

**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**

© 2023 Jingyan Li  
ALL RIGHTS RESERVED

## 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  $138 \text{ fb}^{-1}$ . 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,  $t \rightarrow e\mu\mu$  ( $t \rightarrow e\mu c$ ), of  $0.032 \times 10^{-6}$  ( $0.498 \times 10^{-6}$ ),  $0.022 \times 10^{-6}$  ( $0.369 \times 10^{-6}$ ), and  $0.012 \times 10^{-6}$  ( $0.216 \times 10^{-6}$ ) for tensor, vector, and scalar interactions, respectively.

# Contents

|  |           |
|--|-----------|
| List of Figures . . . . .                                | vii       |
| List of Tables . . . . .                                 | xi        |
| List of Acronyms . . . . .                               | xii       |
| <b>1 Introduction</b>                                    | <b>1</b>  |
| <b>I Theoretical Framework</b>                           | <b>2</b>  |
| <b>2 Electroweak Standard Model</b>                      | <b>3</b>  |
| 2.1 Gauge Theory . . . . .                               | 3         |
| 2.2 Higgs Mechanism . . . . .                            | 3         |
| 2.3 Flavor Sector . . . . .                              | 3         |
| <b>3 Beyond the Standard Model</b>                       | <b>4</b>  |
| 3.1 Overview of the BSM Theories . . . . .               | 4         |
| 3.2 Leptoquark Model . . . . .                           | 4         |
| <b>4 Effective Field Theory</b>                          | <b>5</b>  |
| 4.1 Standard Model EFT . . . . .                         | 5         |
| 4.2 Low Energy EFT . . . . .                             | 5         |
| <b>5 Event Simulation for Colliders</b>                  | <b>6</b>  |
| 5.1 Perturbative Quantum Chromodynamics . . . . .        | 6         |
| 5.2 Nonperturbative Quantum Chromodynamics . . . . .     | 6         |
| <b>II Experimental Apparatus</b>                         | <b>7</b>  |
| <b>6 The Large Hadron Collider</b>                       | <b>8</b>  |
| <b>7 The Compact Muon Solenoid Detector</b>              | <b>9</b>  |
| 7.1 Coordinate System Used in the CMS Detector . . . . . | 9         |
| 7.2 The Inner Tracking System . . . . .                  | 9         |
| 7.3 The Electromagnetic Calorimeter . . . . .            | 9         |
| 7.4 The Hadronic Calorimeter . . . . .                   | 9         |
| 7.5 The Superconducting Magnet . . . . .                 | 9         |
| 7.6 The Muon System . . . . .                            | 9         |
| 7.7 The Trigger System . . . . .                         | 9         |
| <b>8 Object Reconstruction in the CMS detector</b>       | <b>10</b> |
| 8.1 Electron and Photon . . . . .                        | 10        |
| 8.2 Muon . . . . .                                       | 10        |
| 8.3 Tau . . . . .  | 10        |

|            |   |           |
|------------|---|-----------|
| 8.4        | Jets . . . . .  | 10        |
| 8.5        | Energy Sums . . . . .   | 10        |
| <b>9</b>   | <b>Run-3 Operation of the CMS detector</b>                              | <b>11</b> |
| 9.1        | The CMS Run Organisation . . . . .                                      | 11        |
| 9.2        | Central Shifting Crew . . . . .   | 11        |
| 9.3        | Detector On-call Expert . . . . .                                       | 11        |
| <b>10</b>  | <b>The Phase-2 Upgrade of the CMS Detector</b>                          | <b>12</b> |
| 10.1       | Overview of the Upgrade . . . . .                                       | 12        |
| 10.2       | Leve-1 Track Finder . . . . .   | 12        |
| 10.3       | Leve-1 Electron Trigger Algorithm . . . . .                             | 12        |
| <b>III</b> | <b>Search for Flavor-Violating <math>e\mu\tau q</math> Interactions</b> | <b>13</b> |
| <b>11</b>  | <b>Previous searches</b>  | <b>14</b> |
| 11.1       | ATLAS . . . . .   | 14        |
| 11.2       | CMS . . . . .   | 14        |
| <b>12</b>  | <b>Datasets, Simulated Samples and Triggers</b>                         | <b>15</b> |
| 12.1       | Signal Samples . . . . .  | 15        |
| 12.2       | Background Samples . . . . .  | 16        |
| 12.3       | Triggers . . . . .  | 17        |
| <b>13</b>  | <b>Object Selection</b>   | <b>19</b> |
| 13.1       | Leptons . . . . .   | 19        |
| 13.2       | Jets . . . . .  | 20        |
| 13.3       | B-tagging . . . . .   | 20        |
| <b>14</b>  | <b>Event Selection</b>  | <b>21</b> |
| 14.1       | Signal Region . . . . .   | 21        |
| 14.2       | Validation Region . . . . .   | 22        |
| 14.3       | Kinematic Reconstruction . . . . .                                      | 24        |
| <b>15</b>  | <b>Nonprompt Background Estimate</b>                                    | <b>28</b> |
| 15.1       | The Matrix Method . . . . .   | 28        |
| 15.2       | Extraction of Prompt and Nonprompt Efficiencies . . . . .               | 30        |
| 15.3       | Validation of the Matrix Method . . . . .                               | 38        |
| 15.4       | Nonprompt estimate in SR . . . . .                                      | 38        |
| <b>16</b>  | <b>Signal Extraction with Boosted Decision Trees</b>                    | <b>42</b> |
| 16.1       | BDT Setup . . . . .   | 42        |
| 16.2       | Input Variables . . . . .   | 43        |
| 16.3       | BDT Output . . . . .  | 49        |
| <b>17</b>  | <b>Systematic Uncertainties</b>   | <b>51</b> |
| 17.1       | Theoretical Uncertainties . . . . .                                     | 51        |
| 17.2       | Nonprompt Uncertainties . . . . .                                       | 52        |

---

|  |           |
|--|-----------|
| 17.3 Diboson Uncertainties . . . . .                   | 54        |
| 17.4 Other Experimental Uncertainties . . . . .        | 56        |
| <b>18 Statistical Analysis</b>                         | <b>62</b> |
| 18.1 Profile Likelihood Fit . . . . .                  | 62        |
| 18.2 Upper Limits . . . . .                            | 62        |
| <b>IV Outlook for CLFV Searches Using Top Quarks</b>   | <b>67</b> |
| <b>19 Inclusive Signal Generation</b>                  | <b>68</b> |
| 19.1 UFO Model . . . . .                               | 68        |
| 19.2 EFT Reweighting . . . . .                         | 68        |
| <b>20 Object and Event Selection</b>                   | <b>69</b> |
| 20.1 Object Selection . . . . .                        | 69        |
| 20.2 Event Selection . . . . .                         | 69        |
| <b>21 Fake Tau Estimate</b>                            | <b>70</b> |
| 21.1 ABCD Method . . . . .                             | 70        |
| 21.2 Validation . . . . .                              | 70        |
| <b>22 Expected Sensitivity</b>                         | <b>71</b> |
| 22.1 Asimov Fit . . . . .                              | 71        |
| 22.2 Expected Upper Limits . . . . .                   | 71        |
| <b>23 Run-2 Combination</b>                            | <b>72</b> |
| 23.1 Correlation Scheme . . . . .                      | 72        |
| 23.2 Minimal Flavor Violation Interpretation . . . . . | 72        |
| <b>24 Conclusion</b>                                   | <b>73</b> |
| <b>References</b>                                      | <b>74</b> |
| <b>A List of Trigger Paths</b>                         | <b>75</b> |
| <b>B Nuisance Parameter Impact</b>                     | <b>78</b> |
| <b>C More materials</b>                                | <b>82</b> |

## List of Figures

|      |  |    |
|------|--|----|
| 12.1 | Comparison of kinematic distributions produced by different UFOS: LFV electron $p_T$ (left), LFV muon $p_T$ (right). The "Reweight" (shown in blue curve) are produced by applying weights produced by citereweight to EFT samples produced by SMEFTfr. .  | 16 |
| 14.1 | Distributions of different kinematic variables estimated in SR (full run II). Backgrounds are estimated using MC only. From left to right: invariant mass of the $e\mu$ pair, leading lepton $p_T$ . . . . .   | 22 |
| 14.2 | Illustration of event selections applied to different regions . . . . .  | 23 |
| 14.3 | Distributions of different kinematic variables estimated in WZ control region (full run II). Backgrounds are estimated using MC only. From left to right: leading lepton $\eta$ , jet multiplicity. . . . .  | 24 |
| 14.4 | Distributions of different kinematic variables estimated in WZ control region (full run II). Backgrounds are estimated using MC only. From left to right: leading lepton $\eta$ , jet multiplicity. . . . .  | 24 |
| 14.5 | Distributions of different kinematic variables estimated in WZ control region (full run II). Backgrounds are estimated using MC only. From left to right: leading lepton $\eta$ , jet multiplicity. . . . .  | 25 |
| 15.1 | Illustration of the <i>real</i> efficiency $r$ and the <i>fake</i> efficiency $f$ . . . . .  | 29 |
| 15.2 | Distribution of lepton $p_T$ in an electron <i>fake</i> efficiency MR (2017). In this particular example: both $e+e$ (Same-Sign-Same-Flavor) and $e+\mu$ (Same-Sign-Opposite-Flavor) pair are considered. At least one jet and at most one b-tag jet are required (the second jet multiplicity bin). <i>Probe</i> electron is required to have $1.6 <  \eta  < 2.4$ (the third $\eta$ bin). Contamination from prompt backgrounds are estimated with MC simulation. From left to right: <i>tight+not tight</i> electron, only <i>tight</i> electron. . . . . | 35 |
| 15.3 | Fake efficiency measured in 2017 data. 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. From left to right: electron $f$ , muon $f$ . . . . .  | 36 |
| 15.4 | Comparison of the muon <i>fake</i> efficiencies measured with isolation requirement in <i>loose</i> definition (2017 datasets). From left to right: $ \eta  < 0.8$ , $0.8 <  \eta  < 1.6$ , $1.6 <  \eta  < 2.4$ . . . . .   | 37 |
| 15.5 | Comparison of the muon <i>real</i> efficiencies measured with different <i>loose</i> definition (2017 datasets). From left to right: $ \eta  < 0.8$ , $0.8 <  \eta  < 1.6$ , $1.6 <  \eta  < 2.4$ . . . . .  | 37 |
| 15.6 | <i>Real</i> efficiency measured in simulated $t\bar{t}$ events (2017). 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. From left to right: electron $r$ , muon $r$ . . . . .  | 37 |

|  |    |
|--|----|
| 15.7 Distributions of different kinematic variables estimated in VR with three electrons.<br>From left to right: leading lepton $\eta$ , jet multiplicity. . . . .                         | 38 |
| 15.8 Distributions of different kinematic variables estimated in VR with electron, muon<br>and a third light lepton. From left to right: leading lepton $\eta$ , jet multiplicity. . . . . | 39 |
| 15.9 Distributions of different kinematic variables estimated in VR with three muons. From<br>left to right: leading lepton $\eta$ , jet multiplicity. . . . .                             | 39 |
| 15.10 Distributions of different kinematic variables estimated in SR (full run II). From left<br>to right: b-jet multiplicity, missing transverse energy. . . . .                          | 40 |
| 15.11 Distributions of different kinematic variables estimated in SR (full run II). From left<br>to right: b-jet multiplicity, missing transverse energy. . . . .                          | 40 |
| 15.12 Distributions of different kinematic variables estimated in SR (full run II). From left to<br>right: mass of the flavor-violating-top-quark candidate, mass of the Z boson candidate | 41 |
| 16.1 Normalized distribution in SR1. From left to right: LFV top mass, BDT shape . . .   | 42 |
| 16.2 ROC curve with 5-fold cross validation. From left to right: BDT targeting TT (SR1),<br>BDT targeting ST (SR2). . . . .  | 43 |
| 16.3 Normalized distribution of various features in SR. From to left to right: MVA_ $M_{e\mu}$ ,<br>MVA_ $JeDr$ , MVA_ $JeDr$ . . . . .  | 43 |
| 16.4 Normalized distribution of additional features in SR. From to left to right: MVA_ $LFVePt$ ,<br>MVA_ $LFVmuPt$ , MVA_ $LFVTopmass$ . . . . .  | 44 |
| 16.5 Normalized distribution of additional features in SR. From to left to right: MVA_ $Zmass$ ,<br>MVA_ $Mbl2$ , MVA_ $IIDr$ . . . . .  | 45 |
| 16.6 Normalized distribution of additional features in SR. From to left to right: MVA_ $tM$ ,<br>MVA_ $Ht$ , MVA_ $njet$ . . . . .   | 45 |
| 16.7 Normalized distribution of additional features in SR. From to left to right: MVA_ $Jet2Btag$ ,<br>MVA_ $Mbl1$ , MVA_ $nbjet$ . . . . .  | 45 |
| 16.8 List of features ranked by their relative importance. From left to right: BDT targeting<br>TT (SR1), BDT targeting ST (SR2) . . . . .   | 46 |
| 16.9 Correlation matrices (SR1), from left to right: background correlation, signal correlation.   | 46 |
| 16.10 Correlation matrices (SR2), from left to right: background correlation, signal correlation.  | 47 |
| 16.11 BDT input validation, from left to right: MVA_ $M_{e\mu}$ , MVA_ $LFVePt$ , MVA_ $LFVmuPt$   | 48 |
| 16.12 BDT input validation, from left to right: MVA_ $Zmass$ , MVA_ $LFVTopmass$ , MVA_ $nbjet$  | 48 |
| 16.13 BDT input validation, from left to right: MVA_ $Topmass$ , MVA_ $JetHt$ , MVA_ $tM$ .  | 48 |
| 16.14 BDT input validation, from left to right: MVA_ $njet$ , MVA_ $Ht$ , MVA_ $Met$ . . . . .   | 49 |
| 16.15 BDT input validation, from left to right: MVA_ $SSZmass$ , MVA_ $JeDr$ , MVA_ $JmuDr$  | 49 |
| 16.16 BDT input validation, from left to right: MVA_ $Mbl1$ , MVA_ $Mbl2$ , MVA_ $BaPt$ . .  | 49 |
| 16.17 BDT input validation, from left to right: MVA_ $Jet2Btag$ , MVA_ $IIDr$ . . . . .  | 50 |
| 16.18 Distribution of BDT output, from left to right TT enriched SR (SR1), ST enriched<br>SR (SR2) . . . . .   | 50 |



|  |    |
|--|----|
| 17.1 Comparison of different components of the uncertainties associated to the <i>fake</i> efficiency measured in 2017 dataset ( $n_{\text{jet}}=0$ bin, $ \eta  < 0.8$ bin). From left to right: electron $f$ uncertainty, muon $f$ uncertainty. Note: black uncertainty bars represent statistical uncertainties. . . . .  | 52 |
| 17.2 Comparison of different components of the uncertainties associated to the <i>fake</i> efficiency measured in 2017 dataset ( $n_{\text{jet}} > 0$ bin, $ \eta  < 0.8$ bin). From left to right: electron $f$ uncertainty, muon $f$ uncertainty. Note: black uncertainty bars represent statistical uncertainties. . . . .  | 53 |
| 17.3 Comparison of different components of the uncertainties associated to the <i>real</i> efficiency measured in 2017 dataset ( $n_{\text{jet}}=0$ bin, $ \eta  < 0.8$ bin). From left to right: electron $r$ uncertainty, muon $r$ uncertainty. Note: black uncertainty bars represent statistical uncertainties. . . . .  | 54 |
| 17.4 The impact of matching leptons to trigger objects on non-prompt estimate. From left to right: non-prompt estimate in ST enriched SR, non-prompt estimate in TT enriched SR. The nominal configuration of the matrix method is to match the leading lepton with trigger objects. Matching the 2nd-leading with the trigger objects is taken as an alternative to evaluate the robustness of the non-prompt estimate. The uncertainty band only covers the variation of the non-prompt estimate as a result of varying leading lepton $f$ by 50 %. Uncertainty bars only include statistical uncertainties. . . . . | 55 |
| 17.5 Diboson control region, from left to right: 2016, 2017 and 2018 datasets. . . . .   | 56 |
| 17.6 Scale factors derived from diboson control region, from left to right: 2016, 2017 and 2018 datasets. . . . .  | 56 |
| 17.7 Distributions of relative uncertainties on total expected backgrounds as a function of BDT output in ST enriched SR (left), TT enriched SR (right). Luminosity and normalization uncertainties are not included in these plots. JES, JER and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table ?? are combined into "BtagSF". From top to bottom: 2016, 2017 and 2018 datasets. . . .   | 60 |
| 17.8 Distributions of relative uncertainties on signal ( $C_{e\mu\tau}^{\text{vector}}$ is used as an example) as a function of BDT output in ST enriched SR (left), TT enriched SR (right). Luminosity and normalization uncertainties are not included in these plots. JES, JER and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table ?? are combined into "BtagSF". From top to bottom: 2016, 2017 and 2018 datasets. . . .   | 61 |
| 18.1 Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with vector-like operator involving an Up quark. The three data-taking years are aggregated for better visualization. . . . .  | 63 |
| 18.2 Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with vector-like operator involving a Charm quark. . . . .   | 63 |

|      |   |    |
|------|---|----|
| 18.3 | Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with tensor-like operator involving an Up quark. . . . .                         | 64 |
| 18.4 | Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with tensor-like operator involving a Charm quark. . . . .                       | 64 |
| 18.5 | Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with scalar-like operator involving an Up quark. . . . .                         | 65 |
| 18.6 | Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with scalar-like operator involving a Charm quark. . . . .                       | 65 |
| 18.7 | Two-dimensional upper limits on the Wilson coefficients (left) and the branching ratios (right). . . . .  | 66 |
| B.1  | Impacts of nuisance 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. . . . . | 79 |
| B.2  | Impacts of nuisance 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. . . . . | 80 |
| B.3  | Expected impact with an expected signal strength at 1. From top to bottom: tensor, vector, scalar. From left to right: $e\mu tu$ , $e\mu tc$ . . . . .  | 81 |

