considered for all the signals and major prompt backgrounds (i.e. WZ, $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$).

QCD scale uncertainties are evaluated by varying the renormalization scale μ_R and factorization scales μ_F in ME. A total of six variations are considered: varying μ_R by a factor of 2 and 0.5, varying μ_F by a factor of 2 and 0.5, and varying μ_R and μ_F simultaneously by a factor of 2 and 0.5.

Similar to PDF uncertainty, the normalization effects of each variation are removed. An envelope that covers all six variations is used to represent the scale uncertainty. This uncertainty is treated as uncorrelated between different processes but correlated across the years. This uncertainty is considered for all the signals and major prompt backgrounds (i.e. WZ, $t\bar{t}W$, $t\bar{t}Z$, and $t\bar{t}H$).

Uncertainties associated with the Parton Shower (PS) are evaluated by varying the renormalization scale μ_R in the initial and final state radiations, which effectively changes the strong coupling constant in the PS. Similarly, μ_R is varied by a factor of 2 and 0.5, and the normalization effects of each variation are removed. This uncertainty is treated as uncorrelated between different processes but correlated across the years. This uncertainty is only considered for signal events.

17.2 Nonprompt Uncertainties

There are several sources of uncertainties associated with the determination of the *nonprompt* efficiency f. One of these uncertainties comes from the estimate of *prompt* contamination in MR. As is discussed in Chapter 15, *prompt* backgrounds (estimated with MC) are subtracted from total event yields measured in data. A flat 20 % uncertainty (α in Equation 17.1) is assigned to the event yields of the *prompt* background and the resulting variation of f is taken as the uncertainty.

$$f = \frac{n_{data}^{tag+tight} - (1+\alpha)n_{MC(prompt)}^{tag+tight}}{n_{data}^{tag+loose} - (1+\alpha)n_{MC(prompt)}^{tag+loose}}. \tag{17.1}$$

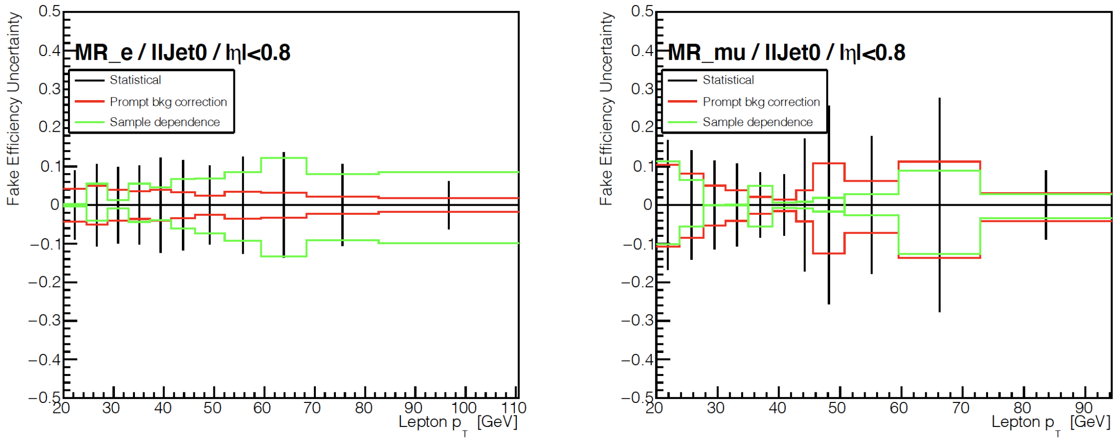


Figure 17.1: Comparison of different components of the uncertainties associated to the *nonprompt* efficiency measured in 2017 dataset (njet=0 bin, $|\eta|$ <0.8 bin). From left to right: electron f uncertainty, muon f uncertainty.

Another source of uncertainty associated with the determination of *nonprompt* efficiency f is concerned with the observation that f exhibits a flavor dependency, as is shown in Figure 15.3. This can happen when different physics processes enter MRs with different lepton flavor composites, which lead to differences in *nonprompt* lepton behaviors. This type of uncertainty, referred to as "sample dependence", is estimated by introducing a variation factor β between the proportions of same-flavor and different-flavor pairs in MR. For example, electron f can be calculated as (prompt background correction is ignored from the equation),

$$f_{\mathrm{e}} = \frac{(1+\beta)n_{\mathrm{e+e}}^{tag+tight} + (1-\beta)n_{\mathrm{e+\mu}}^{tag+tight}}{(1+\beta)n_{\mathrm{e+e}}^{tag+loose} + (1-\beta)n_{\mathrm{e+\mu}}^{tag+loose}}. \tag{17.2}$$

A 20% variation (β) is assigned the resulting variation of f is taken as the uncertainty.

Statistical uncertainty is also considered when determining f. A comparison of different sources of uncertainties is shown in Figure 17.1 and Figure 17.2. All sources of uncertainties are added in quadrature to form the final uncertainty on f.

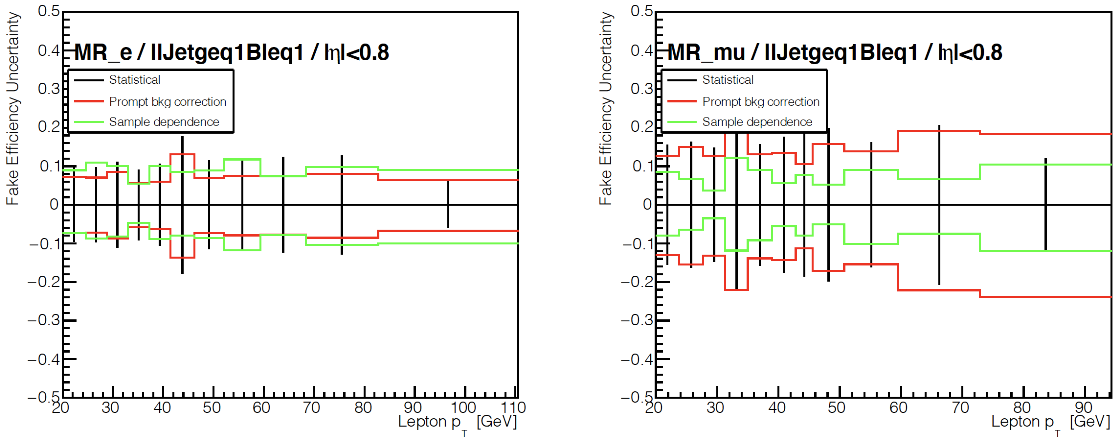


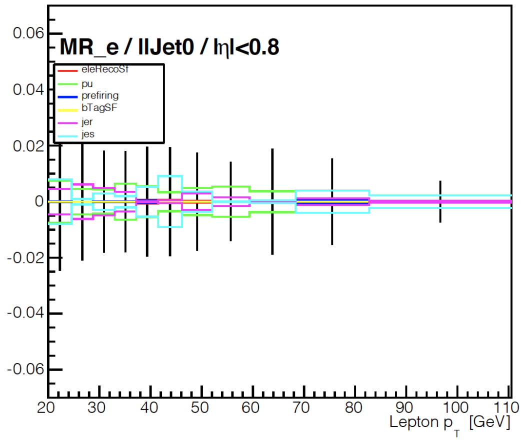
Figure 17.2: Comparison of different components of the uncertainties associated to the *non-prompt* efficiency measured in 2017 dataset (njet>0 bin, $|\eta|$ <0.8 bin). From left to right: electron f uncertainty, muon f uncertainty.

Since the *prompt* efficiency r is measured in simulated $t\bar{t}$ events, MC uncertainties described in Section 17.4 are propagated to r as the uncertainties. Additionally, statistical uncertainty is added in quadrature to the MC uncertainties to form the final uncertainty on r.

The uncertainties associated with the *prompt* efficiency are relatively small when compared to the *nonprompt* efficiency uncertainties. A comparison of different sources of *prompt* efficiency uncertainties is shown in Figure 17.3.

Uncertainties associated to r and f are determined separately for electron and muon. Therefore, there are four independent uncertainties: r_{e}, r_{μ}, f_{e} and f_{μ}.

A fifth uncertainty is considered that accounts for the potential bias caused by the way the generalized *matrix method*is implemented. Four out of the eight ARs that appear on the lefthand side of the Equation 15.6 (i.e. $N^{\overline{T}TT}$, $N^{\overline{T}T\overline{T}}$, $N^{\overline{T}\overline{T}T}$, $N^{\overline{T}\overline{T}\overline{T}}$) are selected by requiring the



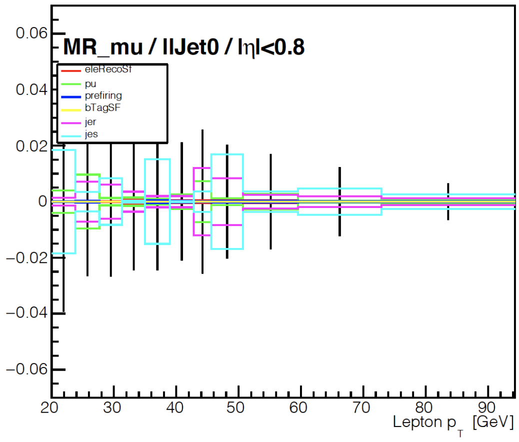


Figure 17.3: Comparison of different components of the uncertainties associated to the *prompt* efficiency measured in 2017 dataset (njet=0 bin, $|\eta|$ <0.8 bin). From left to right: electron r uncertainty, muon r uncertainty.

leading lepton in p_{T} to fail the *tight* criteria described in Table 15.2. Effectively this means that the isolation requirement is reversed for leading lepton that enter these four ARs. Selecting the leading lepton by a loose requirement is not ideal since the leading lepton is required to match with iso-triggers. To account for this bias, a 50 % uncertainty is assigned to the f_1 (*nonprompt* efficiency associated with the leading lepton) for events that enter these four ARs. The variation of the *nonprompt* estimate due to trigger matching is largely covered by this uncertainty, as is shown in Figure 17.4.

The five components of the uncertainties discussed in this section are propagated through the matrix inversion. The resulting variations of the *nonprompt* estimates are taken as the uncertainties, which contain both normalization and differential effects to the BDT templates. These uncertainties are treated uncorrelated between different components but correlated across the years. In addition to these five uncertainties, an overall normalization uncertainty of 10% is assigned to cover any other potential variations of the *nonprompt* backgrounds.

17.3 Diboson Uncertainties

Mismodeling of the jet multiplicity is observed in WZ control region, as is shown in Figure 14.3. This is largely due to the fact WZ process is modeled at LO with one extra parton in the ME. Any other extra jets are modeled by the parton shower, which is suboptimal when compared to the modeling from ME. To take this into account, a dedicated jet-dependent uncertainty is assigned to each event. This uncertainty is determined using diboson VR that has the same OnZ requirement as the WZ VR, no jet multiplicity requirement, a MET $>$ 85 GeV requirement, and a requirement of no b-tagged jets with $p_{\mathrm{T}} > 20$ GeV. Unlike for the WZ VR, events with

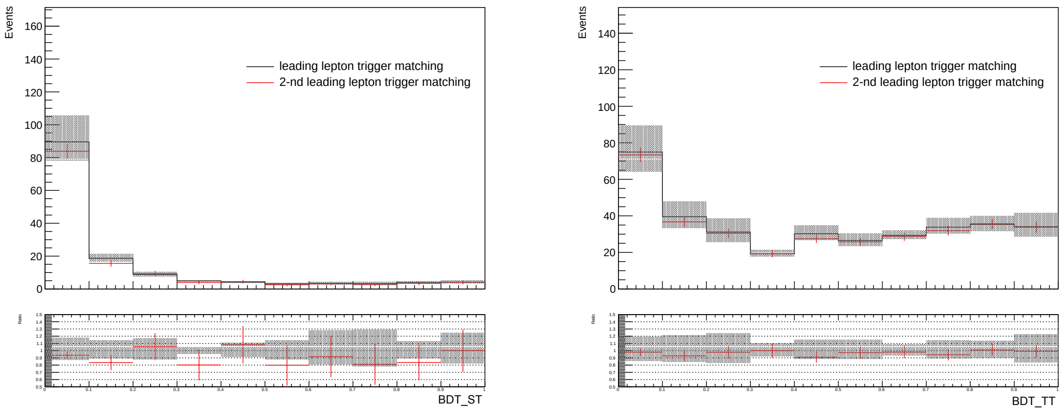


Figure 17.4: The impact of matching leptons to trigger objects on *nonprompt* estimate. From left to right: *nonprompt* estimate in top production enriched SR, *nonprompt* estimate in top decay enriched SR. The nominal configuration of the *matrix method*is to match the leading lepton with trigger objects. Matching the sub-leading with the trigger objects is taken as an alternative to evaluating the robustness of the *nonprompt* estimate. The uncertainty band only covers the variation of the *nonprompt* estimate as a result of varying leading lepton f by 50 %. Uncertainty bars only include statistical uncertainties.

different lepton flavor compositions are combined.

The jet multiplicity distributions in diboson VR are shown in Figure 17.5. For each year, a scale factor parameterized as bins of jet multiplicity is derived,

$$\epsilon = \frac{N_{data} - N_{\mathrm{VVV}} - N_{\mathrm{t\bar{t}+X(X)}} - N_{\mathrm{t\bar{t}}} - N_{others}}{N_{\mathrm{VV}}}. \tag{17.3}$$

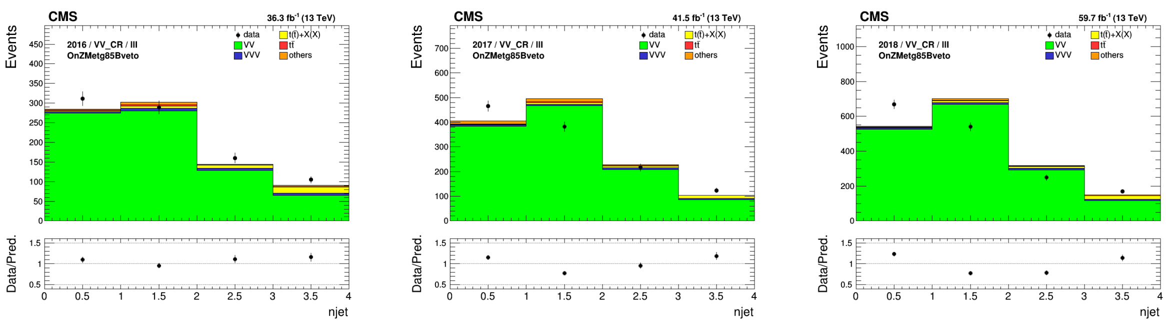


Figure 17.5: The diboson VRs, from left to right: 2016, 2017 and 2018 datasets.

The scale factor ϵ is used to estimate the uncertainty, denoted by Δ,

$$\Delta = |1 - \epsilon| \tag{17.4}$$

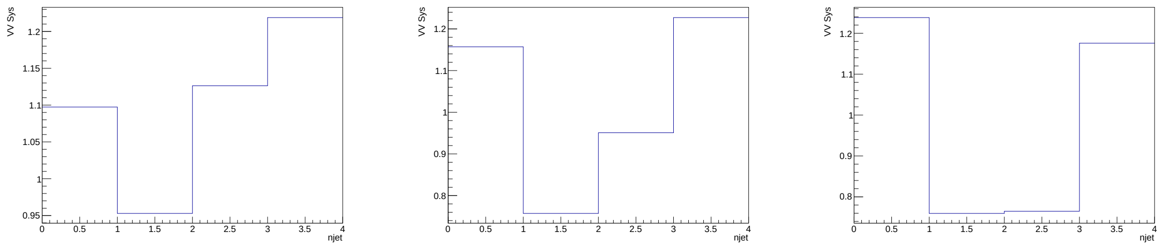


Figure 17.6: Scale factors derived from the diboson VRs, from left to right: 2016, 2017 and 2018 datasets.

This uncertainty shifts the predictions of WZ and ZZ processes by up to 20%, as is shown in Figure 17.6.

17.4 Other Experimental Uncertainties

Uncertainties of 1.2, 2.3, and 2.5% are assigned to the integrated luminosity for 2016, 2017, and 2018, respectively [35, 36, 37]. These uncertainties affect the normalization of the BDT templates of all signals as well as *prompt* backgrounds. The correlation between these uncertainties is taken into account when combining the 2016-2018 datasets.

PU distributions of all signals and *prompt* backgrounds are reweighted using per-event scale factors to recreate the PU profile measured in data. The uncertainties associated with these scale factors are evaluated by varying the inelastic pp cross-section by ± 4.6 [38]. These uncertainties are considered correlated across the years.

Calibrations of the reconstruction of electrons and muons are done centrally at CMS by using a tag-and-probe approach in DY enriched dilepton events. Per-object scale factors are used to correct for the discrepancy between reconstruction efficiencies measured in data and MC. Limited sample size as well as the choice of fit models contribute to the uncertainties associated with these scale factors. These uncertainties are considered correlated across the years.

The TOP LeptonMVA covers both identification and isolation of *prompt* leptons. Similar to lepton reconstruction, the calibration of TOP LeptonMVA is done using a tag-and-probe approach in DY enriched dilepton events. Per-object scale factors are used to correct for the discrepancy between reconstruction efficiencies measured in data and MC. Uncertainties of these scale factors are divided into two separate uncertainties: the statistical components of these uncertainties are treated as uncorrelated across the years while the other components are merged and treated as fully correlated across the years. For high p_T electrons and muons ($p_T > 200$ GeV), an additional uncertainty, denoted by "eleIDHighPt/muIDHighPt", is assigned and it increases linearly from 0 to 10% (200 GeV-1000 GeV) and is capped at 10% after 1000 GeV. These additional uncertainties are introduced because the efficiency calibration is largely done in low p_T phase space. This additional uncertainty is considered correlated across the years.

Calibrations of energy scale and resolution of electrons are done centrally at CMS [39] and no uncertainties are considered as they are largely negligible. Calibrations of muon energy scale and resolution are done using the "Rochester algorithm" [40] for muons with $p_T < 200$ GeV. "MuonScale" is used to denote the uncertainties associated with this correction, which comes primarily from a limited sample size. For muons with $p_T > 200$ GeV, no corrections are applied as there are not enough events for a robust correction from the "Rochester algorithm". An additional uncertainty, also denoted by "MuonScale" is assigned to the momentum of these high p_T muons using the "Generalized Endpoint method" [41]. The "MuonScale" uncertainty is considered correlated across the years.

No calibration is done for trigger efficiency as they are generally close to 1 in both data and MC. A flat 2% uncertainty is assigned to all signals and *prompt* backgrounds to cover statistical fluctuations. This uncertainty is treated as uncorrelated across the years.

Calibrations of the DeepJet scores are described in Section 13.2. Uncertainties associated with the calibrations are divided into 8 different sources to properly account for the correlations, which are summarized in Table 17.1. For b and udsg jets, lf, hf, hfstats1/2, and lfstats1/2 uncertainties are applied. For c jets, cferr1/2 uncertainties are applied.

Table 17.1: A hyphen (–) denotes that a source is not correlated across the years.