## List of Tables

|      |   |    |
|------|---|----|
| 12.1 | Theoretical cross sections for top production and decay for each CLFV coupling. Uncertainties related to PDF and QCD scale are given ( $\sigma_{-scale}^{+scale} \pm \text{PDF}$ ). . . . .   | 17 |
| 12.2 | Summary of information related to Monte-Carlo (MC) sample generation. . . . .   | 18 |
| 14.1 | Summary of the cuts applied to different regions . . . . .  | 23 |
| 15.1 | Parent datasets (2016) used to generate TOP PAG NanoAOD-v6 samples used for $r$ and $f$ measurement. (X = "SingleElectron" and "SingleMuon".) . . . . .   | 31 |
| 15.2 | Parent datasets (2017) used to generate TOP PAG NanoAOD-v6 samples used for $r$ and $f$ measurement. (X = "SingleElectron" and "SingleMuon".) . . . . .   | 32 |
| 15.3 | Parent datasets (2018) used to generate TOP PAG NanoAOD-v6 samples used for $r$ and $f$ measurement. (X = "Egamma" and "SingleMuon".) . . . . .   | 33 |
| 15.4 | Summary of the triggers used in the measurement of $r$ and $f$ . . . . .  | 34 |
| 15.5 | Summary of the lepton selections needed for the measurement of $r$ and $f$ . Please note: a) the minimum $p_T$ cut for <i>tag</i> electron in 2016 dataset is reduced to 30 GeV to adjust for the trigger threshold; b) the <i>tight</i> selection here is the same as the lepton selection listed in autorefsec:reconstruction.1. . . . .  | 35 |
| 15.6 | Summary of the cuts applied to the $r/f$ measurement region. . . . .  | 38 |
| 15.7 | 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 $C_{e\mu\tau\nu}^{\text{vector}}/\Lambda^2 = 1\text{TeV}^{-2}$ , is also listed for reference. The signal yields include contributions from both top production and decay modes. . . . . | 41 |
| 17.1 | A hyphen (–) denotes that a source is not correlated between the different years. . . . .   | 57 |
| 17.2 | Summary of the sources of uncertainty related to the JECs. A hyphen (–) denotes that a source is not correlated between the different years. . . . .  | 59 |
| 18.1 | Upper limit on the LFV signal using the full Run 2 data set. . . . .  | 66 |
| A.1  | Triggers used to record events during data taking in 2016. . . . .  | 75 |
| A.2  | Triggers used to record events during data taking in 2017. . . . .  | 76 |
| A.3  | Triggers used to record events during data taking in 2018. . . . .  | 77 |

## List of Acronyms

**ATLAS** A Toroidal LHC Apparatus

**CMS** Compact Muon Solenoid

**EFT** Effective Field Theory

**HLT** High Level Trigger

**HL-LHC** High Luminosity-LHC

**LHC** Large Hadron Collider

**L1** Level-1

**MC** Monte-Carlo

**PD** Primary Dataset

**SM** Standard Model

**UFO** Universal Feynman Rule

# 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 $e\mu\tau q$ Interactions**

# CHAPTER 11

## Previous searches

### 11.1 ATLAS

### 11.2 CMS

# CHAPTER 12

## Datasets, Simulated Samples and Triggers

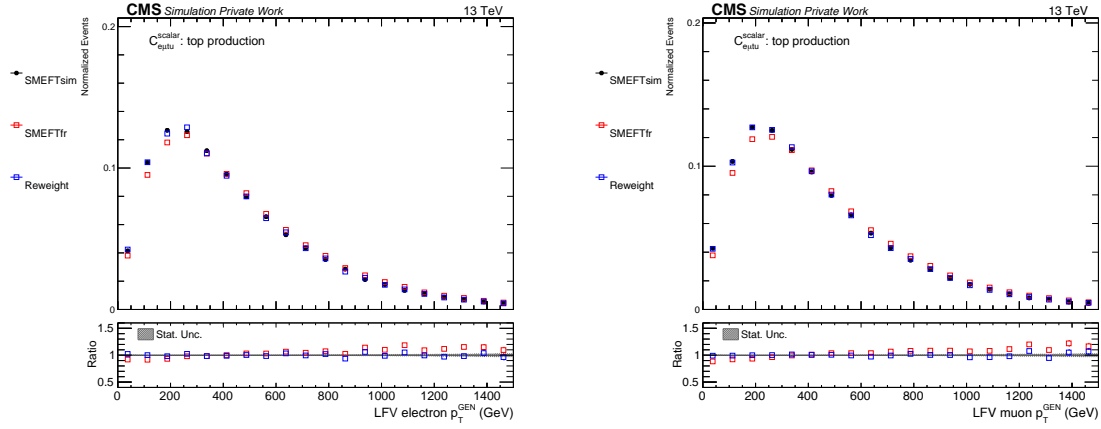
This analysis is based on data collected by the Compact Muon Solenoid (CMS) experiment in 2016-2018 from  $pp$  collisions at a center-of-mass energy of 13 TeV corresponding to an integrated luminosity of  $138 \text{ fb}^{-1}$ . There were approximately 30 simultaneous  $pp$  collisions occurring per 25 ns. Based on online selection criteria, fully reconstructed collision data that contain high-level physics objects are divided in “Primary Datasets (PDs)”, which include “DoubleEG”, “DoubleMu”, “MuonEG”, “SingleElectron”, and “SingleMuon” for 2016 and 2017 data taking era. In 2018, “SingleElectron”, “DoubleEG” are replaced by “EGamma”. The names of these PDs reflect the selection criteria. The data taking conditions in 2016-2018 were difference across the years. To account for this, all MC samples are generated separately for each year separately.

### 12.1 Signal Samples

This analysis made use of signal samples that were initially produced for the LFV analysis in the dilepton channel `citeDilepton`, which used SMEFTfr v2 UFO `citeSMEFTsim`. The datacards used to generate these samples can be found at `github citedatacards`. It was first reported by ATLAS authors `citeLHCTOP` that SMEFTfr v2 was buggy and we decided to adopt SMEFTsim v3 UFO `citeSMEFTsim` as the new CMS standard moving forward. However, it would be too late to reproduce all the samples at the time of this decision, and we decided to implement the following changes:

The difference in kinematic distributions are shown in Fig12.1. A reweighting tool `citereweight` is used to emulate the kinematic distributions produced by SMEFTsim. In addition to the shape components, we updated the top production cross sections from `citeDilepton`, which are summarized in `Table reftab:signal`. The top decay cross sections are kept the same because they do not come from UFO.

All but the  $Q_{lq}^{(3)ijkl}$  operator in `Table reftab:dimension6` are considered because of the equivalence of  $Q_{lq}^{(3)ijkl}$  and  $Q_{lq}^{(1)ijkl}$ . To reduce the number of free parameters, we assume 6 independent Wilson coefficients and apply condition `refeq1:1-refeq1:6`.



**Figure 12.1:** Comparison of kinematic distributions produced by different UFOs: LFV electron  $p_T$  (left), LFV muon  $p_T$  (right). The "Reweight" (shown in blue curve) are produced by applying weights produced by citereweight to EFT samples produced by SMEFTfr.

$$C_{lq} = C_{lq}^{(1)1213} + C_{lq}^{(1)2113} + C_{lq}^{(1)1231} + C_{lq}^{(1)1213}, \quad (12.1)$$

$$C_{lu} = C_{lu}^{1213} + C_{lu}^{2113} + C_{lu}^{1231} + C_{lu}^{1213}, \quad (12.2)$$

$$C_{eq} = C_{eq}^{1213} + C_{eq}^{2113} + C_{eq}^{1231} + C_{eq}^{1213}, \quad (12.3)$$

$$C_{eu} = C_{eu}^{1213} + C_{eu}^{2113} + C_{eu}^{1231} + C_{eu}^{1213}, \quad (12.4)$$

$$C_{lequ}^{(1)} = C_{lequ}^{(1)1213} + C_{lequ}^{(1)2113} + C_{lequ}^{(1)1231} + C_{lequ}^{(1)1213}, \quad (12.5)$$

$$C_{lequ}^{(3)} = C_{lequ}^{(3)1213} + C_{lequ}^{(3)2113} + C_{lequ}^{(3)1231} + C_{lequ}^{(3)1213} \quad (12.6)$$

The signal is simulated using MADGRAPH5\_aMC@NLO with SMEFT\_2l2q\_UFO\_unitary as the UFO. The signal samples are produced with an assumption that  $C_a=1$ ,  $\Lambda=1\text{TeV}$ . Since all vector-like operators show similar results, we combine them and call them vector LFV signals.

## 12.2 Background Samples

For 2016 MC, PDF set NNPDF3.0 and PYTHIA8 tune CUETP8M1 are used. In 2017 and 2018, NNPDF3.1 and PYTHIA8 tune CP5 are used. Simulated processes are divided into the following categories:  $t\bar{t}$ , WZ production, multi-boson production (other than WZ), the associated production of  $t(\bar{t})$  and a boson or/and a quark (e.g.  $t\bar{t}W$ ,  $t\bar{t}Z$  and  $t\bar{t}H$ ) and others (e.g. W jets, DY jets). MC samples used in this analysis are summarized in Table 12.2. The cross sections of the prompt backgrounds used for this analysis are the same as the cross sections listed in citeXSref. The nonprompt backgrounds in this analysis will be modeled with data-driven technique. The MC samples listed in "nonprompt" prompt category of Table 12.2 are only used

**Table 12.1:** Theoretical cross sections for top production and decay for each CLFV coupling. Uncertainties related to PDF and QCD scale are given ( $\sigma_{-scale}^{+scale} \pm \text{PDF}$ ).

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

as reference for initial study. The charge-flip backgrounds constitute a very small contribution to the signal region.

### 12.3 Triggers

To achieve a trigger optimal acceptance, a combination of single-lepton, di-lepton and tri-lepton triggers are used to select events. These triggers are summarized in Table reftab:triggers16-reftab:triggers18. The following trigger logic was implemented to remove the overlap between different datasets:

- Events in MC datasets are required to fire at least one of the triggers listed in Table A.1-A.3.
- Events in Single Muon datasets are required to fire at least one of the triggers listed under “SingleMuon”.
- Events in Double Muon 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 “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

**Table 12.2:** Summary of information related to MC sample generation.

|                        | Process       | Event Generator | Perturbative QCD | Tune          | XS precision |
|------------------------|---------------|-----------------|------------------|---------------|--------------|
| Prompt backgrounds     | WZ            | Madgraph        | NLO              | CUETP8M1(CP5) | NLO [1]      |
|                        | ZZ            | Powheg          | NLO              | CUETP8M1(CP5) | NLO [1]      |
|                        | VVV           | Madgraph        | NLO              | CUETP8M1(CP5) | NLO          |
|                        | ttW, ttZ      | Madgraph        | NLO              | CUETP8M1(CP5) | NLO [3, 4]   |
|                        | ttH           | Powheg          | NLO              | CUETP8M1(CP5) | NLO [4]      |
|                        | tZq           | Madgraph        | NLO              | CP5           | NLO          |
|                        | tHq, tHW, tWZ | Madgraph        | LO               | CUETP8M1(CP5) | LO           |
| Non-prompt backgrounds | tt            | Powheg          | NLO              | CUETP8M1(CP5) | NNLO [2]     |
|                        | DYM50         | Madgraph        | NLO              | CUETP8M1(CP5) | NNLO [5]     |
|                        | DYM10to50     | Madgraph        | LO               | CUETP8M1(CP5) | NLO [5]      |

listed under “SingleMuon” or “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” or “DoubleMu” or “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” or “DoubleMu” or “MuonEG”.

# CHAPTER 13

## Object Selection

### 13.1 Leptons

#### Electron

We apply the following selections to electron objects:

- $p_T > 20$  GeV,
- $|\eta| < 2.4$ ,
- the gap ( $1.4442 < |\eta_{SC}| < 1.566$ ) is removed,
- $|d_{xy}| < 0.05$  cm,  $|d_z| < 0.1$  cm,
- missing inner hits  $< 2$ ,
- significance of the 3D impact parameter  $SIP_{3D} < 8$ ,
- pass the TOP MVA ID Tight WP  $citemvaTOP (> 0.9 \text{ across all years})$ ,
- $minisolation < 0.12$ .

Electron energy scale corrections [citeEcalScale](#) are applied.

#### Muon

Similarly, we apply the following selections to muon objects:

- $p_T > 20$  GeV,
- $|\eta| < 2.4$ ,
- $|d_{xy}| < 0.05$  cm,  $|d_z| < 0.1$  cm,
- significance of the 3D impact parameter  $SIP_{3D} < 8$ ,
- pass the cut-based muon ID  $citemuID$  (medium working point),
- pass the TOP MVA ID Tight WP  $citemvaTOP (> 0.9 \text{ across all years})$ ,
- $minisolation < 0.12$ .

Muon energy (Rochester) corrections are applied to low  $p_T$  muons ( $p_T < 200 \text{ GeV}$ ).

The TOP MVA ID is used to suppress the contamination of non-prompt leptons. Both electron and muon selection in the analysis is synchronized with.

The isolation variable "minisolation" used to select both electrons and muons is also used as input to the TOP MVA training. It is similar to a normal isolation variable but with a  $p_T$  dependent cone size:

$$R = \max(0.05, \min(0.2, \frac{10 \text{ GeV}}{p_T})). \quad (13.1)$$

This analysis requires three leptons selected by the criteria above. The  $p_T$  threshold on the leading lepton is 38 GeV.

## 13.2 Jets

Jets are reconstructed from particle flow, PF, candidates using the Anti-kT clustering algorithm with a cone parameter  $R = 0.4$  (AK4). The Charged Hadron Subtraction algorithm, CHS, is applied to the jets, AK4PFCHS jets, in order to groom them, removing extra radiation from pileup. After this preliminary reconstruction the groomed jets must satisfy the below conditions:

- $p_T > 30 \text{ GeV}$ ,
- $|\eta| < 2.4$ ,
- TightLepVeto citejetID,
- Pile up jet ID loose citepujetID, if  $p_T < 50 \text{ GeV}$ .

Jets are also required to be isolated from selected leptons. A cone of  $\Delta R < 0.4$  around each jet candidate is defined and jets are removed if any selected leptons are found within such a cone.

Jets from MC are corrected in order to match the data which is skewed by pileup (L1FastJet) and L2L3 MC-truth corrections. In order to correct the MC jets, we applied the recommended Jet Energy Corrections. For reference the versions of the JECs we used for

## 13.3 B-tagging



# CHAPTER 14

## Event Selection

We require events to contain exactly three selected leptons. Events with different lepton compositions are further categorized into different channels:  $eee$ ,  $\mu\mu\mu$ ,  $e\mu l$ . In all three channels, the sum of the charges of three leptons are required to be 1 or -1. The leading leptons in all selected events are required to be matched with trigger objects via  $\Delta R < 0.2$ . Within each channel, different regions are defined to further understand signal and background.

$e\mu l$  is the channel where the presence of the signal is expected. We require at least one pair of leptons of Opposite-Sign-Opposite-Flavor (OSOF). Such a requirement is achieved if:

- Both positively charged and negatively charged leptons are present.
- Both electrons and muons are present.

Events with the presence of the same flavor leptons are categorized into  $eee$  or  $\mu\mu\mu$  channels to study the background composition.

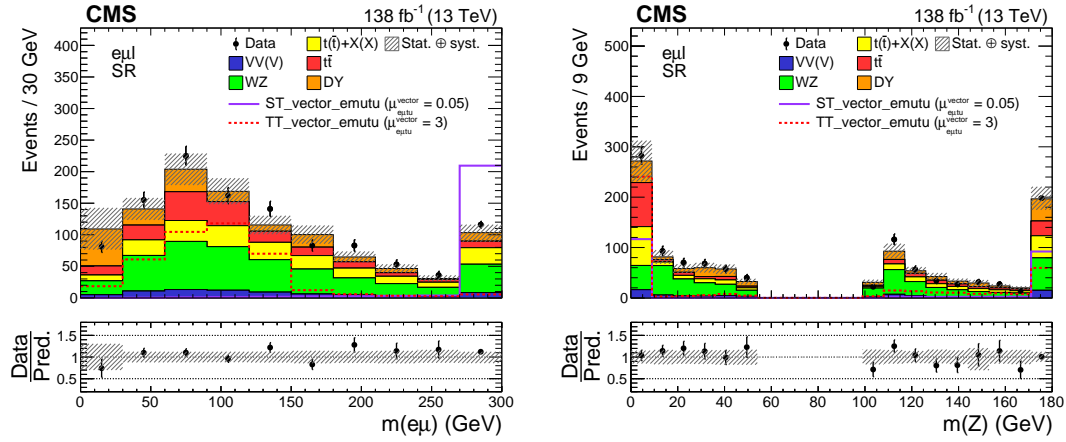
### 14.1 Signal Region

In  $e\mu l$  channel, we set up our signal region (SR) targeting LFV signal processes. Apart from the OSOF pair, we require the events to have big MET ( $MET > 20$  GeV). Events with an Opposite-Sign-Same-Flavor (OSSF) lepton pair, the invariant mass of which is between 50 GeV and 106 GeV, are removed to suppress contribution from Z production. To further suppress background contribution, we require at least one jet and at most one b-tag jet.

Distributions of leading lepton  $p_T$  and invariant mass of Opposite-Sign  $e\mu$  pair are shown in Figure 14.1. Please note: Although we intend to deploy a data-driven technique to estimate non-prompt background in the actual analysis, all backgrounds in Figure 14.1 are estimated using MC simulation only. This serves as a preliminary check to understand the components of different backgrounds in SR.

Using the invariant mass of the Opposite-Sign  $e\mu$  pair, SR is further divided into two subsets to create ST and TT enriched regions:

- SR1,  $M_{e\mu} < 150$  GeV: TT enriched.



**Figure 14.1:** Distributions of different kinematic variables estimated in SR (full run II). Backgrounds are estimated using MC only. From left to right: invariant mass of the  $e\mu$  pair, leading lepton  $p_T$ .

- SR2,  $M_{e\mu} > 150\text{GeV}$ : ST enriched.

## 14.2 Validation Region

Events are required to pass trigger requirements listed in Table A.1-A.3. The leading lepton is required to have a  $p_T > 38\text{ GeV}$ . The sum of the charges of the three selected leptons is required to be either 1 or -1. Selections applied to different regions are summarized in Table 14.1 and are illustrated in Figure 14.2.

- Acronyms
  - $C_i$ : electric charge of the  $i$ -th lepton
  - OnZ: requiring the presence of at least one Opposite-Sign-Same-Flavor lepton pair with an invariant mass between 50 GeV and 106 GeV.
  - OffZ: veto events that pass OnZ
  - MET: missing transverse energy
  - $n_{\text{jet}}$ : number of jets that pass jet selection
  - $n_{\text{bjet}}$ : number of b-tag jets that pass jet selection
  - SR: signal region
  - VR: validation region for data-driven method
  - WZ CR: WZ control region

| Channel     | Region       | $ \Sigma_i C_i =1$ | 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        | ✓                  | -   | ✓    | ✓            | ✓       | ✓        |

Table 14.1: Summary of the cuts applied to different regions

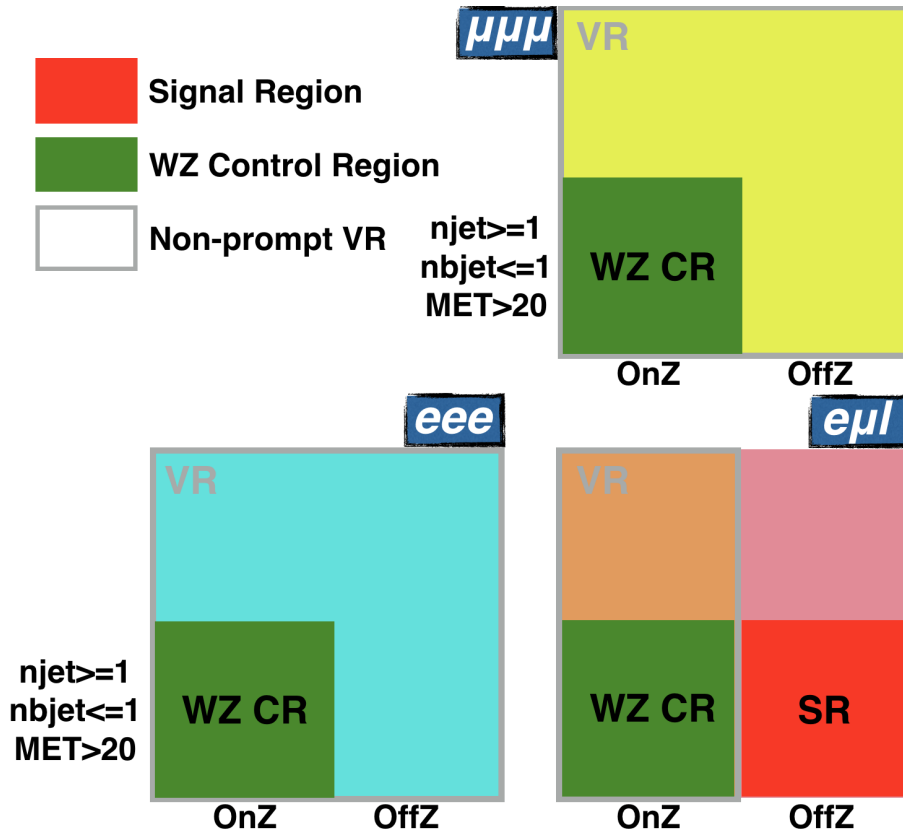
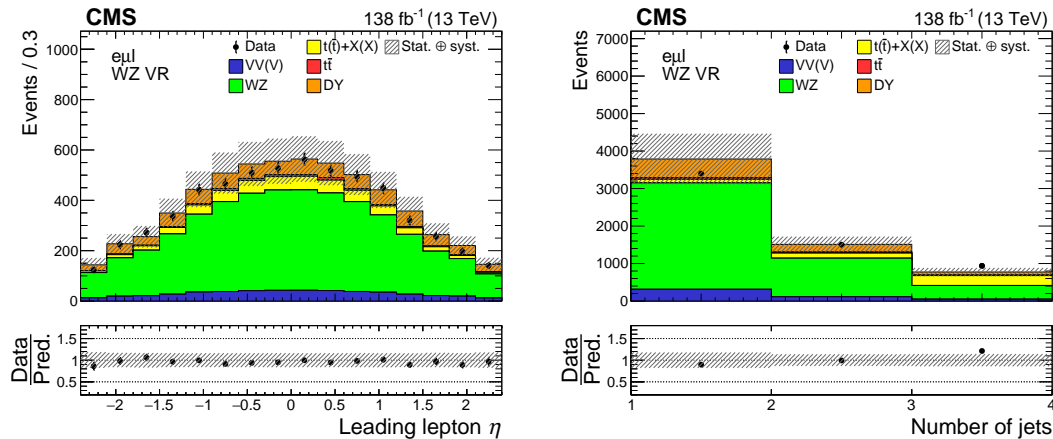
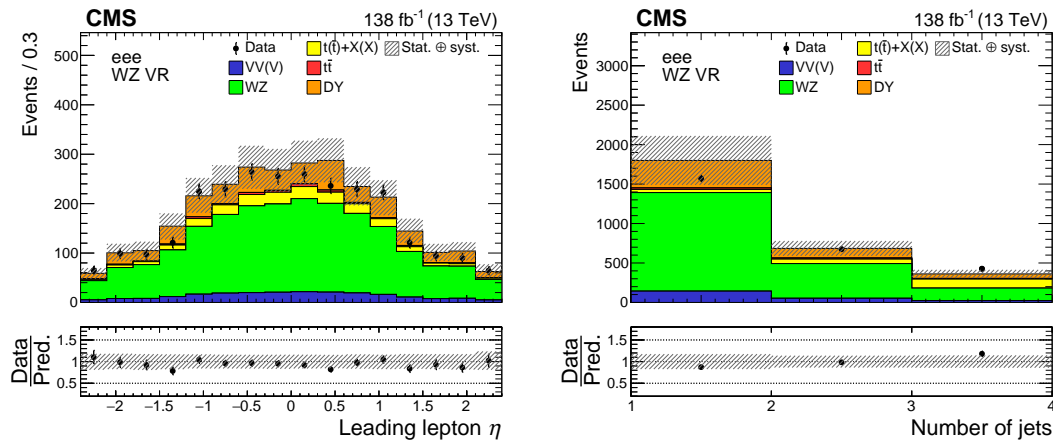
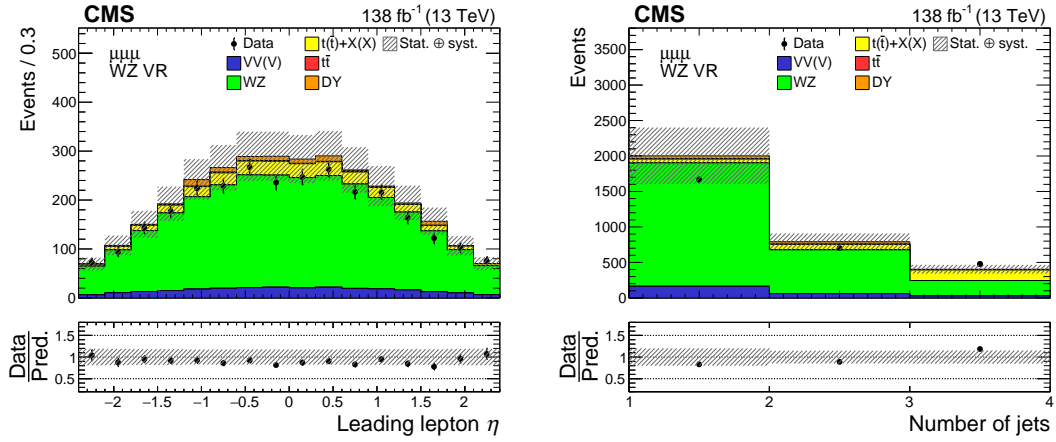


Figure 14.2: Illustration of event selections applied to different regions

There are two types of signal-depleted regions defined within  $eee/\mu\mu\mu$  channel: validation region (VR) and WZ control region. The distinction between these two regions comes down to





**Figure 14.5:** Distributions of different kinematic variables estimated in WZ control region (full run II). Backgrounds are estimated using MC only. From left to right: leading lepton  $\eta$ , jet multiplicity.

the SM top-quark.

- We take the x and y component of MET 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 (bachelor lepton+neutrino) must be equal to W boson mass.
- If there is no real solution, we take the real part of the complex solution.
- If there is more than one real solution, we chose the solution that is the closest to the  $p_z$  of the bachelor lepton.

In events where there is more than one candidate of bachelor lepton, the lepton that gives a top mass that is the closest to the the SM top-quark mass is chosen as the bachelor lepton.

Once the bachelor lepton has been determined, the OSOF pair is combined with jets One-By-One to reconstruct LFV top 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 candidates, the candidate that gives a top mass that is the closest to the the SM top-quark mass is chosen.

Kinematic reconstruction of heavy objects like top-quark is carried out using selected leptons, jets, and MET.

#### • Z boson candidate

- Z boson candidate is reconstructed by combining two Opposite-Sign-Same-Flavor leptons

- Whenever possible, Z candidate with an invariant mass that is closest to Z boson mass is chosen

- **Standard model top-quark candidate**

- Standard model top-quark candidate is reconstructed by combining a jet, a lepton, and MET
- Jet with the highest b-tagging score is always chosen
- In  $e\mu/\mu\mu$  channel, flavor-violating leptons are excluded from reconstruction
  - \* Whenever more than one lepton is available due to possible combinatorics of flavor-violating  $e\mu$  pair, lepton that forms a top-quark candidate with an invariant mass that is closest to the top-quark mass is chosen
- In  $eee/\mu\mu\mu$  channel, leptons that form Z candidate are excluded
- Lepton is combined with MET to reconstruct W candidate
  - \* Transverse component of the neutrino momentum is taken from MET
  - \* Neutrino  $p_z$  is obtained by W mass constrain set on the lepton-neutrino system

- **Flavor-violating top-quark candidate**

- Flavor-violating top-quark candidate is reconstructed by combining a jet and an Opposite-Sign  $e\mu$  lepton pair
- Jet with the highest b-tagging score is excluded from reconstruction unless it fails b-tagging
- Whenever possible, the jet that forms a flavor-violating top-quark candidate with an invariant mass that is the closest to the top-quark mass is chosen
- Lepton that forms the standard model top-quark candidate is excluded

- **The first b-jet+lepton system**

- Jet with the highest b-tagging score is always chosen
- All three leptons can form this system with the chosen jet
  - \* Lepton that forms a jet-lepton system with the lowest invariant mass is chosen

- **The second b-jet+lepton system**

- Jet with the highest b-tagging score is excluded unless it passes b-tagging
- Whenever possible (after the previous step), jet with the highest b-tagging score is chosen
- Lepton that forms the first jet-lepton system is excluded

- Leptons that have the same sign as the lepton that forms the first jet-lepton system is excluded
- Whenever possible, lepton that forms a jet-lepton system with the lowest invariant mass is chosen

# CHAPTER 15

## Nonprompt Background Estimate

In this analysis, the term *real* leptons refers to prompt leptons that originate from the primary interaction, like the decays of vector bosons that are produced via hard collisions. *Fake* leptons refer to non-prompt leptons that are mostly produced in the decays of b-quarks, c-quarks, pions, or from photon conversion. A relatively low ratio of *fake* leptons in a lepton selection is desired and can be achieved by applying stringent isolation requirements and/or an MVA-based lepton identification specifically trained against *fake* leptons.

Non-prompt backgrounds are defined to be backgrounds with at least one *fake* lepton passing the analysis selection, in this case generally dominated by Drell-Yan and  $t\bar{t}$  production. An accurate estimation of non-prompt backgrounds is difficult to achieve through MC modeling. Therefore, a data-driven technique called the 3D matrix method is used to estimate the non-prompt backgrounds.