Source	Correlated	Description
lf	✓	udsg+c jets in heavy flavor region
hf	✓	b+c jets in light flavor region
hfstats1	-	Linear fluctuations of c jets
hfstats2	-	Quadratic fluctuations of c jets
lfstats1	-	Linear fluctuations of udsg jets
lfstats2	-	Quadratic fluctuations of udsg jets
cferr1	✓	Linear fluctuations of c jets
cferr2	✓	Quadratic fluctuations of c jets

Calibrations of JES and JER are done centrally at CMS [42]. Uncertainties associated with the JES calibrations are divided into 27 sources to properly account for correlations, which are summarized in Table 17.2. Uncertainties associated with the calibrations of JER are combined into a separate uncertainty, which is considered uncorrelated across the years. Variations of JES and JER due to these uncertainties are propagated to the MET and calibrations of the DeepJet scores: scale factors used to correct DeepJet scores and the MET vector are recomputed for each of the jet energy variations and treated as uncertainties that are fully correlated to the respective jet energy variation.

Table 17.2: Summary of the sources of uncertainty related to the JECs. A hyphen (−) denotes that a source is not correlated across the years.

Source	Correlated	Source	Correlated
AbsoluteStat	-	RelativePtHF	✓
AbsoluteScale	✓	RelativeBal	✓
AbsoluteMPFBias	✓	RelativeSample	-
Fragmentation	✓	RelativeFSR	-
SinglePionECAL	✓	RelativeStatFSR	✓
SinglePionHCAL	✓	RelativeStatEC	-
FlavorQCD	✓	RelativeStatHF	-
TimePtEta	-	PileUpDataMC	✓
RelativeJEREC1	-	PileUpPtRef	✓
RelativeJEREC2	-	PileUpPtBB	✓
RelativeJERHF	✓	PileUpPtEC1	✓
RelativePtBB	✓	PileUpPtEC2	✓
RelativePtEC1	-	PileUpPtHF	✓
RelativePtEC2	-		

One additional uncertainty is assigned to the unclustered MET is considered [43], and is treated as uncorrelated across the years.

In the 2016 and 2017 runs, L1 ECAL triggers fired early [44] causing many uninteresting events to be recorded while the later interesting events were rejected. Since this effect is not present in the MC simulation, a correction is applied to all signals and *prompt* backgrounds. This correction is varied by 20% and the resulting change in per-event weight is taken as the uncertainty. This uncertainty is treated as correlated across the years.

In the 2018 runs, two HCAL modules whose power supply died in the middle of the data taking. This affected the measurement of jet energy and MET. No correction is applied as this effect is not well-understood and likely not significant relative to other corrections. Nevertheless, an uncertainty, denoted by "HEM", is assigned to cover the variations of jet energy and MET caused by those two dead modules.

A comparison of different sources of systematic uncertainties of the background estimates in the SRs is shown in Figure 17.7.

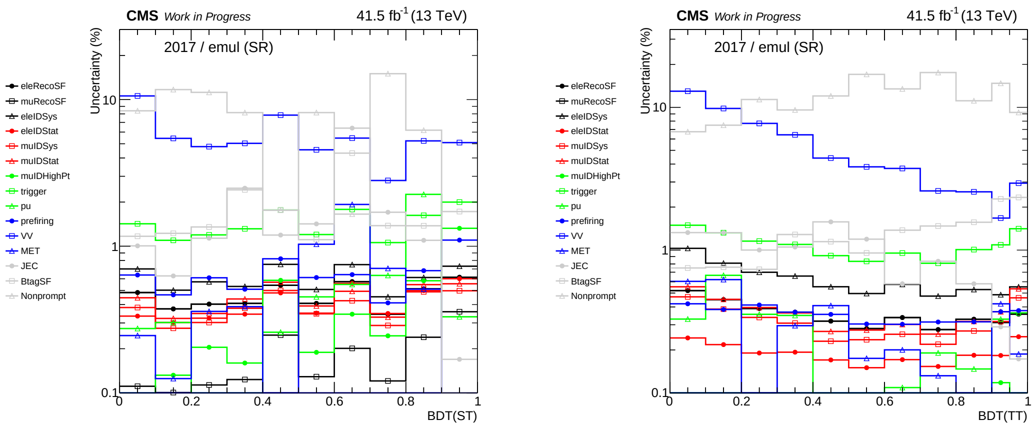
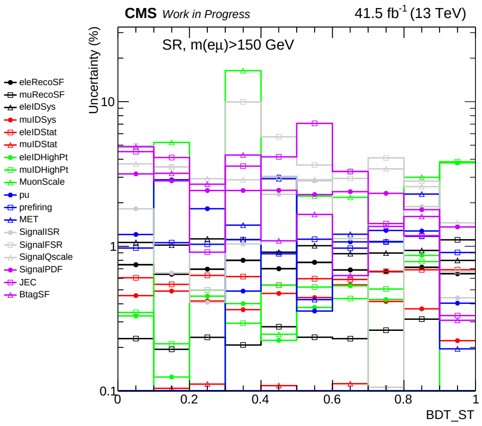


Figure 17.7: Distributions of relative uncertainties on total expected backgrounds as a function of BDT output in top production enriched SR (left), top decay enriched SR (right). Luminosity and cross-section uncertainties are not included in these plots. JES, JER, and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table 17.1 are combined into "BtagSF".

A comparison of different sources of systematic uncertainties of the signal estimates in the SRs is shown in Figure 17.8.

A representative range of systematic uncertainties on $t\bar{t}$ background and signal are summarized in Table 17.3. These uncertainties are extracted from pre-fit BDT templates shown in Figure 16.12.



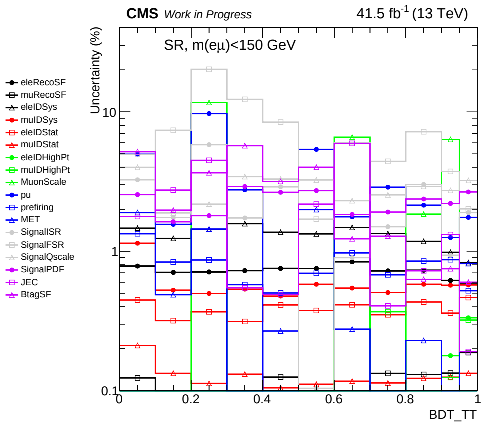


Figure 17.8: Distributions of relative uncertainties on signal ($C_{e\mu tu}^{vector}$ is used as an example) as a function of BDT output in top production enriched SR (left), top decay enriched SR (right). Luminosity and cross-section uncertainties are not included in these plots. JES, JER, and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table 17.1 are combined into "BtagSF".

Table 17.3: Representive range of systematic uncertainties extracted from 2017 dataset.

Sys. Unc.	BDT(TT)		BDT(ST)	
	$t\bar{t}W$	LFV signal	$t\bar{t}W$	LFV signal
eleRecoSF	0.5%-0.55%	0.6%-0.65%	0.59%-0.64%	0.64%-0.72%
muRecoSF	0.12%-0.16%	0.14%-0.23%	0.25%-0.35%	0.31%-1.13%
eleIDSys	0.72%-0.87%	0.82%-0.99%	0.73%-0.76%	0.79%-0.95%
eleIDStat	0.3%-0.32%	0.4%-0.49%	0.57%-0.62%	0.69%-0.73%
eleIDHighPt	0.03%-0.06%	0.17%-0.61%	0.46%-0.87%	0.89%-3.95%
muIDSys	0.68%-0.69%	0.54%-0.61%	0.37%-0.45%	0.22%-0.37%
muIDStat	0.15%-0.15%	0.12%-0.13%	0.08%-0.1%	0.05%-0.08%
muIDHighPt	0.06%-0.12%	0.09%-0.53%	0.43%-1.12%	0.84%-3.97%
MuonScale	0.09%-1.12%	0.1%-2.61%	1.04%-4.82%	0.01%-0.99%
pu	0.55%-1.89%	0.01%-0.53%	0.22%-0.72%	0.15%-0.57%
prefiring	0.65%-0.7%	0.66%-0.71%	0.86%-0.96%	0.88%-1.18%
MET	0.09%-1.18%	0.12%-0.89%	0.07%-1.9%	0.08%-0.14%
ISR	-	0.65%-1.11%	-	0.29%-0.66%
FSR	-	3.79%-4.87%	-	0.18%-2.58%
Qscale	11.16%-13.08%	2.88%-3.28%	14.82%-16.71%	1.35%-2.33%
PDF	0.07%-0.09%	2.33%-2.66%	0.18%-0.18%	1.28%-1.6%
VV	0.0%-0.0%	-	0.0%-0.0%	-
JEC	0.75%-1.74%	0.57%-1.83%	0.61%-2.28%	0.23%-1.08%
BtagSF	4.39%-5.65%	0.31%-1.2%	1.42%-2.29%	0.44%-0.83%

Chapter 18

Statistical Analysis

In the absence of significant excess over the SM prediction, the observed distributions of the BDT discriminator are used to test various hypophyses, where the coexistence of the CLFV signals and backgrounds are assumed. A statistical method called "profile likelihood" is used to move the focus on the cross sections of the CLFV signals while also keeping track of the systematic uncertainties. The profile likelihood fit performed on the distributions of the BDT discriminator is discussed in Section 18.1. Upper limits on various WCs and branching fractions established by this analysis are presented in Section 18.2.

18.1 Profile Likelihood Fit

A binned likelihood function $\mathcal{L}(\mu, \theta)$ is constructed to perform the statistical analysis using the BDT discriminator distributions. The top quark production and decay signal modes are combined. The signal strength μ, defined previously in Equation 12.14, governs the cross-section of the two signal modes simultaneously.

All systematic uncertainties are incorporated into the likelihood function as nuisance parameters, denoted by θ. The uncertainties that affect the shape of the BDT discriminator distributions utilize Gaussian distributions while other uncertainties that only affect the normalizations utilize log-normal distributions. The "Barlow-Beeston lite" method [45] is used to incorporate the statistical uncertainties in the signal and background predictions.