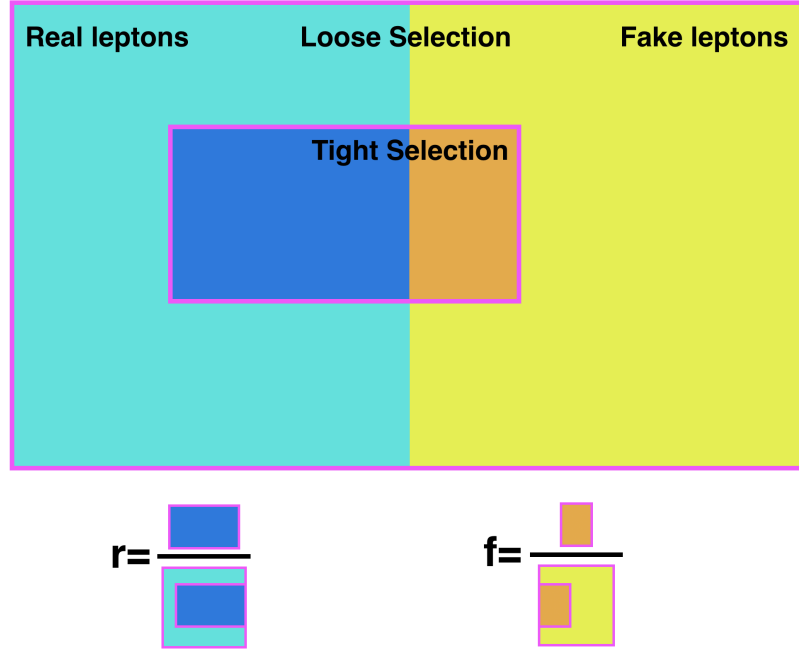
A brief description of the matrix method is given in this section followed by details on its implementation. In [autorefyou](#) you can find more information on the generalization of the matrix method to 2D and 3D. The appendices also contain a validation of the method and its performance is compared to the results of the fake factor method, another commonly used data-driven technique.

### 15.1 The Matrix Method

The matrix method [citematrix2012](#) is a data driven technique used to estimate the fraction of *fake* leptons that pass a given lepton selection, referred to as "*tight*". The *tight* selection usually incorporates tight lepton identification and isolation requirements and should be similar if not exactly the same as lepton selection in the actual analysis. The *loose* selection is obtained by loosening the *tight* selection. The *loose* selection is used as a baseline such that any leptons that pass it will fall into one of the two exclusive categories: *tight* or *not tight*. The matrix method treats *real* and *fake* leptons separately. As a result, *real* and *fake* efficiencies are introduced (see [Figure 15.1](#)).

In a simplified scenario with only one lepton in the final state, the *real* efficiency  $r$  measures the likelihood of *real* leptons pass *tight* selection. It is treated as an observable that can be obtained





**Figure 15.1:** Illustration of the *real* efficiency  $r$  and the *fake* efficiency  $f$ .

through measurement:

$$r = \frac{n_R^T}{n_R^T + n_R^{\bar{T}}}, \quad (15.1)$$

in which  $n_R^T/n_R^{\bar{T}}$  denotes the number of events with a *real* lepton that is *tight/not tight*.

Similarly, *fake* efficiency  $f$  can be expressed as

$$f = \frac{n_F^T}{n_F^T + n_F^{\bar{T}}}, \quad (15.2)$$

in which  $n_F^T/n_F^{\bar{T}}$  denotes the number of events with a *fake* lepton that is *tight/not tight*.

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

In our 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:

$$\begin{aligned} N^T &= N_R^T + N_F^T \\ N^{\bar{T}} &= N_R^{\bar{T}} + N_F^{\bar{T}}, \end{aligned} \tag{15.3}$$

in which capital letter "N" is used to indicate that these numbers are referring to events in a region that is different from MR.  $N_R^{\bar{T}}/N_F^{\bar{T}}$  can be expressed in terms of  $r/f$  and  $N_R^T/N_F^T$  according to Equation 15.1/15.2 and the assumption that  $r/f$  is transferable to other regions:

$$\begin{aligned} N^T &= r \frac{N_R^T}{r} + f \frac{N_F^T}{f} \\ N^{\bar{T}} &= (1-r) \frac{N_R^T}{r} + (1-f) \frac{N_F^T}{f}. \end{aligned} \tag{15.4}$$

Equation 15.4 can also be expressed in the form of matrix:

$$\begin{pmatrix} N^T \\ N^{\bar{T}} \end{pmatrix} = \begin{pmatrix} r & f \\ 1-r & 1-f \end{pmatrix} \begin{pmatrix} N_R^T/r \\ N_F^T/f \end{pmatrix}. \tag{15.5}$$

Since those two numbers that appear on the lefthand side of the equation can be obtained through measurement, a matrix inversion can give us an estimation of *fake* leptons that pass *tight* selection, denoted by  $N_F^T$ .

## 15.2 Extraction of Prompt and Nonprompt Efficiencies

The description in previous section deals with a scenario where only one lepton is studied. This analysis uses a generalized version of the matrix method (i.e. 3D matrix method. See `autorefsec:appendixnonprompt` for more details), where all three leptons are considered to be possibly *fake*.

The measurement of both  $r$  and  $f$  are performed in dilepton MRs using the Tag-and-Probe approach. Datasets used to measure  $r$  and  $f$  are listed in Table 15.1-15.3. More information about MC samples listed in Table 15.1-15.3 can be found in Table ?? and ??. Events are required to fire either single-electron or single-muon trigger listed in Table 15.4.

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,

**Table 15.1:** Parent datasets (2016) used to generate TOP PAG NanoAOD-v6 samples used for  $r$  and  $f$  measurement. (X = "SingleElectron" and "SingleMuon".)

|     | Data type | Samples  |
|-----|-----------|--|
|     | data      | /X/Run2016B-17Jul2018_ver2-v1/MINIAOD<br>/X/Run2016C-17Jul2018-v1/MINIAOD<br>/X/Run2016D-17Jul2018-v1/MINIAOD<br>/X/Run2016E-17Jul2018-v2/MINIAOD<br>/X/Run2016F-17Jul2018-v1/MINIAOD<br>/X/Run2016G-17Jul2018-v1/MINIAOD<br>/X/Run2016H-17Jul2018-v1/MINIAOD  |
| $f$ | MC        | /WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8<br>/ZZTo4L_13TeV_powheg_pythia8<br>/ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8<br>/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8<br>/TTZToLL_M-1to10_TuneCUETP8M1_13TeV-madgraphMLM-pythia8<br>/THQ_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1<br>/THW_Hincl_13TeV-madgraph-pythia8_TuneCUETP8M1<br>/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8<br>/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8<br>/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8<br>/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8<br>/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8<br>/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8 |
| $r$ | MC        | /TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8   |

**Table 15.2:** Parent datasets (2017) used to generate TOP PAG NanoAOD-v6 samples used for  $r$  and  $f$  measurement. (X = "SingleElectron" and "SingleMuon".)

|     | Data type | Samples  |
|-----|-----------|--|
| $f$ | data      | /X/Run2017B-31Mar2018-v1/MINIAOD                           |
|     |           | /X/Run2017C-31Mar2018-v1/MINIAOD                           |
|     |           | /X/Run2017D-31Mar2018-v1/MINIAOD                           |
|     |           | /X/Run2017E-31Mar2018-v1/MINIAOD                           |
|     |           | /X/Run2017F-31Mar2018-v1/MINIAOD                           |
|     | MC        | /WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8               |
|     |           | /ZZTo4L_13TeV_powheg_pythia8                               |
|     |           | /ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8              |
|     |           | /TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8           |
|     |           | /WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8                     |
|     |           | /WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8                     |
|     |           | /WZZ_TuneCP5_13TeV-amcatnlo-pythia8                        |
|     |           | /ZZZ_TuneCP5_13TeV-amcatnlo-pythia8                        |
|     |           | /TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8            |
|     |           | /THQ_4f_Hincl_13TeV_madgraph_pythia8                       |
|     |           | /THW_5f_Hincl_13TeV_madgraph_pythia8                       |
|     |           | /tZq_ll_4f_ckm_NLO_TuneCP5_PWeights_13TeV-amcatnlo-pythia8 |
|     |           | /TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8   |
| $r$ | MC        | /TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8                    |

**Table 15.3:** Parent datasets (2018) used to generate TOP PAG NanoAOD-v6 samples used for  $r$  and  $f$  measurement. (X = "Egamma" and "SingleMuon".)

|     | Data type | Samples  |
|-----|-----------|--|
| $f$ | data      | /X/Run2018A-17Sep2018-v1/MINIAOD<br>/X/Run2018B-17Sep2018-v1/MINIAOD<br>/X/Run2018C-17Sep2018-v1/MINIAOD<br>/X/Run2018D-PromptReco-v2/MINIAOD  |
|     | MC        | /WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8<br>/ZZTo4L_13TeV_powheg_pythia8<br>/ttHTtoNonbb_M125_TuneCP5_13TeV-powheg-pythia8<br>/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8<br>/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8<br>/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8<br>/WZZ_TuneCP5_13TeV-amcatnlo-pythia8<br>/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8<br>/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8<br>/THQ_4f_Hincl_13TeV_madgraph_pythia8<br>/THW_5f_Hincl_13TeV_madgraph_pythia8<br>/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8<br>/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8 |
| $r$ | MC        | /TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8  |

| Channel  | path                  | dataset   | 2016 | 2017 | 2018 |
|----------|-----------------------|-----------|------|------|------|
| Electron | HLT_Ele27_WPTight_Gsf | data & MC | ✓    | -    | -    |
|          | HLT_Ele35_WPTight_Gsf | data & MC | -    | ✓    | ✓    |
| Muon     | HLT_IsoMu27           | data & MC | ✓    | ✓    | ✓    |

**Table 15.4:** Summary of the triggers used in the measurement of  $r$  and  $f$ .

- 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:  $\{0, 0.8, 1.6, 2.4\}$ ,
- Jet multiplicity:  $\{0 \text{ jet, at least 1 jet}\}$ .

The addition of jet multiplicity as a dependency is motivated by the unideal assumption that both  $r$  and  $f$  are independent of physics processes. We know that *fake* leptons of different origins are often reconstructed at a different rate. For instance, *fake* electrons coming from the decays of b-quarks have very different kinematics than *fake* electrons that originate from photon conversion. Therefore, the probabilities of these two types of *fake* electrons to pass any kinematic-dependent selections are going to differ. In other words, the proportion of  $t\bar{t}$  events that constitutes the non-prompt backgrounds has a direct impact on  $r$  and  $f$ . To achieve a better representation of the target region (in terms of background composition),  $r$  and  $f$  are linked to the number of jets in events.

The *fake* efficiency is measured in a same-sign dilepton region, in which the leading  $p_T$  lepton, used as the *tag* (See Table 15.5 for definition), is required to be matched with trigger objects via  $\Delta R < 0.2$ . The sub-leading lepton is required to pass the *loose* selection and is taken as the *probe*. Two orthogonal regions are defined within the dilepton MR, which account for the two bins of jet multiplicity:

- Events in dilepton MR with at 0 jets.
- Events in dilepton MR with at least 1 jet, and at most 1 b-tag jet.

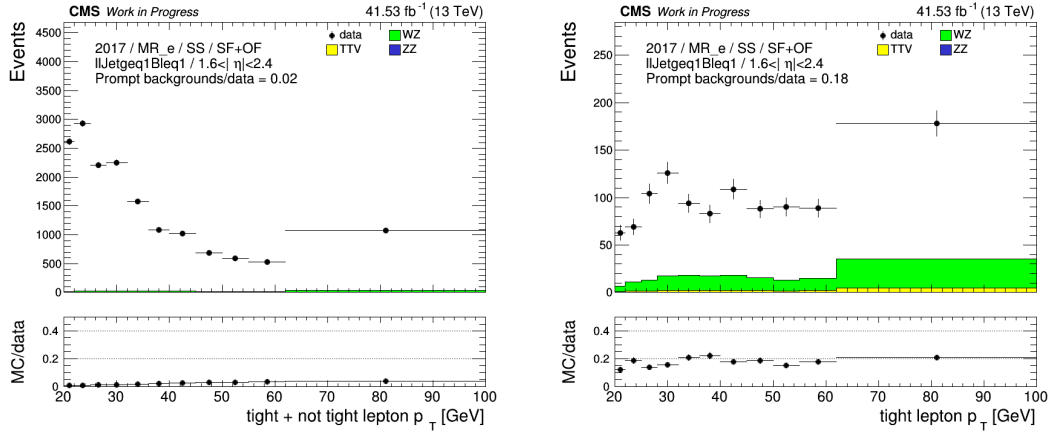
The selection of the second MR is motivated by the selection of the SR, in which at least one jet and at most one b-tag jet are required.

The contribution from prompt backgrounds, estimated from MC simulation, are subtracted from data (see Figure 15.2). Therefore, the fake efficiency  $f$  is calculated as:

$$f = \frac{n_{data}^T - n_{Prompt\ MC}^T}{n_{data}^{T+\bar{T}} - n_{Prompt\ MC}^{T+\bar{T}}}. \quad (15.6)$$

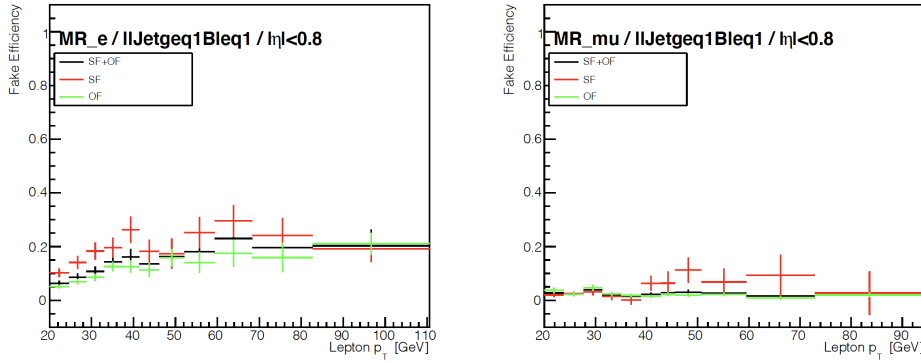
| Lepton   | Selection                  | Loose              | Tag                  | Tight <sup>b</sup> (Passing-Probe) |
|----------|----------------------------|--------------------|----------------------|------------------------------------|
| Electron | $p_T$                      | $> 20 \text{ GeV}$ | $> 38 \text{ GeV}^a$ | $> 20 \text{ GeV}$                 |
|          | $I_{iso}^{mini}$           | $< 0.4$            | $< 0.1$              | $< 0.12$                           |
|          | TOP MVA ID                 | $> -0.9$           | $> 0.95$             | $> 0.9$                            |
|          | Match with trigger objects | -                  | ✓                    | -                                  |
|          |                            |                    |                      |                                    |
| Muon     | $p_T$                      | $> 20 \text{ GeV}$ | $> 30 \text{ GeV}$   | $> 20 \text{ GeV}$                 |
|          | $I_{iso}^{mini}$           | $< 0.4$            | $< 0.1$              | $< 0.12$                           |
|          | cut-Based ID               | -                  | Medium WP            | Medium WP                          |
|          | TOP MVA ID                 | $> 0.5$            | $> 0.9$              | $> 0.9$                            |
|          | Match with trigger objects | -                  | ✓                    | -                                  |

**Table 15.5:** Summary of the lepton selections needed for the measurement of  $r$  and  $f$ . Please note: a) the minimum  $p_T$  cut for *tag* electron in 2016 dataset is reduced to 30 GeV to adjust for the trigger threshold; b) the *tight* selection here is the same as the lepton selection listed in autorefsec:reconstruction.1.