A profile likelihood fit is performed simultaneously in six regions (three data-taking years and two SRs) by maximizing the likelihood function $\mathcal{L}(\mu, \hat{\theta}_\mu)$, where $\hat{\theta}_\mu$ are the values of the nuisance parameters that maximize the likelihood for a specific signal strength. The post-fit distributions of the BDT discriminators are shown in Figure 18.1. The largest post-fit uncertainties are the statistical uncertainties from the limited number of simulated events.

The impacts of the nuisance parameters on the profile likelihood fit are quantified and a representative ranking of the impacts is shown in Figure 18.2. In general, the most prominent uncertainties affecting the likelihood fit are the statistical uncertainties that arise from limited sample size. A full collection of nuisance parameter impacts can be found in Appendix C.

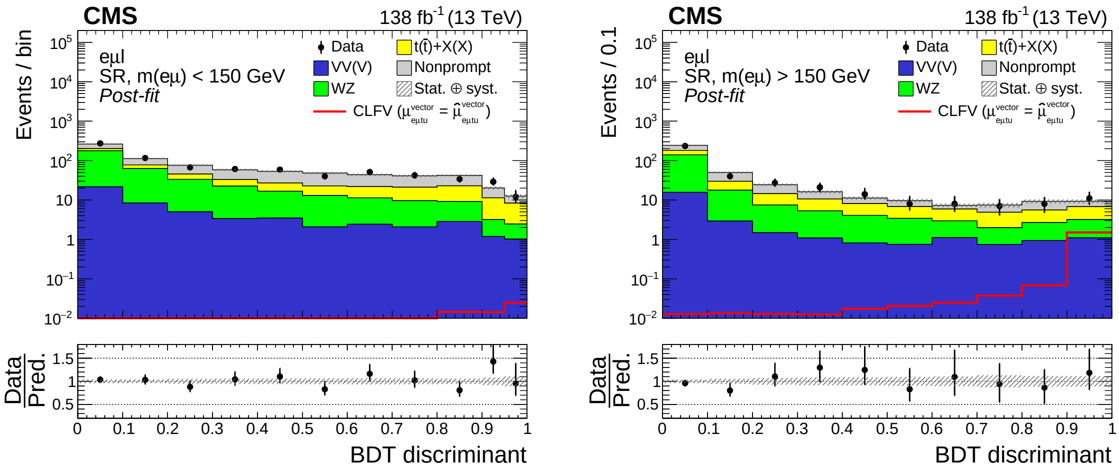


Figure 18.1: Distributions of the post-fit BDT discriminator targeting the CLFV top quark decay (left) and production (right) signal. Contributions from the two signal modes (production and decay) are combined within each SR and are shown as the solid red line. The post-fit signal strength ($\mu^{\text{vector}}_{\text{e}\mu\text{tu}} = \hat{\mu}^{\text{vector}}_{\text{e}\mu\text{tu}}$) is used to normalise the signal cross sections. The hatched bands indicate post-fit uncertainties (statistical and systematic) for the SM background predictions.

18.2 Upper Limits

The compatibility between the data and the combined signal plus background expectation under the hypothesized value of the signal strength μ is quantified by a test statistic that considers the profile likelihood ratio:

$$q(\mu) = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}. \tag{18.1}$$

Using the asymptotic modified frequentist CL_s method [46, 5, 47] with the profile likelihood ratio as the test statistic, upper limits are placed on μ at 95% CL. The one-dimensional upper limits on a given WC, C_a/Λ^2, are obtained by taking the square root of the corresponding signal strength μ_a while setting other WCs to zero. The branching fractions, $\mathcal{B}(\text{t} \to \text{e}\mu\text{q})$ with q=u or c, are obtained assuming a top quark mass (width) of 172.5 (1.33) GeV in Equation 12.13 taken from [15].

The resulting one-dimensional limits are summarized in Table 18.1. Assuming a linear relationship between $\mathcal{B}(\text{t} \to \text{e}\mu\text{u})$ and $\mathcal{B}(\text{t} \to \text{e}\mu\text{c})$ in the case of nonvanishing signals, the two-dimensional limits can be obtained through interpolation and are shown in Figure 18.3.

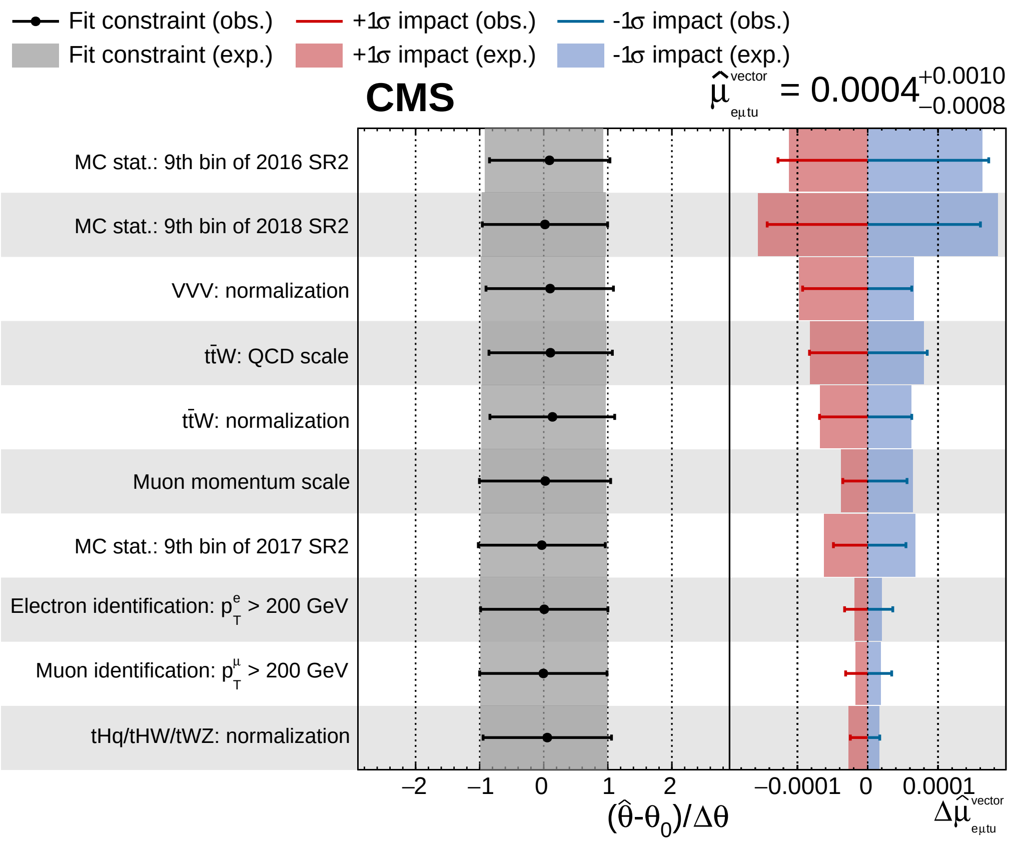
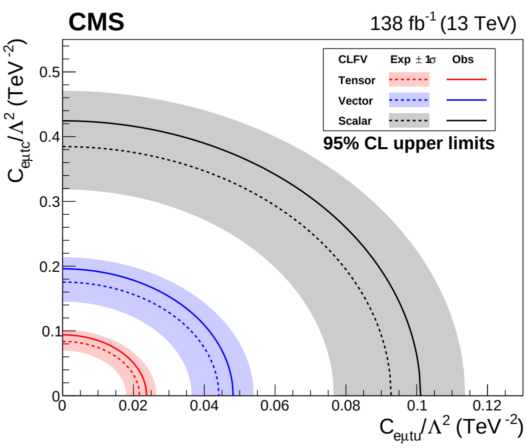


Figure 18.2: The nominal value of the observed signal strength $\hat{\mu}$ and its uncertainty is shown in the top right corner. Ranking of the nuisance parameters according to their observed impacts on $\hat{\mu}$ (represented with error bars) is shown in the right panel. Only the 10 nuisance parameters with the largest observed impacts are shown. The expected impacts (represented with red and blue rectangles) are derived using Asimov fits, where data is replaced by a background-only template (i.e. the nominal value of the expected $\hat{\mu}$ is 0). The impact of each nuisance parameter, $\Delta\hat{\mu}$, is calculated as the difference between the nominal $\hat{\mu}$ and the value of $\hat{\mu}$ when the corresponding nuisance parameter is fixed to $\hat{\theta} \pm \sigma$, where $\hat{\theta}$ (σ) is its post-fit value (uncertainty). The left panel shows the pulls (represented with black dots) and uncertainties (represented with error bars and grey rectangles) of the nuisance parameters in units of the pre-fit uncertainties. The pulls are calculated as the difference between the nominal and the post-fit values of the nuisance parameters. The “SR2" quoted in the label corresponds to the top quark production enriched signal region.

Table 18.1: Upper limits at 95% CL on WCs and the branching fractions. The expected and observed upper limits are shown in regular and bold fonts, respectively. The intervals that contain 68% of the distribution of the expected upper limits are shown in parentheses.