**Figure 15.2:** Distribution of lepton  $p_T$  in an electron *fake* efficiency MR (2017). In this particular example: both  $e + e$  (Same-Sign-Same-Flavor) and  $e + \mu$  (Same-Sign-Opposite-Flavor) pair are considered. At least one jet and at most one b-tag jet are required (the second jet multiplicity bin). *Probe* electron is required to have  $1.6 < |\eta| < 2.4$  (the third  $\eta$  bin). Contamination from prompt backgrounds are estimated with MC simulation. From left to right: *tight+not tight* electron, only *tight* electron.

The dependency of  $f$  on background composition is also reflected in the difference between  $f$  measured with different flavor compositions. For electrons, the *fake* efficiency measured with  $e + \mu$  selection is visibly different than the *fake* efficiency measured with  $e + e$  selection, which



**Figure 15.3:** Fake efficiency measured in 2017 data. 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. From left to right: electron  $f$ , muon  $f$ .

could suggest:

- *Fake* electrons that originate from the decays of  $b$ -quarks have a lower probability of being selected by *tight* selection than *fake* electrons coming from  $Z$ +jets events.
- The Same-Sign  $e + e$  selection is contaminated with *real* electrons and, to a less extent, the Same-Sign  $e + \mu$  selection.

Events that have two same-sign electrons with an invariant mass between 76 GeV and 106 GeV are removed from MR to suppress the charge-flip effect of electrons. No such cut has been introduced to muon MR due to the negligible rate of charge-flip in muons.

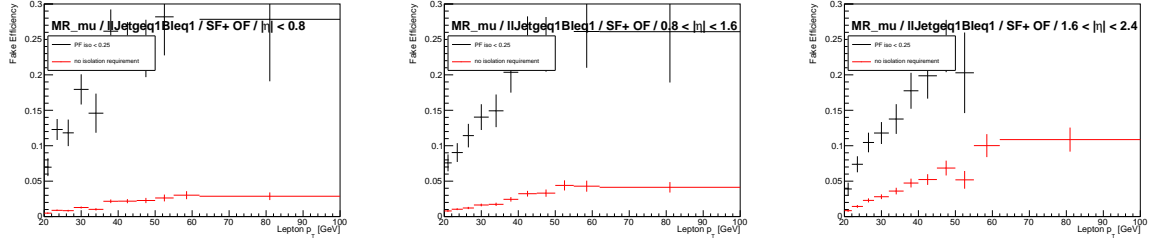
Lepton selections listed in Table 15.5 are referred to as *baseline*, which has been optimized to provide a robust non-prompt estimate. Variation of the *baseline* selection has also been studied.

There is a trade-off between muon  $r$  and  $f$  when changing the isolation requirement (See Figure 15.4-15.5). While it is ideal to reduce statistical uncertainties by loosening the *loose* definition, a low *real* efficiency should also be avoided. This change proves to have a minor effect on the overall estimate.

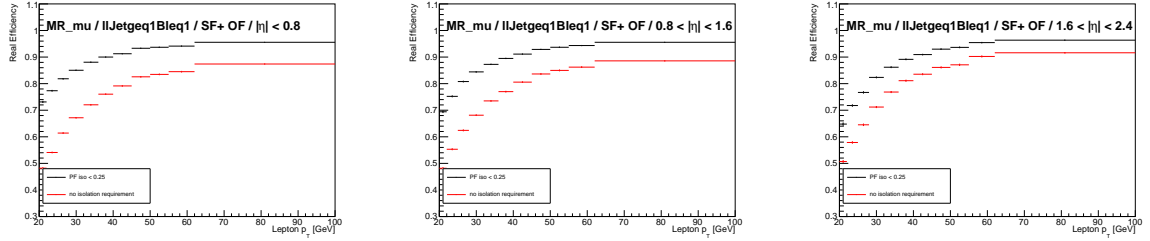
The *real* efficiency  $r$  is estimated with simulated  $t\bar{t}$  events in an Opposite-Sign dilepton region. The same lepton selection (See Table 15.5) is used to perform Tag-and-Probe. The leading lepton in  $p_T$  is used as a *tag* while the sub-leading lepton (with opposite sign) is taken as a *probe*. The variation of  $r$  between  $e + e$  and  $e + \mu$  flavor combination is negligible (See Figure 15.6). Therefore, only  $e + \mu$  events are used to measure *real* efficiency in order to minimize the contamination of *fake* leptons.

The selection criteria for  $r/f$  measurement region is summarized in Table 15.6. The latest version

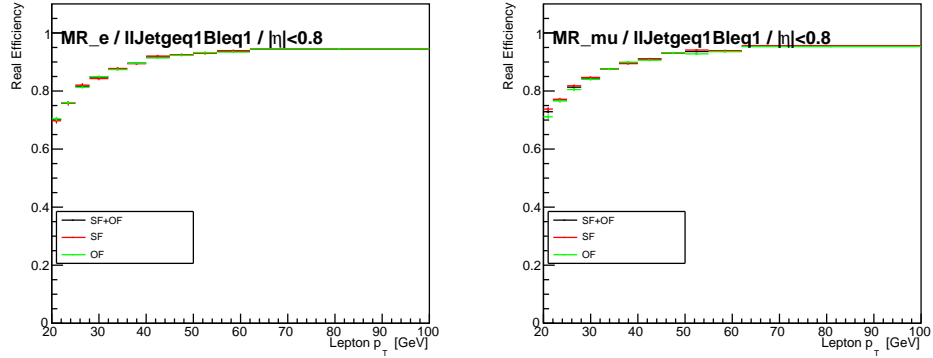




**Figure 15.4:** Comparison of the muon *fake* efficiencies measured with isolation requirement in loose definition (2017 datasets). From left to right:  $|\eta| < 0.8$ ,  $0.8 < |\eta| < 1.6$ ,  $1.6 < |\eta| < 2.4$ .



**Figure 15.5:** Comparison of the muon *real* efficiencies measured with different loose definition (2017 datasets). From left to right:  $|\eta| < 0.8$ ,  $0.8 < |\eta| < 1.6$ ,  $1.6 < |\eta| < 2.4$ .



**Figure 15.6:** *Real* efficiency measured in simulated  $t\bar{t}$  events (2017). 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. From left to right: electron  $r$ , muon  $r$ .

of the *real* and *fake* efficiency is summarized in `autorefsec:RandF`.

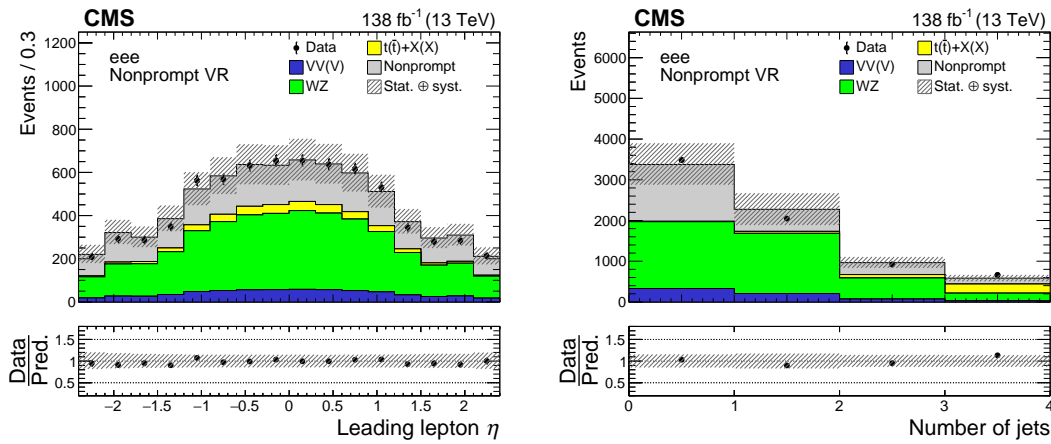
| Observable | jet bin       | # of selected leptons | lepton flavor composite | $ \Sigma_i C_i $ | OffZ | njet $\geq$ 1 | nbjet $\leq$ 1 |
|------------|---------------|-----------------------|-------------------------|------------------|------|---------------|----------------|
| $f$        | 0 jet         | 2                     | any                     | 1                | ✓    | -             | -              |
|            | 1 or more jet | 2                     | any                     | 1                | ✓    | ✓             | ✓              |
| $r$        | 0 jet         | 2                     | $e\mu$ only             | 0                | ✓    | -             | -              |
|            | 1 or more jet | 2                     | $e\mu$ only             | 0                | ✓    | ✓             | ✓              |

**Table 15.6:** Summary of the cuts applied to the  $r/f$  measurement region.

### 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 with MC simulation while non-prompt backgrounds are estimated with the matrix method. A summary of the selections applied to these VRs is given in `autorefsec:selection.4`.

Distribution of lepton lepton  $\eta$  and jet multiplicity are shown in Figure 15.7-15.9. Distributions of other variables are included in `autorefsec:VRmatrix`

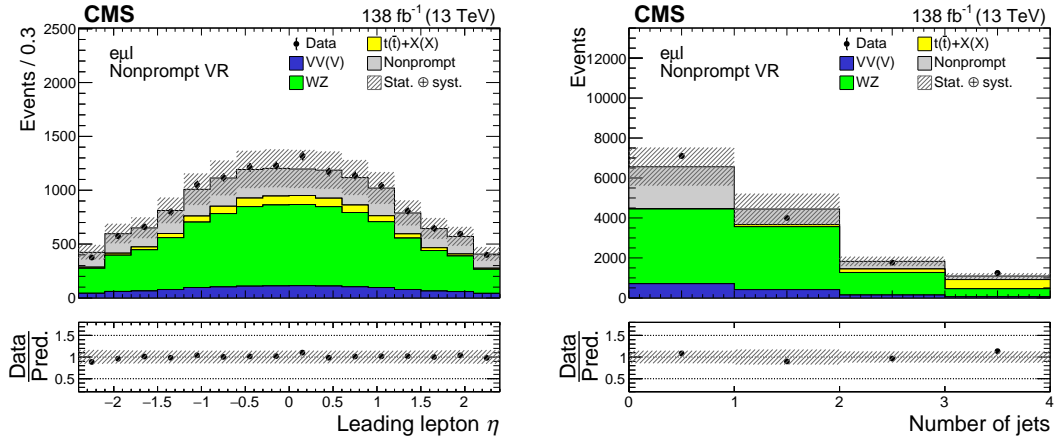


**Figure 15.7:** Distributions of different kinematic variables estimated in VR with three electrons. From left to right: leading lepton  $\eta$ , jet multiplicity.

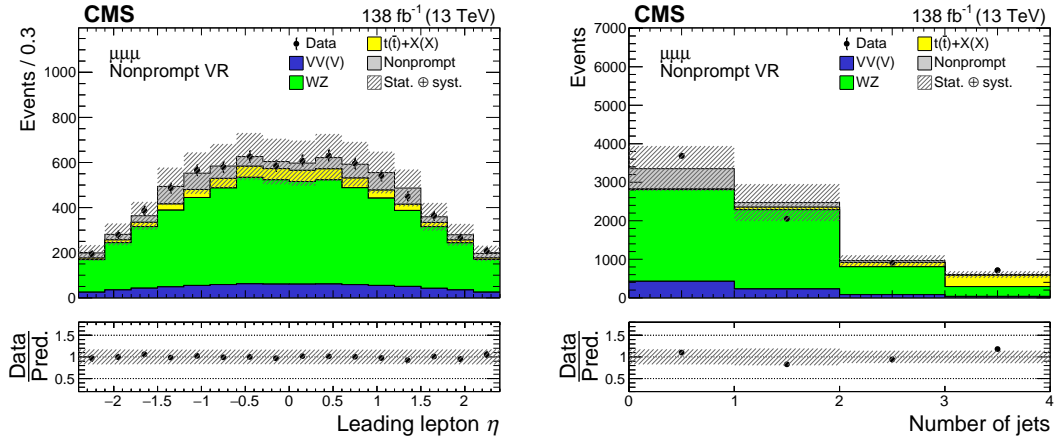
### 15.4 Nonprompt estimate in SR

The matrix method is used to estimate non-prompt backgrounds in the signal region due to its superior performance over other alternative methods (e.g. fake factor method discussed in `autorefsec:appendixnonprompt`). Distributions of leading lepton  $\eta$ , jet multiplicity, n-jet multiplicity, missing transverse energy, LFV-Top mass and Z mass are shown in Figure 15.10-15.11. Distributions of other variables are included in `autorefsec:SRDataDriven`.

The number of expected events from various kinds of backgrounds are shown in Table 15.7.

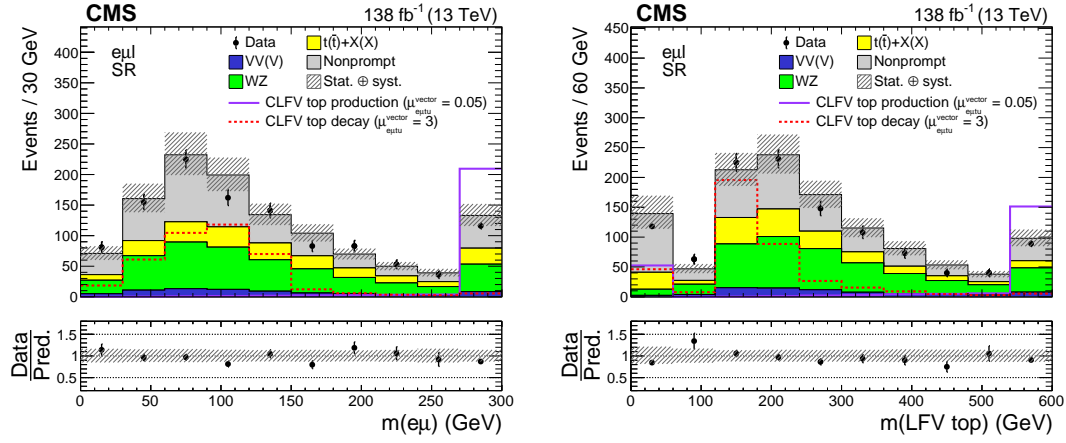


**Figure 15.8:** Distributions of different kinematic variables estimated in VR with electron, muon and a third light lepton. From left to right: leading lepton  $\eta$ , jet multiplicity.

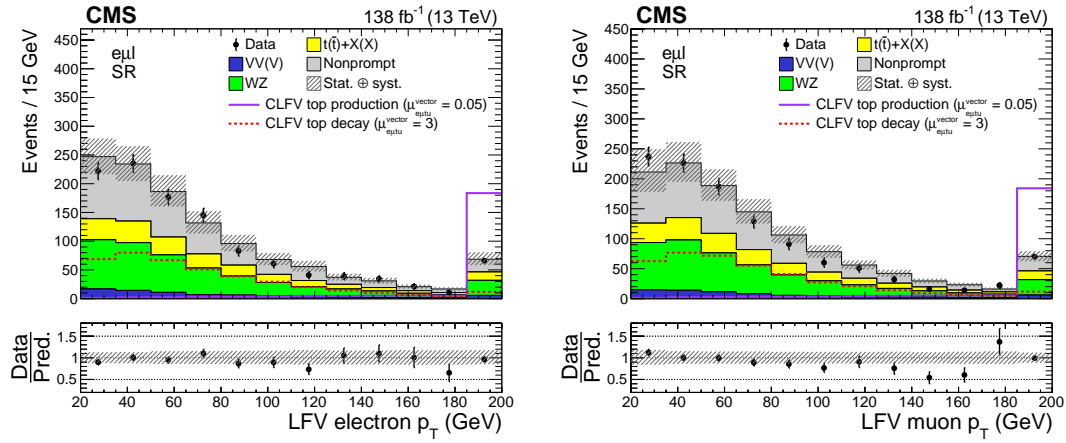


**Figure 15.9:** Distributions of different kinematic variables estimated in VR with three muons. From left to right: leading lepton  $\eta$ , jet multiplicity.