CLFV coupling	Lorentz structure	$C_{e\mu tq}/\Lambda^2$ (TeV^{-2}) exp (68% range)	obs	$\mathcal{B}(t \to e\mu q) \times 10^{-6}$ exp (68% range)	obs
eμtu	tensor	0.022 (0.018–0.026)	**0.024**	0.027 (0.018–0.040)	**0.032**
	vector	0.044 (0.036–0.054)	**0.048**	0.019 (0.013–0.028)	**0.022**
	scalar	0.093 (0.077–0.114)	**0.101**	0.010 (0.007–0.016)	**0.012**
eμtc	tensor	0.084 (0.069–0.102)	**0.094**	0.396 (0.272–0.585)	**0.498**
	vector	0.175 (0.145–0.214)	**0.196**	0.296 (0.203–0.440)	**0.369**
	scalar	0.385 (0.318–0.471)	**0.424**	0.178 (0.122–0.266)	**0.216**



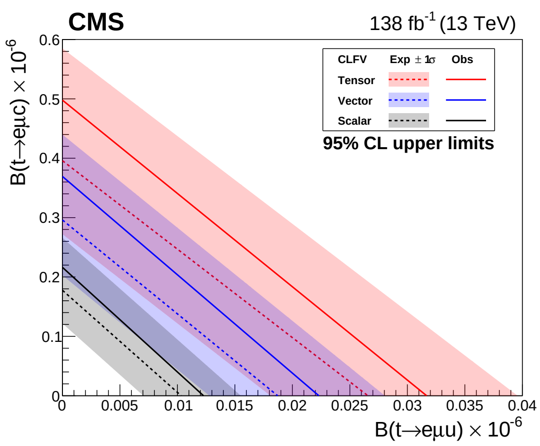


Figure 18.3: Two-dimensional 95% CL upper limits on the WCs (left) and the branching fractions (right). The observed (expected) upper limits for tensor-, vector-, and scalar-like CLFV interactions are shown in red, blue, and black solid (dotted) lines, respectively. The shaded bands contain 68% of the distribution of the expected upper limits.

Part IV

Outlook for CLFV Searches Using Top Quarks

The Part IV of this dissertation describes an ongoing physics analysis that is an extension of the previous analysis documented in Part III. Under the supervision of Prof. Louise Skinnari, Emily Minyun Tsai and I started working on this analysis in January 2023. Materials presented in the Part IV have not yet been fully approved by the CMS Collaboration and shall not be considered finalized. The Part IV is organized as follows. Chapter 19 gives a description of the signal processes targeted by this analysis and how they are generated. Object- and event-level selection criteria used by this analysis are described in Chapter 20. Chapter 21 presents results of a preliminary sensitivity study.

Chapter 19

Inclusive CLFV Signals

19.1 Targeted Signals

19.2 Signal Event Generation

Chapter 20

Object and Event Selection

20.1 Object Selection

20.2 Event Selection

Chapter 21

Expected Sensitivity

21.1 Asimov Fit

21.2 Expected Upper Limits

Chapter 22

Summary and Conclusions

References

[1] **ATLAS** Colloboration. *Search for charged lepton-flavour violation in top-quark decays at the LHC with the ATLAS detector.* **ATLAS-CONF-2018-044** (2018). URL: https://cds.cern.ch/record/2638305. (Cited on pages vii, 14, and 16.)

[2] **ATLAS** Colloboration. *Search for charged-lepton-flavour violating $\mu\tau qt$ interactions in top-quark production and decay with the ATLAS detector at the LHC.* **ATLAS-CONF-2023-001** (2023). URL: https://cds.cern.ch/record/2845451. (Cited on pages vii and 15.)

[3] **CMS** Collaboration. *Search for charged-lepton flavor violation in top quark production and decay in pp collisions at $\sqrt{s}$ = 13 TeV*. JHEP **06**, 082 (2022). arXiv:2201.07859, doi:10.1007/JHEP06(2022)082. (Cited on pages vii and 16.)

[4] **CMS** Colloboration. *Search for charged lepton flavor violation in the top quark sector in trilepton final states with the CMS detector at* $\sqrt{s} = 13$ TeV **CMS-PAS-TOP-22-005** (2023). URL: http://cds.cern.ch/record/2851002. (Cited on page 13.)

[5] Alexander Read. *Presentation of search results: The CL_s technique*. J. Phys. G **28**, 2693 (2002). doi:10.1088/0954-3899/28/10/313. (Cited on pages 15 and 65.)

[6] Sacha Davidson, Michelangelo L. Mangano, Stephane Perries, and Viola Sordini. *Lepton Flavour Violating top decays at the LHC*. Eur. Phys. J. C **75(9)**, 450 (2015). arXiv:1507.07163, doi:10.1140/epjc/s10052-015-3649-5. (Cited on page 15.)

[7] Mikael Chala, Jose Santiago, and Michael Spannowsky. *Constraining four-fermion operators using rare top decays*. JHEP **04**, 014 (2019). arXiv:1809.09624, doi:10.1007/JHEP04(2019)014. (Cited on page 15.)

[8] D. Barducci et al. *Interpreting top-quark LHC measurements in the standard-model effective field theory* (2018). arXiv:1802.07237. (Cited on page 17.)

[9] A. Dedes, M. Paraskevas, J. Rosiek, K. Suxho, and L. Trifyllis. *SmeftFR – Feynman rules generator for the Standard Model Effective Field Theory*. Comput. Phys. Commun. **247**, 106931 (2020). arXiv:1904.03204, doi:10.1016/j.cpc.2019.106931. (Cited on page 18.)

[10] Celine Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. *UFO - The Universal FeynRules Output*. Comput. Phys. Commun. **183**, 1201–

1214 (2012). arXiv:1108.2040, doi:10.1016/j.cpc.2012.01.022. (Cited on page 18.)

[11] Neil D. Christensen and Claude Duhr. *FeynRules - Feynman rules made easy*. Comput. Phys. Commun. **180**, 1614–1641 (2009). arXiv:0806.4194, doi:10.1016/j.cpc.2009.02.018. (Cited on page 18.)

[12] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. JHEP **07**, 079 (2014). arXiv:1405.0301, doi:10.1007/JHEP07(2014)079. (Cited on page 18.)

[13] Ilaria Brivio, Yun Jiang, and Michael Trott. *The SMEFTsim package, theory and tools*. JHEP **12**, 070 (2017). arXiv:1709.06492, doi:10.1007/JHEP12(2017)070. (Cited on page 18.)

[14] Michal Czakon and Alexander Mitov. *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*. Comput. Phys. Commun. **185**, 2930 (2014). arXiv:1112.5675, doi:10.1016/j.cpc.2014.06.021. (Cited on pages 19 and 22.)

[15] Jennifer Kile and Amarjit Soni. *Model-Independent Constraints on Lepton-Flavor-Violating Decays of the Top Quark*. Phys. Rev. D **78**, 094008 (2008). arXiv:0807.4199, doi:10.1103/PhysRevD.78.094008. (Cited on pages 20 and 65.)

[16] **NNPDF** Collaboration. *Parton distributions for the LHC Run II*. JHEP **04**, 040 (2015). arXiv:1410.8849, doi:10.1007/JHEP04(2015)040. (Cited on pages 21 and 53.)

[17] **NNPDF** Collaboration. *Parton distributions from high-precision collider data*. Eur. Phys. J. C **77(10)**, 663 (2017). arXiv:1706.00428, doi:10.1140/epjc/s10052-017-5199-5. (Cited on pages 21 and 53.)

[18] Stefano Frixione, Paolo Nason, and Carlo Oleari. *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*. JHEP **11**, 070 (2007). arXiv:0709.2092, doi:10.1088/1126-6708/2007/11/070. (Cited on page 21.)

[19] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. *An introduction to PYTHIA 8.2*. Comput. Phys. Commun. **191**, 159–177 (2015). arXiv:1410.3012, doi:10.1016/j.cpc.2015.01.024. (Cited on page 21.)

[20] **CMS** Colloboration. *Event generator tunes obtained from underlying event and multiparton scattering measurements*. Eur. Phys. J. C **76(3)**, 155 (2016). arXiv:1512.00815, doi:10.1140/epjc/s10052-016-3988-x. (Cited on page 21.)

[21] **CMS** Colloboration. *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*. Eur. Phys. J. C **80(1)**, 4 (2020). arXiv:1903.12179, doi:10.1140/epjc/s10052-019-7499-4. (Cited on page 21.)

[22] John M. Campbell, R. Keith Ellis, and Ciaran Williams. *Vector boson pair production at the LHC*. JHEP **07**, 018 (2011). arXiv:1105.0020, doi:10.1007/JHEP07(2011)018. (Cited on pages 22 and 53.)

[23] Rikkert Frederix and Ioannis Tsinikos. *On improving NLO merging for* $\mathrm{t\bar{t}W}$ *production*. JHEP **11**, 029 (2021). arXiv:2108.07826, doi:10.1007/JHEP11(2021)029. (Cited on pages 22 and 53.)

[24] Anna Kulesza, Leszek Motyka, Daniel Schwartländer, Tomasz Stebel, and Vincent Theeuwes. *Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy*. Eur. Phys. J. C **80(5)**, 428 (2020). arXiv:2001.03031, doi:10.1140/epjc/s10052-020-7987-6. (Cited on pages 22 and 53.)

[25] Ye Li and Frank Petriello. *Combining QCD and electroweak corrections to dilepton production in FEWZ*. Phys. Rev. D **86**, 094034 (2012). arXiv:1208.5967, doi:10.1103/PhysRevD.86.094034. (Cited on page 22.)

[26] **CMS** Colloboration. *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at* $\sqrt{s}$ *=13 TeV*. Phys. Rev. Lett. **122(13)**, 132003 (2019). arXiv:1812.05900, doi:10.1103/PhysRevLett.122.132003. (Cited on page 23.)

[27] **CMS** Colloboration. *Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*. JHEP **02**, 107 (2022). arXiv:2111.02860, doi:10.1007/JHEP02(2022)107. (Cited on page 23.)

[28] **TMVA** Colloboration. *TMVA - Toolkit for Multivariate Data Analysis* (2007). arXiv:physics/0703039. (Cited on page 23.)

[29] Matteo Cacciari and Gavin P. Salam. *Pileup subtraction using jet areas*. Phys. Lett. B **659**, 119–126 (2008). arXiv:0707.1378, doi:10.1016/j.physletb.2007.09.077. (Cited on page 24.)