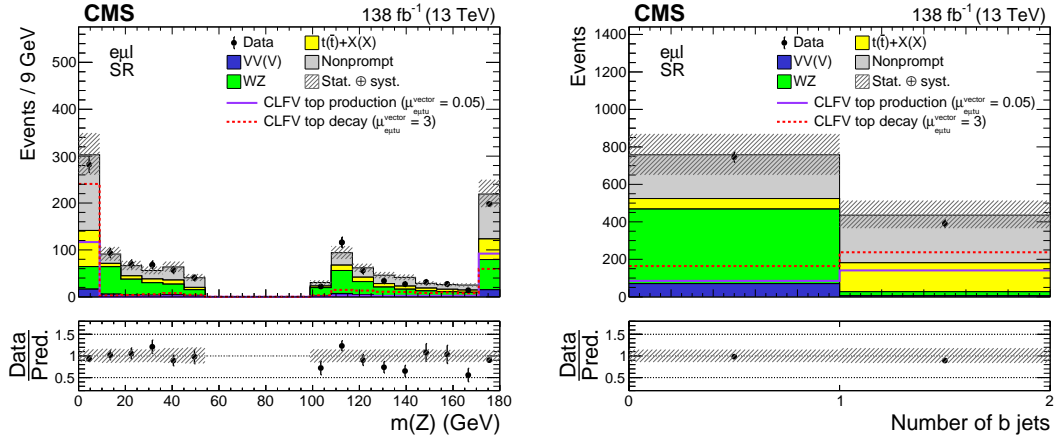
Representative bin yields and their statistical uncertainties are summarized in `citeStatError`.



**Figure 15.10:** Distributions of different kinematic variables estimated in SR (full run II). From left to right: b-jet multiplicity, missing transverse energy.



**Figure 15.11:** Distributions of different kinematic variables estimated in SR (full run II). From left to right: b-jet multiplicity, missing transverse energy.



**Figure 15.12:** Distributions of different kinematic variables estimated in SR (full run II). From left to right: mass of the flavor-violating-top-quark candidate, mass of the Z boson candidate

**Table 15.7:** 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  $C_{e\mu tu}^{\text{vector}}/\Lambda^2 = 1\text{TeV}^{-2}$ , is also listed for reference. The signal yields include contributions from both top production and decay modes.

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

# CHAPTER 16

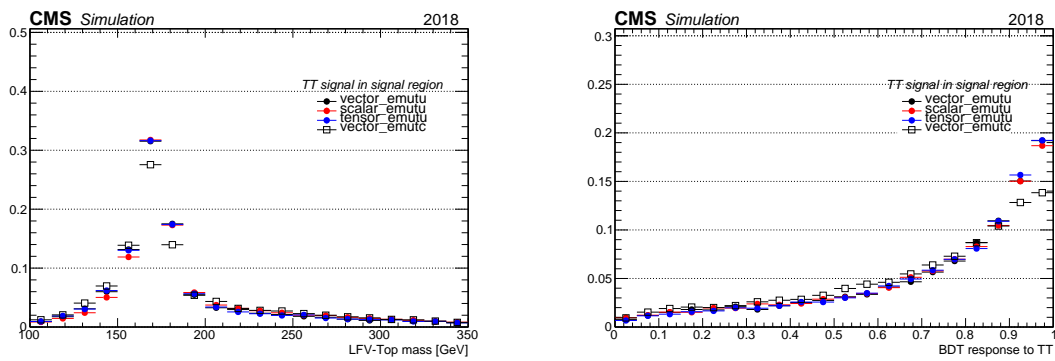
## Signal Extraction with Boosted Decision Trees

To further separate the LFV signal from the standard model backgrounds in our signal region, a Boosted Decision Tree (BDT) was employed using the xgBoost package citeChen:2016:XST:2939672.2939785.

### 16.1 BDT Setup

Since the kinematic distributions of final state particles in our two signal channels (single top production and top decay) differ significantly (see Figure 16.3-16.7), the SR is divided into two parts (SR1:  $M_{e\mu} \leq 150\text{GeV}$ ; SR2:  $M_{e\mu} > 150\text{GeV}$ ) in order to achieve the optimal sensitivity. Within each sub-region of the SR, a binary BDT is trained.

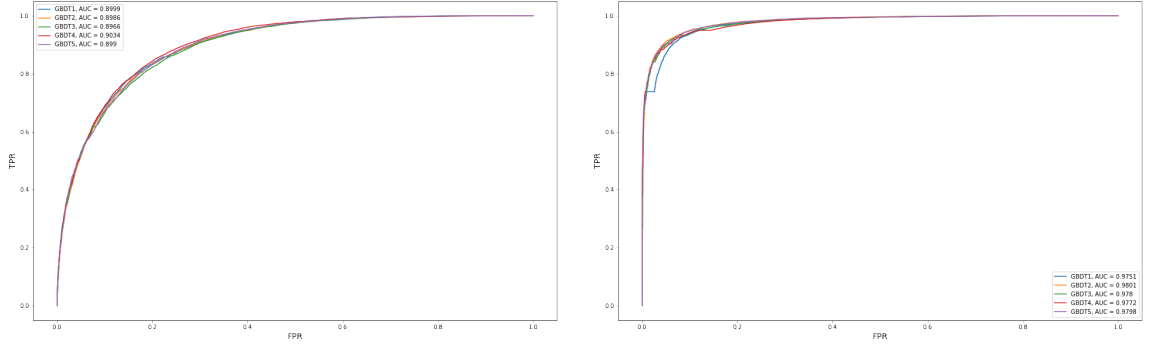
The signal datasets used for BDT training are MC samples that correspond to single top LFV production (ST) and top LFV decay (TT). The Lorentz structure and the flavor of the up-type quark involved in LFV interaction were shown to have a minor impact on the kinematics of final state particles (See Figure 16.1). Therefore, they were not further distinguished.



**Figure 16.1:** Normalized distribution in SR1. From left to right: LFV top mass, BDT shape

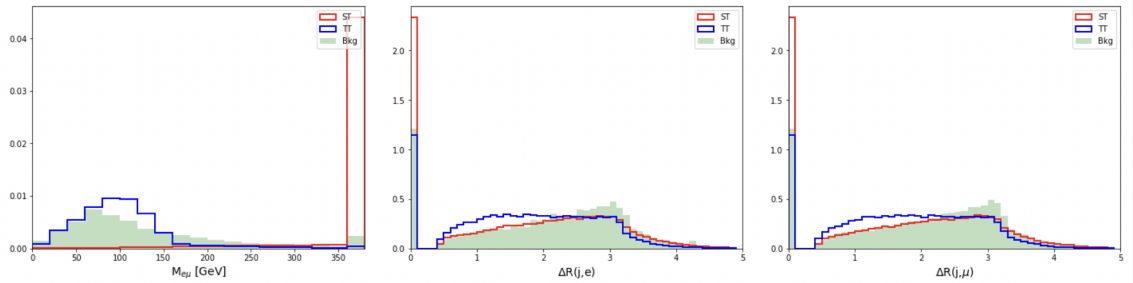
The background MC samples include all the samples listed in Table ??-??. It was determined that splitting the background into separate classes (prompt vs non-prompt backgrounds) did not increase the signal discrimination power and caused overtraining which is why in the backgrounds are combined.

The MC dataset is divided into 5 subsets evenly. A 5-fold cross-validation scheme is deployed:



**Figure 16.2:** ROC curve with 5-fold cross validation. From left to right: BDT targeting TT (SR1), BDT targeting ST (SR2).

the training set consists of 3 out of the 5 subsets, while validation/testing set each consists of 1 subset. The number of estimators is set to 1000, max depth 5, and sub-sample 0.8 with the evaluation metric set to "mlogloss". The results of the cross-validation study are shown in Figure 16.2. The trained BDT model targeting TT(ST) signal is used to evaluate each event in SR1(SR2): a BDT weight is assigned to each event, giving the likelihood that it originated from TT(ST) signal.



**Figure 16.3:** Normalized distribution of various features in SR. From to left to right: MVA\_  $M_{e\mu}$ , MVA\_  $J_{eDr}$ , MVA\_  $J_{eDr}$ .

## 16.2 Input Variables

The observables used in training are referred to as "features" in this analysis. The following features are used for both ST and TT classifier:

- **MVA\_  $M_{e\mu}$ :** invariant mass of the Opposite-Sign  $e\mu$  pair
- **MVA\_  $LFVePt$ :**  $p_T$  of the flavor-violating electron
- **MVA\_  $LFVmuPt$ :**  $p_T$  of the flavor-violating muon
- **MVA\_  $LFVTopmass$ :** invariant mass of the flavor-violating 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 b-jet+lepton system
- **MVA\_njet:** number of jets
- **MVA\_nbjet:** number of b-jets
- **MVA\_tM:** transverse mass of the W boson candidate (from standard model top-quark)
- **MVA\_lIDr:**  $\Delta R$  between flavor-violating electron and muon
- **MVA\_SSee\_Zmass:** invariant mass of a Same-Sign di-electron pair
- **MVA\_Topmass:** invariant mass of the standard model top-quark candidate
- **MVA\_Met:** missing transverse energy

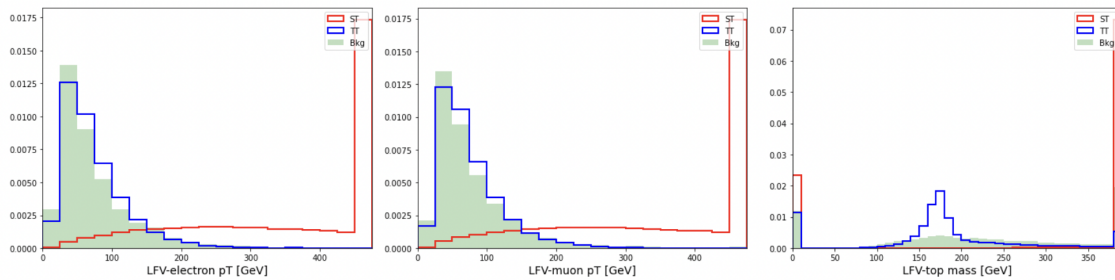
The following features are only used for TT classifier:

- **MVA\_Ht:** scalar sum of the  $p_T$  of all objects
- **MVA\_Mbl1:** invariant mass of the second bjet+lepton system
- **MVA\_JeDr:**  $\Delta R$  between flavor-violating electron and a light jet (non-b-jet)
- **MVA\_JmuDr:**  $\Delta R$  between flavor-violating muon and a light jet (non-b-jet)

The following features are used for ST classifier:

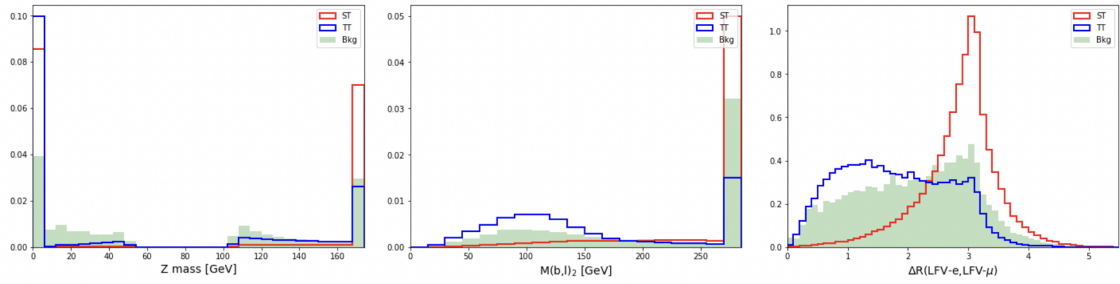
- **MVA\_BaPt:**  $p_T$  of the standard model lepton
- **MVA\_JetHt:** scalar sum of the  $p_T$  of jets

Distributions of selected features are shown in Figure 16.3-16.7.

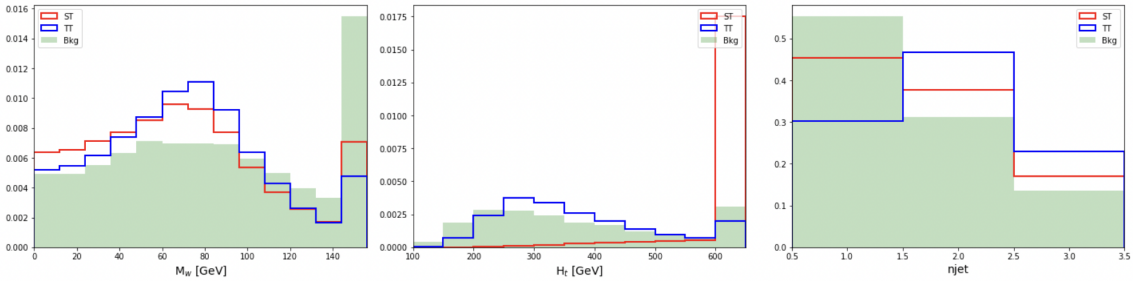


**Figure 16.4:** Normalized distribution of additional features in SR. From to left to right: MVA\_LFVePt, MVA\_LFVmuPt, MVA\_LFVTopmass.

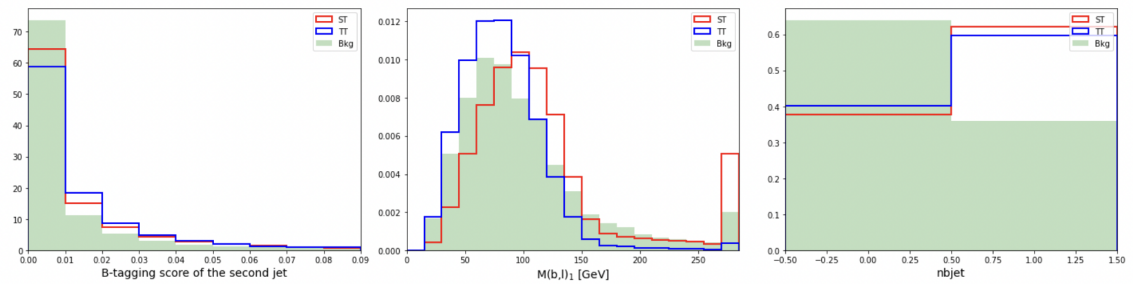




**Figure 16.5:** Normalized distribution of additional features in SR. From to left to right: MVA\_Zmass, MVA\_Mbl2, MVA\_IIDr.



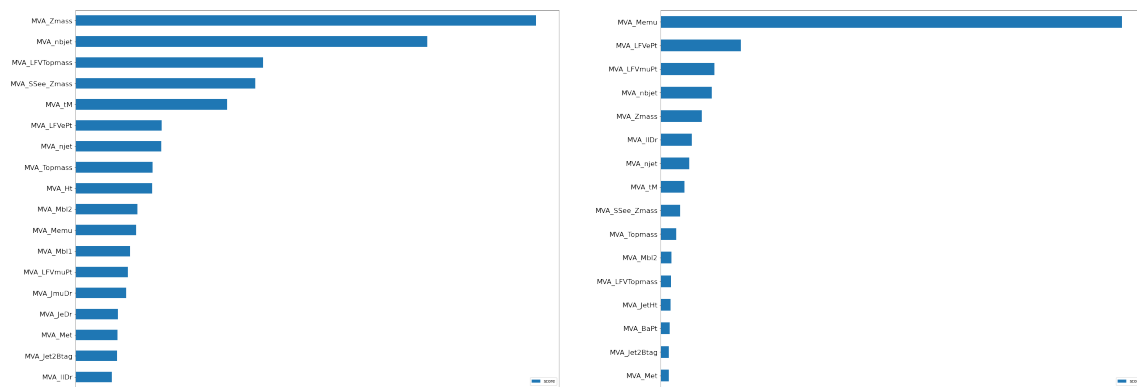
**Figure 16.6:** Normalized distribution of additional features in SR. From to left to right: MVA\_tM, MVA\_Ht, MVA\_njet.



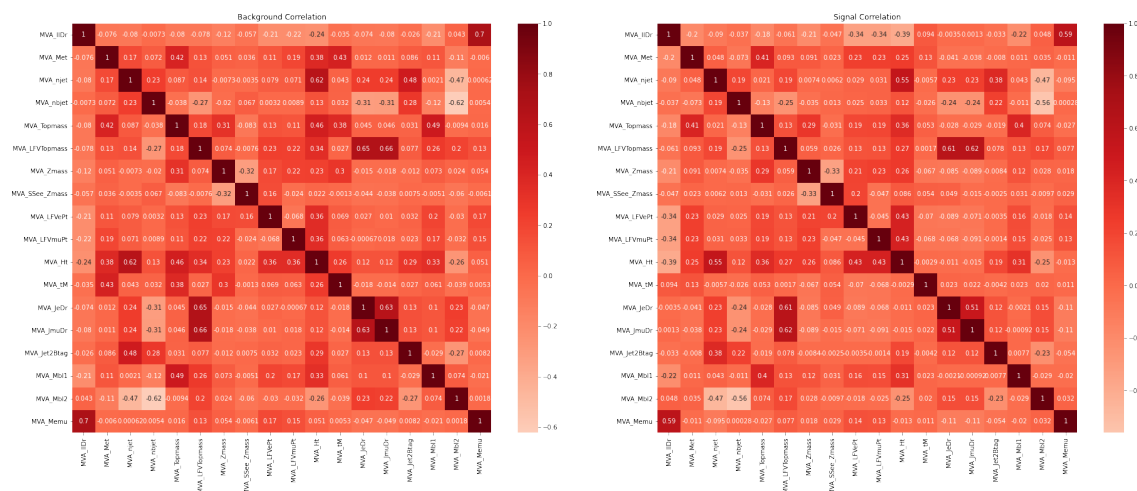
**Figure 16.7:** Normalized distribution of additional features in SR. From to left to right: MVA\_Jet2Btag, MVA\_Mbl1, MVA\_nbjet.

### 16.2.1 Input ranking and correlation

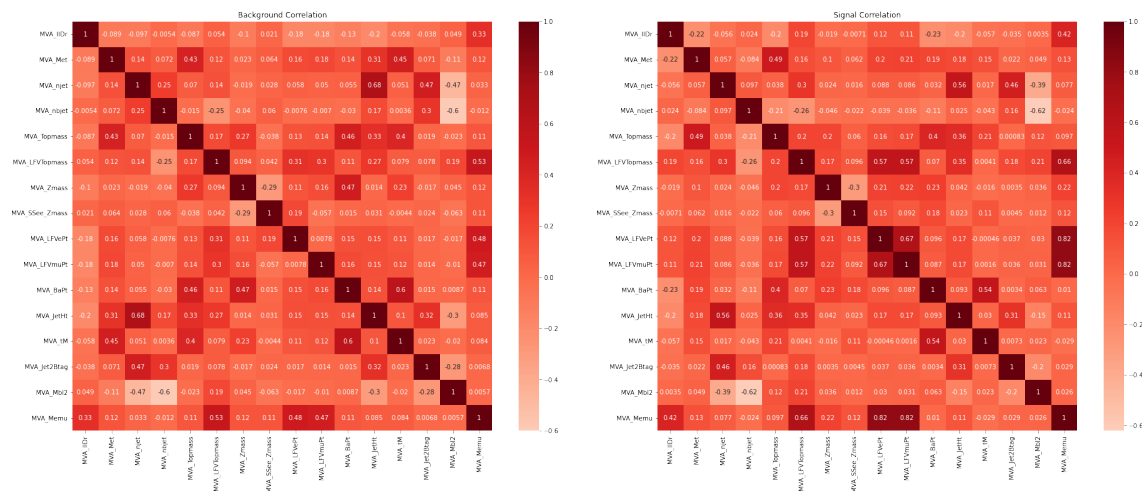
To determine the relative importance of these features, feature importance is extracted using the "gain" method and is shown in Figure 16.10.



**Figure 16.8:** List of features ranked by their relative importance. From left to right: BDT targeting TT (SR1), BDT targeting ST (SR2)



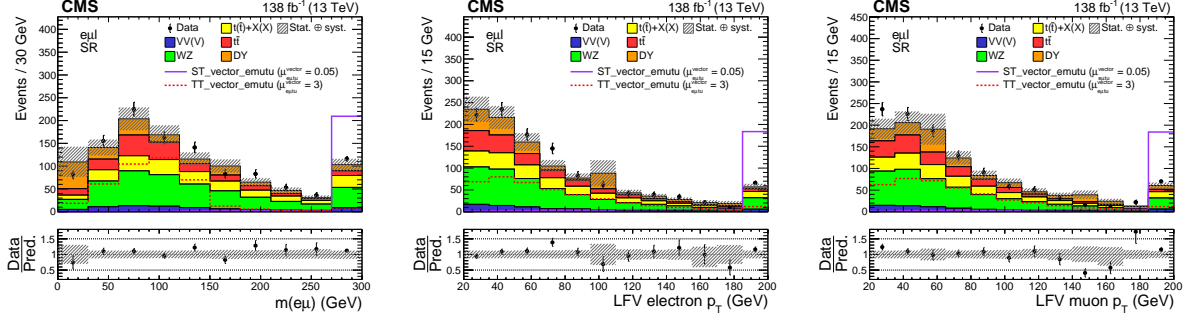
**Figure 16.9:** Correlation matrices (SR1), from left to right: background correlation, signal correlation.



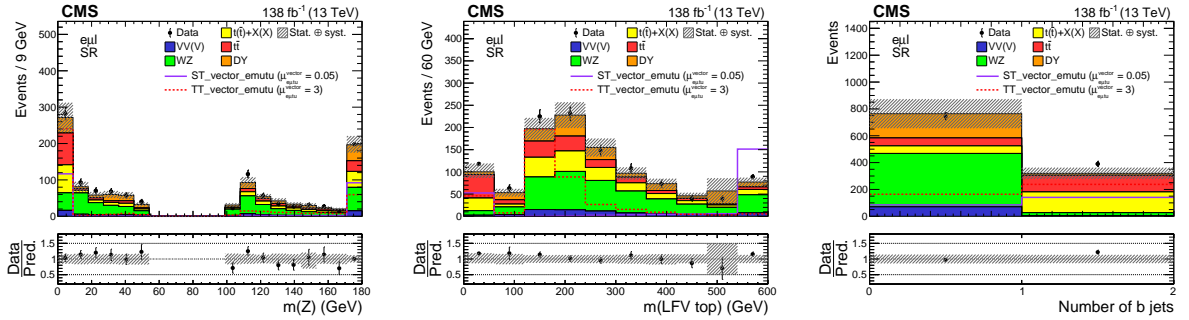
**Figure 16.10:** Correlation matrices (SR2), from left to right: background correlation, signal correlation.

### 16.2.2 Input validation

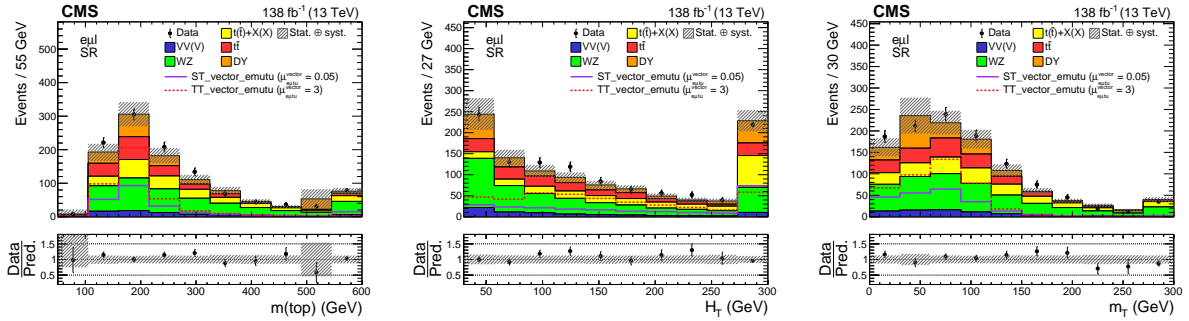
The background modeling of the input features are checked in SR. The input features are shown in Figure 16.11-16.17. Note: all input samples come from MC simulation.



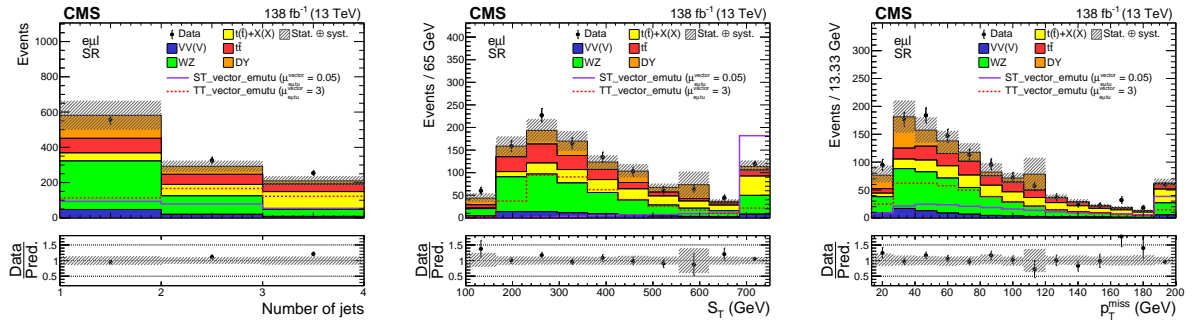
**Figure 16.11:** BDT input validation, from left to right: MVA\_ $M_{e\mu}$ , MVA\_ $LFV_{ePt}$ , MVA\_ $LFV_{\mu Pt}$



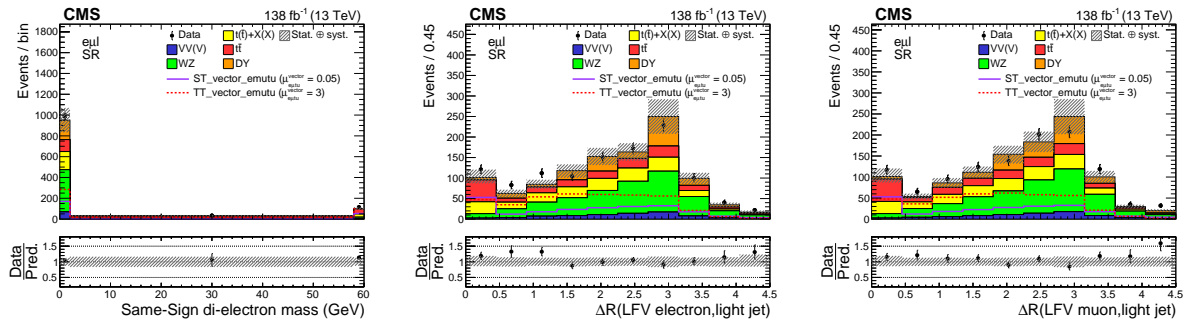
**Figure 16.12:** BDT input validation, from left to right: MVA\_ $Z_{mass}$ , MVA\_ $LFV_{Topmass}$ , MVA\_ $nbjet$



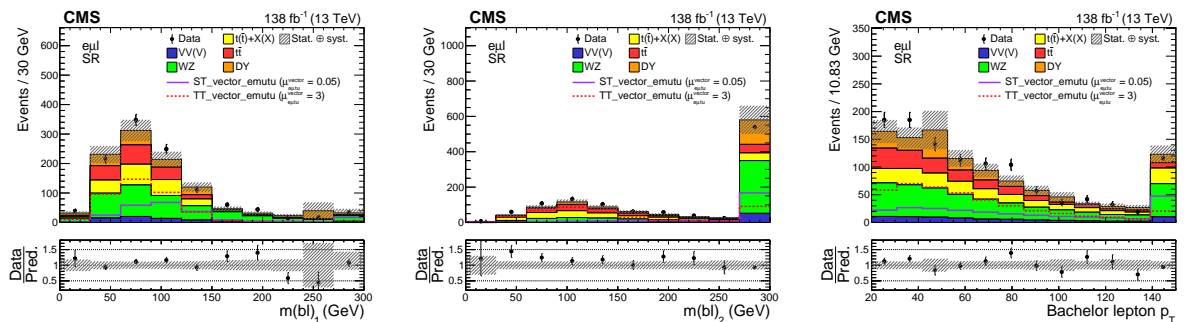
**Figure 16.13:** BDT input validation, from left to right: MVA\_ $Topmass$ , MVA\_ $JetHt$ , MVA\_ $tM$



**Figure 16.14:** BDT input validation, from left to right: MVA\_njet, MVA\_Ht, MVA\_Met



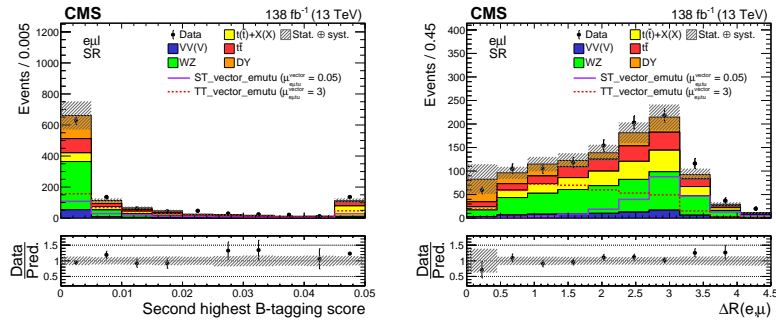
**Figure 16.15:** BDT input validation, from left to right: MVA\_SSZmass, MVA\_JeDr, MVA\_JmuDr



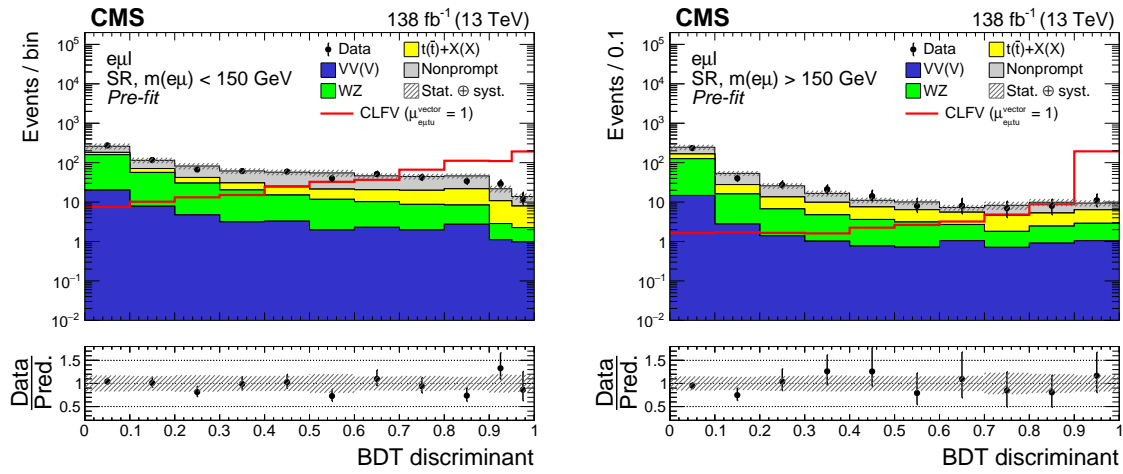
**Figure 16.16:** BDT input validation, from left to right: MVA\_Mbl1, MVA\_Mbl2, MVA\_BaPt

### 16.3 BDT Output

The output of the BDTs are shown in Figure 16.18. Note: all but the nonprompt backgrounds are estimated with MC simulation. The nonprompt background backgrounds are estimated with the matrix method.