[30] Emil Bols, Jan Kieseler, Mauro Verzetti, Markus Stoye, and Anna Stakia. *Jet Flavour Classification Using DeepJet*. JINST **15(12)**, P12012 (2020). arXiv:2008.10519, doi:10.1088/1748-0221/15/12/P12012. (Cited on page 26.)

[31] **CMS** Colloboration. *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*. JINST **13(05)**, P05011 (2018). arXiv:1712.07158, doi:10.1088/1748-0221/13/05/P05011. (Cited on page 26.)

[32] Thomas P. S. Gillam and Christopher G. Lester. *Improving estimates of the number of 'fake' leptons and other mis-reconstructed objects in hadron collider events: BoB's your UNCLE*. JHEP **11**, 031 (2014). arXiv:1407.5624, doi:10.1007/JHEP11(2014)031. (Cited on page 34.)

[33] Tianqi Chen and Carlos Guestrin. *XGBoost: A scalable tree boosting system*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining page 785 (2016). arXiv:1603.02754, doi:10.1145/2939672.2939785. (Cited on pages 45 and 47.)

[34] **CMS** Colloboration. *Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s}$ = 7 TeV*. JINST **7**, P10002 (2012). arXiv:1206.4071, doi:10.1088/1748-0221/7/10/P10002. (Cited on page 53.)

[35] **CMS** Colloboration. *Precision luminosity measurement in proton-proton collisions at $\sqrt{s}$ = 13 TeV in 2015 and 2016 at CMS*. Eur. Phys. J. C **81(9)**, 800 (2021). arXiv:2104.01927, doi:10.1140/epjc/s10052-021-09538-2. (Cited on page 58.)

[36] **CMS** Colloboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV*. **CMS-PAS-LUM-17-004** (2018). URL: https://cds.cern.ch/record/2621960/. (Cited on page 58.)

[37] **CMS** Colloboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV*. **CMS-PAS-LUM-18-002** (2019). URL: https://cds.cern.ch/record/2676164/. (Cited on page 58.)

[38] **CMS** Colloboration. *Measurement of the inelastic proton-proton cross section at $\sqrt{s}$ = 13 TeV*. JHEP **07**, 161 (2018). arXiv:1802.02613, doi:10.1007/JHEP07(2018)161. (Cited on page 58.)

[39] **CMS** Colloboration. *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s}$ = 7 TeV*. JINST **8**, P09009 (2013). arXiv:1306.2016, doi:10.1088/1748-0221/8/09/P09009. (Cited on page 59.)

[40] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov. *Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments*. Eur. Phys. J. C **72**, 2194 (2012). arXiv:1208.3710, doi:10.1140/epjc/s10052-012-2194-8. (Cited on page 59.)

[41] **CMS** Colloboration. *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV*. JINST **13(06)**, P06015 (2018). arXiv:1804.04528, doi:10.1088/1748-0221/13/06/P06015. (Cited on page 59.)

[42] **CMS** Colloboration. *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*. JINST **12(02)**, P02014 (2017). arXiv:1607.03663, doi:10.1088/1748-0221/12/02/P02014. (Cited on page 59.)

[43] **CMS** Colloboration. *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*. JINST **14(07)**, P07004 (2019). arXiv:1903.06078, doi:10.1088/1748-0221/14/07/P07004. (Cited on page 61.)

[44] **CMS** Colloboration. *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*. JINST **15(10)**, P10017 (2020). arXiv:2006.10165, doi:10.1088/1748-0221/15/10/P10017. (Cited on page 61.)

[45] Roger J. Barlow and Christine Beeston. *Fitting using finite Monte Carlo samples*. Comput. Phys. Commun. **77**, 219 (1993). doi:10.1016/0010-4655(93)90005-W. (Cited on page 64.)

[46] Thomas Junk. *Confidence level computation for combining searches with small statistics*. Nucl. Instrum. Meth. A **434**, 435 (1999). arXiv:hep-ex/9902006, doi:10.1016/S0168-9002(99)00498-2. (Cited on page 65.)

[47] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. *Asymptotic formulae for likelihood-based tests of new physics*. Eur. Phys. J. C **71**, 1554 (2011). arXiv:1007.1727, doi:10.1140/epjc/s10052-011-1554-0. (Cited on page 65.)

Appendix A

List of Trigger Paths

Table A.1: Triggers used to record events during data taking in 2016.

Dataset	Trigger Path
SingleMuon	HLT_IsoMu22_eta2p1, HLT_IsoTkMu22_eta2p1 HLT_IsoMu24, HLT_IsoTkMu24 HLT_Mu50, HLT_TkMu50, HLT_Mu45_eta2p1
SingleElectron	HLT_Ele25_eta2p1_WPTight_Gsf HLT_Ele27_WPTight_Gsf HLT_Ele27_eta2p1_WPTight_Gsf HLT_Ele32_eta2p1_WPTight_Gsf HLT_Ele105_CaloIdVT_GsfTrkIdT HLT_Ele115_CaloIdVT_GsfTrkIdT
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ HLT_Mu30_TkMu11, HLT_TripleMu_12_10_5
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL HLT_DoubleEle33_CaloIdL_MW HLT_DoubleEle33_CaloIdL_GsfTrkIdVL HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_MW
MuonEG	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL HLT_Mu30_Ele30_CaloIdL_GsfTrkIdVL HLT_Mu33_Ele33_CaloIdL_GsfTrkIdVL HLT_DiMu9_Ele9_CaloIdL_TrackIdL HLT_Mu8_DiEle12_CaloIdL_TrackIdL

Table A.2: Triggers used to record events during data taking in 2017.

Dataset	Trigger Path
SingleMuon	HLT_IsoMu24_eta2p1
	HLT_IsoMu27
	HLT_Mu50
	HLT_OldMu100
	HLT_TkMu100
SingleElectron	HLT_Ele32_WPTight_Gsf_L1DoubleEG
	HLT_Ele35_WPTight_Gsf
	HLT_Ele115_CaloIdVT_GsfTrkIdT
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
	HLT_Mu19_TrkIsoVVL_Mu9_TrkIsoVVL_DZ_Mass3p8
	HLT_Mu37_TkMu27
	HLT_TripleMu_12_10_5
	HLT_TripleMu_10_5_5_DZ
	HLT_TripleMu_5_3_3_Mass3p8to60_DZ
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
	HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL
	HLT_DoubleEle25_CaloIdL_MW
	HLT_DoubleEle33_CaloIdL_MW
	HLT_DiEle27_WPTightCaloOnly_L1DoubleEG
MuonEG	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu27_Ele37_CaloIdL_MV
	HLT_Mu37_Ele27_CaloIdL_MV
	HLT_DiMu9_Ele9_CaloIdL_TrackIdL
	HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ
	HLT_Mu8_DiEle12_CaloIdL_TrackIdL

Table A.3: Triggers used to record events during data taking in 2018.

Dataset	Trigger Path
SingleMuon	HLT_IsoMu24
	HLT_IsoMu27
	HLT_Mu50
	HLT_OldMu100
	HLT_TkMu100
EGamma	HLT_Ele32_WPTight_Gsf
	HLT_Ele115_CaloIdVT_GsfTrkIdT
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
	HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL
	HLT_DoubleEle25_CaloIdL_MW
	HLT_DiEle27_WPTightCaloOnly_L1DoubleEG
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
	HLT_Mu37_TkMu27
	HLT_TripleMu_12_10_5
	HLT_TripleMu_10_5_5_DZ
	HLT_TripleMu_5_3_3_Mass3p8to60_DZ
MuonEG	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_Mu27_Ele37_CaloIdL_MV
	HLT_Mu37_Ele27_CaloIdL_MV
	HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ
	HLT_Mu8_DiEle12_CaloIdL_TrackIdL

Appendix B

Signal Region Distributions with MC Simulation

Distributions of various variables in SR, which are used in the BDT training. More details on these input features are described in Section 16.2. The data are shown as filled points and the SM background predictions as histograms. The VV(V) background includes ZZ and triboson production, while the $t\bar{t}+X(X)$ component includes $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$, tZq, and smaller backgrounds containing one or two top quarks plus a boson or quark. All backgrounds are estimated using MC simulation. The hatched bands indicate statistical and systematic uncertainties for the SM background predictions. The normalisation of the signal processes is chosen arbitrarily for improved visualisation.

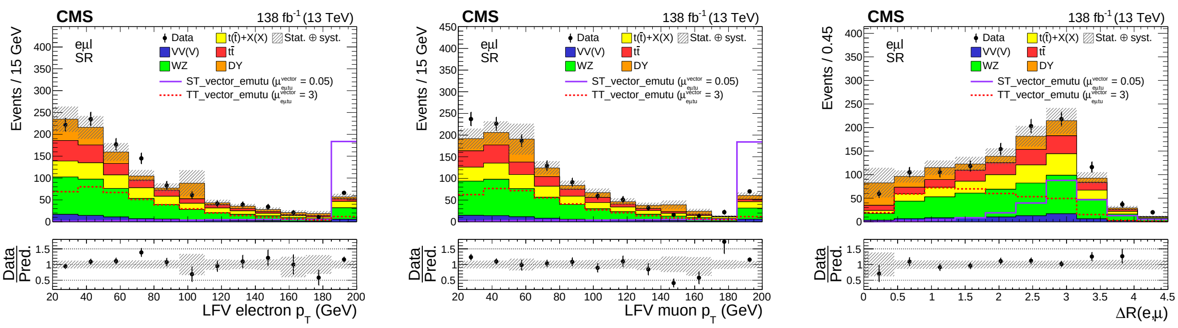


Figure B.1: Distributions of LFV electron p_T (left), LFV muon p_T (middle), and the opening angle between LFV electron and LFV muon (right).

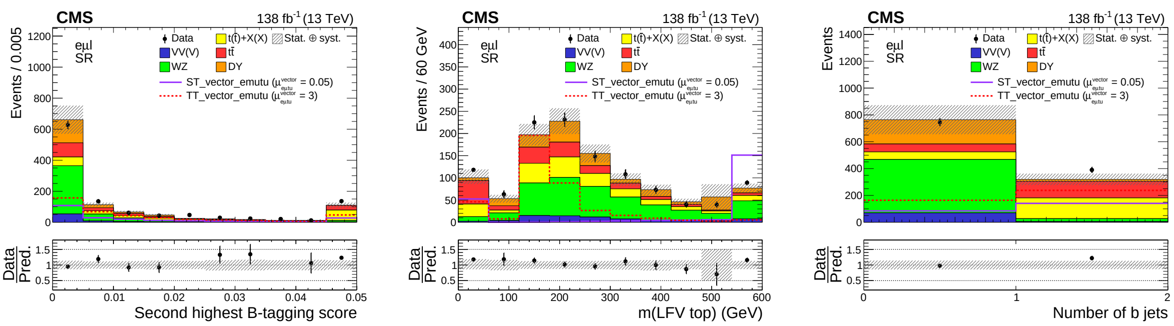


Figure B.2: Distributions of the second highest DeepJet score (left), LFV top mass (middle), b jet multiplicity (right).

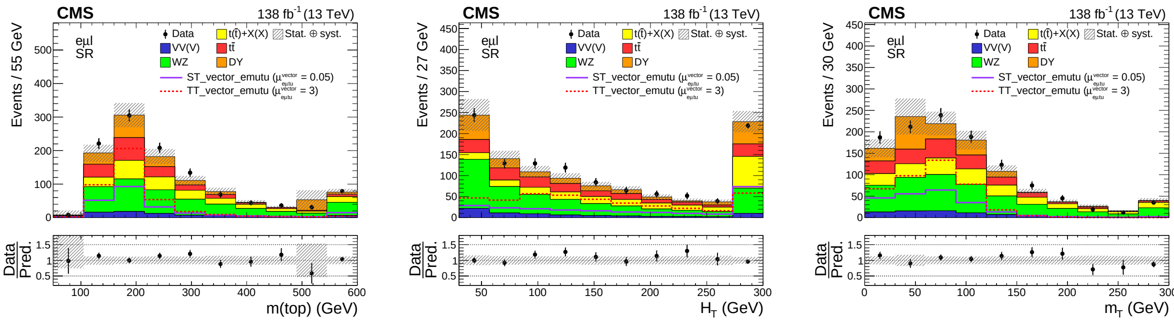


Figure B.3: Distributions of SM top quark mass (left), scalar sum of p_T of all jets (middle), and transverse mass of the W boson (right).

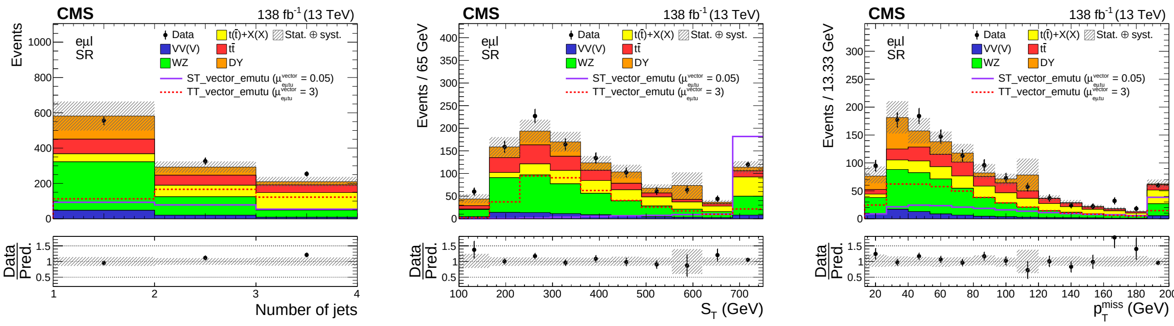


Figure B.4: Distributions of jet multiplicity (left), scalar sum of p_T of all jets and leptons (middle), and MET (right).

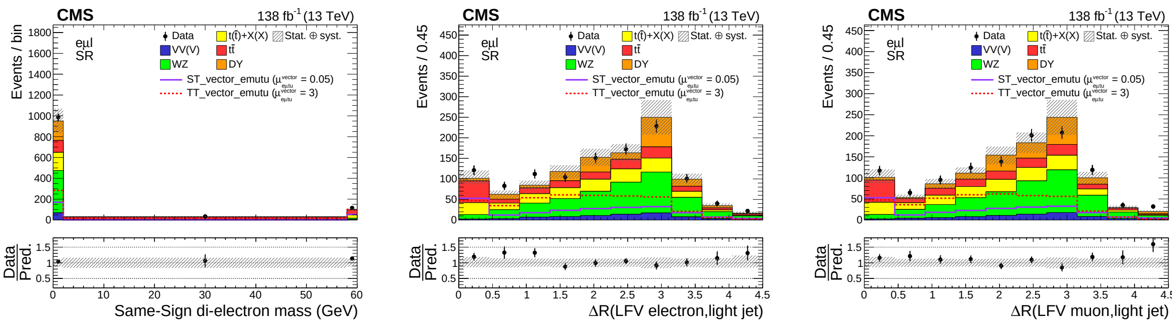


Figure B.5: Distributions of the same-sign di-electron mass (left), the opening angle between LFV electron and a light flavor jet (middle), and the opening angle between LFV muon and a light flavor jet (right).

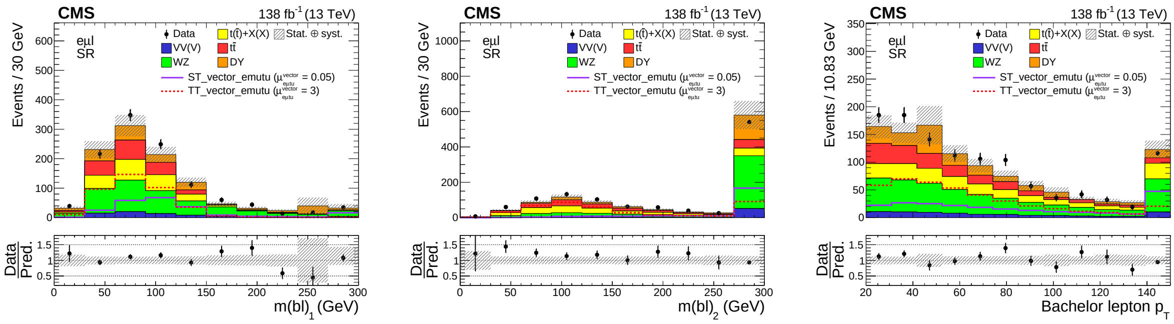


Figure B.6: Distributions of the mass of the first $m_{b\ell}$ system (left), the mass of the second $m_{b\ell}$ system (middle), and standalone lepton p_T (right).

Appendix C

Nuisance Parameter Impact

The observed and expected (μ_{exp} =0) impacts of the nuisance parameters on the profile likelihood fit are shown in Figure C.1-C.2. They were computed using the following commands and plotted below for full run 2.

The expected (μ_{exp} =1) impacts of the nuisance parameters on the profile likelihood fit are shown in Figure C.3.

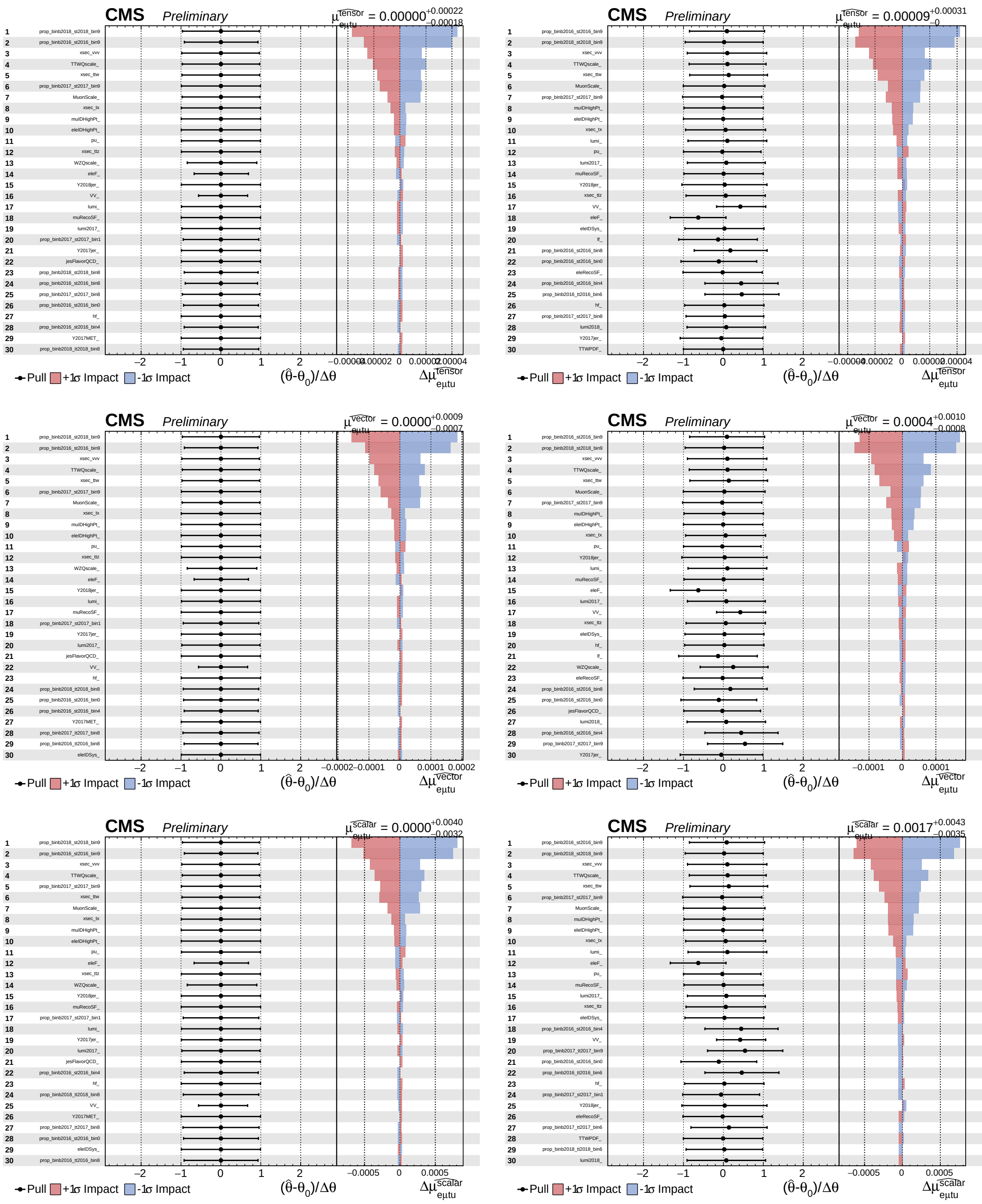


Figure C.1: Impacts of nuissance parameters for run II limit setting. From top to bottom: $e\mu tu$-tensor, $e\mu tu$-vector, $e\mu tu$-scalar. From left to right: expected impact (expected signal strength at 0), observed impact.

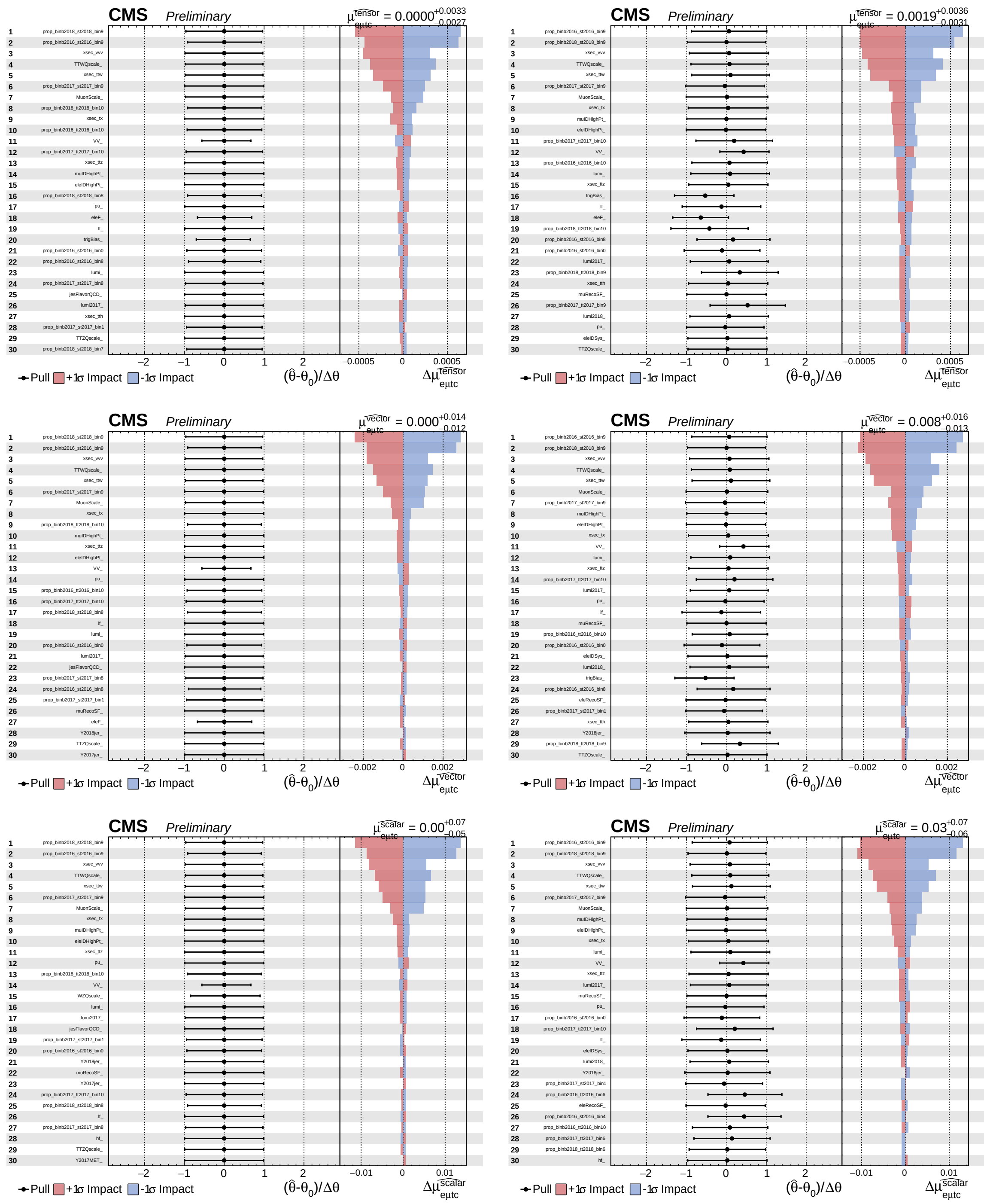


Figure C.2: Impacts of nuissance parameters for run II limit setting. From top to bottom: $e\mu tc$-tensor, $e\mu tc$-vector, $e\mu tc$-scalar. From left to right: expected impact (expected signal strength at 0), observed impact.

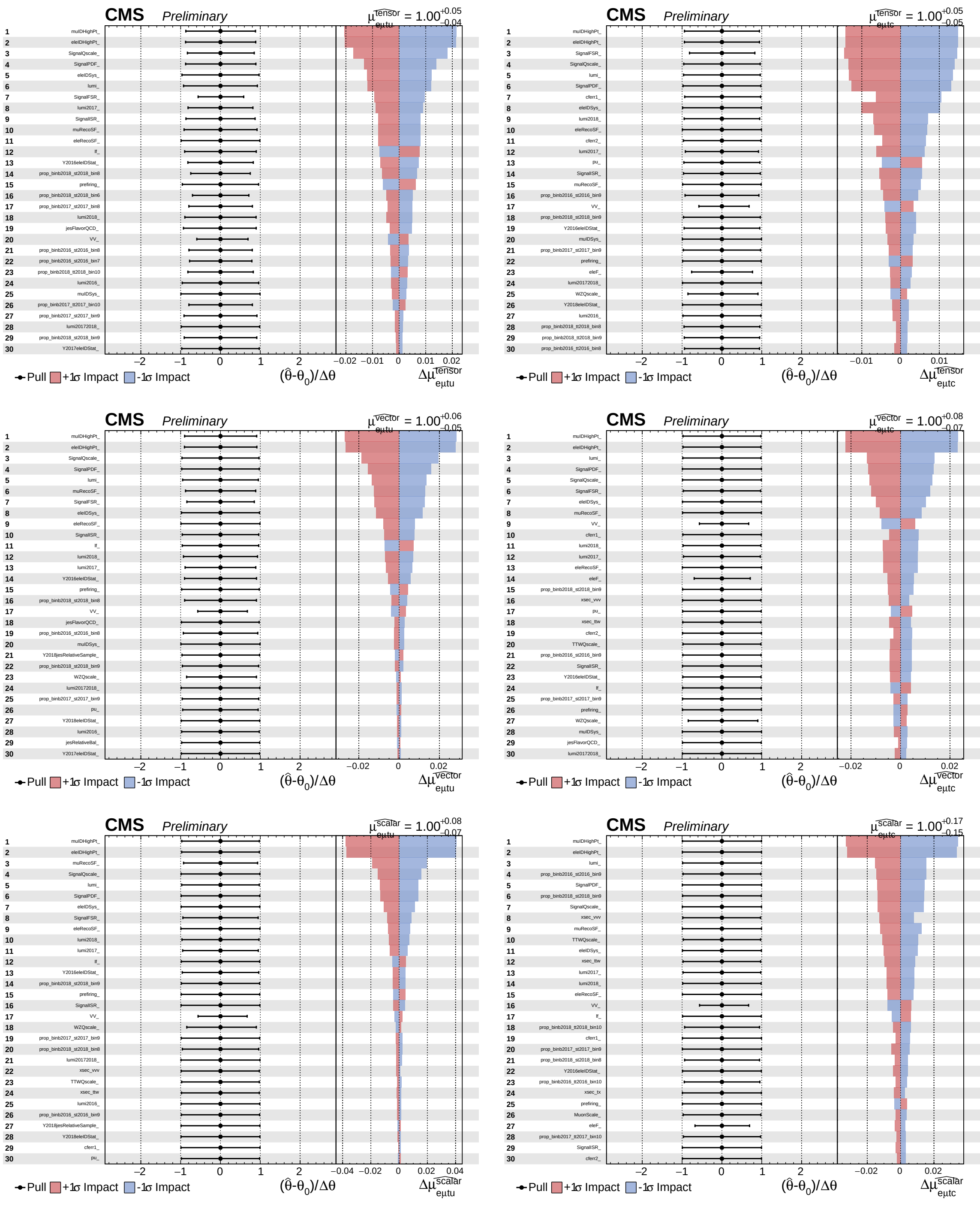


Figure C.3: Expected impact with an expected signal strength at 1. From top to bottom: tensor, vector, scalar. From left to right: $e\mu tu$, $e\mu tc$.