**Figure 16.17:** BDT input validation, from left to right: MVA\_Jet2Btag, MVA\_IIDr



**Figure 16.18:** Distribution of BDT output, from left to right TT enriched SR (SR1), ST enriched SR (SR2)

# CHAPTER 17

## Systematic Uncertainties

For this analysis, we consider the following sources of systematic uncertainties:

### 17.1 Theoretical Uncertainties

- Prompt background cross-sections: These uncertainties arise from the uncertainties in the measured cross-sections of the processes which are used for scaling the MC to the luminosity of data. Systematic uncertainties associated to the prompt backgrounds WZ ( as discussed in autorefsec:CRWZ ), ZZ , ttW/ttZ/ttH, tZq are 6%, 6%, 15% and 20% respectively. Measurements performed on the WZ and ZZ processes currently have uncertainties of order 5-8% citeWZ2019 citeZZ2020. A 50 % normalization uncertainty is assigned to other smaller backgrounds.
- PDF: Using 100 the replicas of the NNPDF 3.1 set, the PDF uncertainties are obtained by taking the root mean square value of the variations. Only shape components of the variation is considered, this is done by normalizing LHE weights of the replicas such that the sum of the LHE wights (in SR) stays the same for each replica. This uncertainty treated as correlated. We consider this uncertainty for all the signals and major prompt backgrounds (i.e. WZ, TTZ, TTW and TTH).
- Renormalization and factorization scale: this uncertainty is evaluated by considering an envelope of 6 variation ( $\mu_R$  2,  $\mu_R$  0.5,  $\mu_F$  2,  $\mu_F$  0.5,  $\mu_R$  and  $\mu_F$  2,  $\mu_R$  and  $\mu_F$  0.5). This uncertainty is treated as correlated and only shape component is considered. We consider this uncertainty for all the signals and major prompt backgrounds (i.e. WZ, TTZ, TTW and TTH).
- ISR and FSR: initial state radiation and final state radiation uncertainty is evaluated by using the default set of variation (2 and 0.5). This uncertainty is treated as correlated and only shape component is considered. We consider this uncertainty for all the signals.

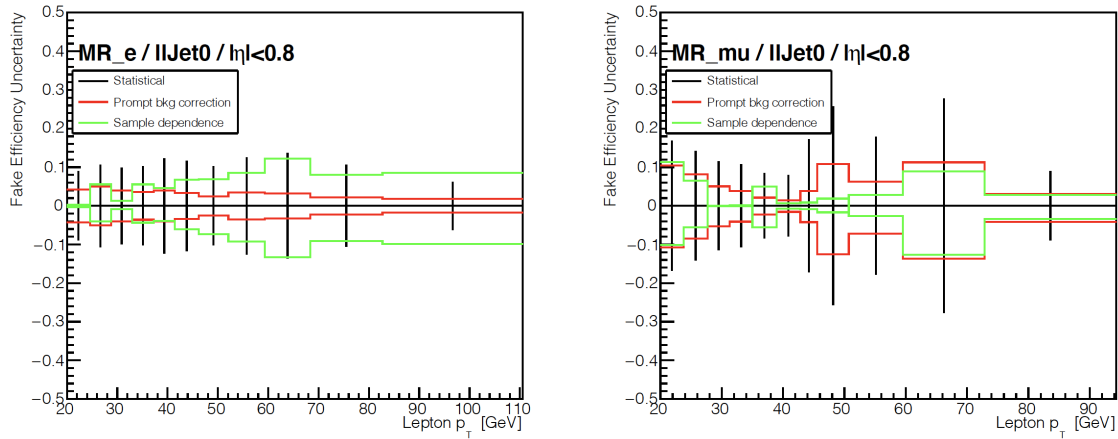
## 17.2 Nonprompt Uncertainties

We consider the following sources of systematic uncertainties (shape) associated with the *fake* efficiency.

- Prompt background correction
- Sample dependence
- Statistical uncertainty

An important systematic uncertainty comes from the estimate of prompt contamination in MR. As is discussed in ??1.3, prompt backgrounds (estimated with MC) are subtracted from total event yields measured in MR (data). A flat 20 % uncertainty ( $\alpha$  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, which is estimated to be between 5 % and 15 %.

$$f = \frac{n_{data}^T - (1 + \alpha)n_{Prompt MC}^T}{n_{data}^{T+\bar{T}} - (1 + \alpha)n_{Prompt MC}^{T+\bar{T}}}. \quad (17.1)$$



**Figure 17.1:** Comparison of different components of the uncertainties associated to the *fake* efficiency measured in 2017 dataset ( $n_{jet}=0$  bin,  $|\eta| < 0.8$  bin). From left to right: electron  $f$  uncertainty, muon  $f$  uncertainty. Note: black uncertainty bars represent statistical uncertainties.

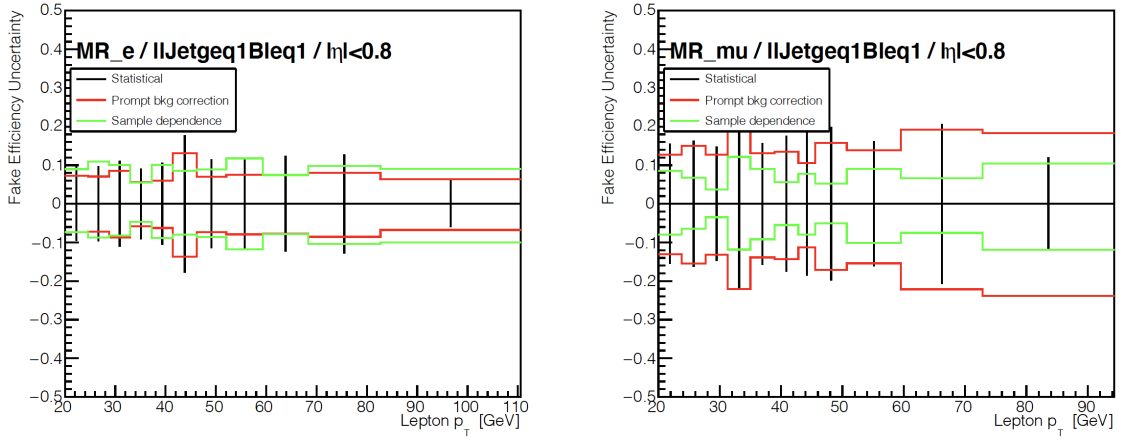
Sample dependence is one of the primary sources of systematic uncertainty and it is concerned with the possibility that the *fake* efficiency  $f$  measured in MR may not reflect characteristics of the non-prompt leptons in SR due to different compositions of backgrounds. This type of uncertainty is estimated by introducing a variation factor  $\beta$  between the proportions of Same-Flavor and Opposite-Flavor pairs in MR. For example, electron  $f$  can be calculated as (prompt background correction is ignored from the equation):



$$f_e = \frac{(1 + \beta)n_{F,e+e}^T + (1 - \beta)n_{F,e+\mu}^T}{(1 + \beta)n_{F,e+e}^{T+\bar{T}} + (1 - \beta)n_{F,e+\mu}^{T+\bar{T}}}. \quad (17.2)$$

A 20% variation ( $\beta$ ) is assigned the resulting variation of  $f$  is taken as the uncertainty, which is estimated to be between 3 % and 15 %.

A comparison of different sources of systematic uncertainties (using 2017 datasets) are shown in Figure 17.1 and Figure 17.2. A complete summary of uncertainties associated to *real* and *fake* efficiency can be found in autorefsec:RandF.

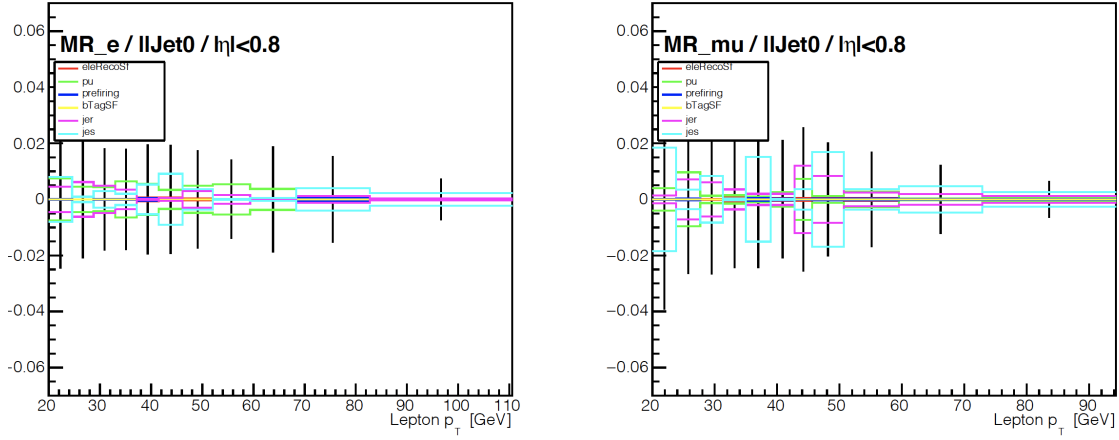


**Figure 17.2:** Comparison of different components of the uncertainties associated to the *fake* efficiency measured in 2017 dataset ( $n_{\text{jet}} > 0$  bin,  $|\eta| < 0.8$  bin). From left to right: electron  $f$  uncertainty, muon  $f$  uncertainty. Note: black uncertainty bars represent statistical uncertainties.

In addition to statistical uncertainties, MC uncertainties are propagated to the *real* efficiency as sources of systematic uncertainties:

- Statistical uncertainty
- MC uncertainty
  - Electron reconstruction
  - b-tagging
  - JES/JER
  - Pile-up reweighting
  - ECAL prefiring

The uncertainties associated to the *real* efficiency are relatively small when compared to the *fake* efficiency uncertainties. A comparison of different sources of *real* efficiency uncertainties are shown in Figure 17.3.



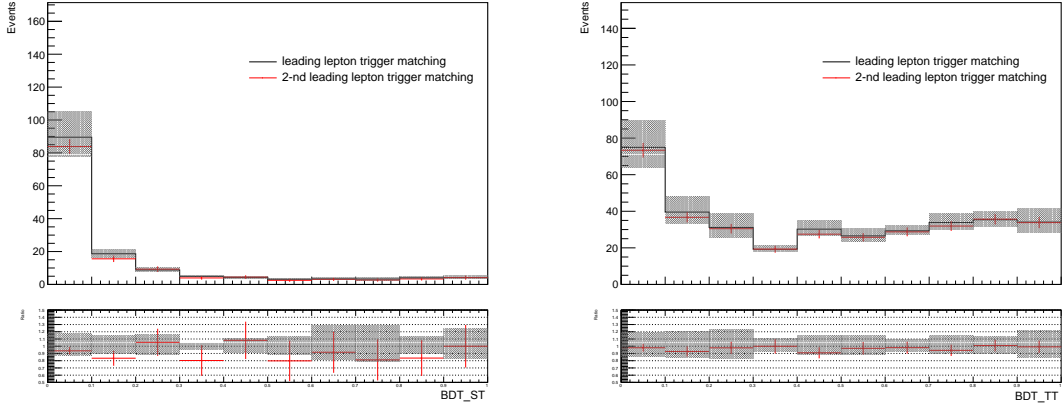
**Figure 17.3:** Comparison of different components of the uncertainties associated to the *real* efficiency measured in 2017 dataset ( $n_{\text{jets}}=0$  bin,  $|\eta| < 0.8$  bin). From left to right: electron  $r$  uncertainty, muon  $r$  uncertainty. Note: black uncertainty bars represent statistical uncertainties.

In addition to the four components (i.e.  $r_e$ ,  $r_\mu$ ,  $f_e$ ,  $f_\mu$ ) of the shape uncertainties associated to the measurement of  $r$  and  $f$ , a fifth component is considered that accounts for the potential bias introduced when implementing 3D matrix method: Four out of the eight regions that appear on the lefthand side of the Equation ?? (i.e.  $N^{\bar{T}TTT}$ ,  $N^{\bar{T}T\bar{T}\bar{T}}$ ,  $N^{\bar{T}\bar{T}TT}$ ,  $N^{\bar{T}\bar{T}\bar{T}\bar{T}}$ ) are selected by requiring the leading lepton to fail the *tight* criteria (See Table 15.5). Effectively this means that the isolation requirement is reversed for leading lepton that enter these four regions. Selecting the leading lepton by a loose/reversed isolation requirement is not ideal since the leading lepton is required to match with iso-trigger. To account for this bias, a 50 % uncertainty is assigned to the  $f_1$  (*fake* efficiency of the leading lepton) for events that enter these four regions. The variation of the non-prompt estimate due to trigger matching is largely covered by this uncertainty (See Figure 17.4).

The five components of the uncertainties discussed in this section are propagated through the matrix inversion. The resulting variations of the non-prompt estimates are taken as the shape uncertainties assigned to non-prompt backgrounds. These uncertainties are treated as correlated between the years. In addition to these shape uncertainties, an overall normalization uncertainty of 10% is assigned in order to cover any other potential bias.

### 17.3 Diboson Uncertainties

Mismodeling of the jet multiplicity is observed in WZ control region (See autorefsec:CRWZ). This is largely due to the fact that jets generated from diboson samples are coming from radiation (due to our trilepton selection). Therefore, the accuracy of the jet multiplicity distribution is suboptimal. To take this into account, a dedicated jet-dependent uncertainty is assigned to each



**Figure 17.4:** The impact of matching leptons to trigger objects on non-prompt estimate. From left to right: non-prompt estimate in ST enriched SR, non-prompt estimate in TT enriched SR. The nominal configuration of the matrix method is to match the leading lepton with trigger objects. Matching the 2nd-leading with the trigger objects is taken as an alternative to evaluate the robustness of the non-prompt estimate. The uncertainty band only covers the variation of the non-prompt estimate as a result of varying leading lepton  $f$  by 50 %. Uncertainty bars only include statistical uncertainties.

event. This uncertainty is determined using diboson control region:

- Exactly three leptons (any flavor composite),
- OnZ,
- $\text{MET} > 85\text{GeV}$ ,
- Veto events with at least one b-tag jet ( $p_T > 20\text{GeV}$ ).

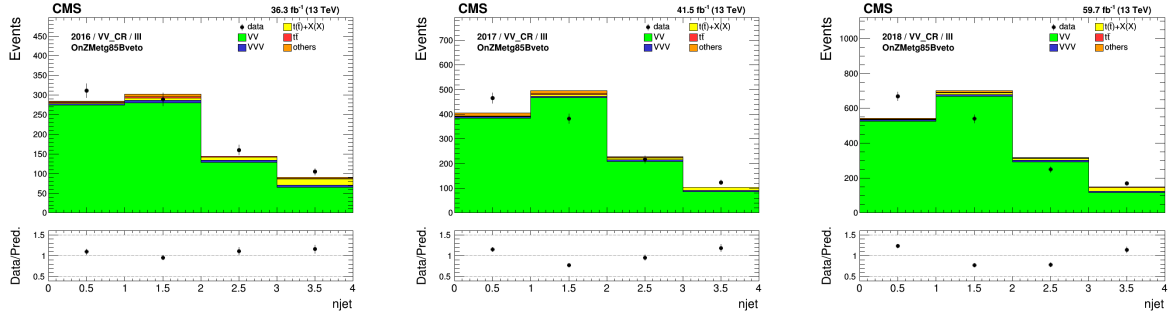
The jet multiplicity distributions are shown in Figure 17.5. For each year, a scale factor parameterized as bins of jet multiplicity is derived:

$$\epsilon = \frac{N_{data} - N_{VVV} - N_{TX} - N_{TTbar} - N_{others}}{N_{VV}}. \quad (17.3)$$

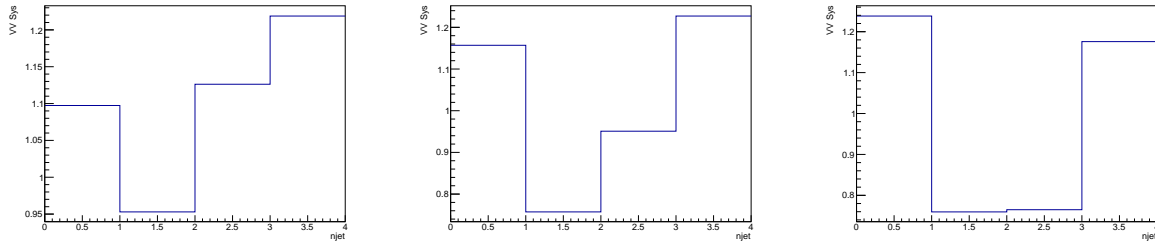
The scale factor  $\epsilon$  is used to estimate the uncertainty, denoted by  $\Delta$ :

$$\Delta = |1 - \epsilon| \quad (17.4)$$

This uncertainty is about 10-20%, as is shown in Figure 17.6.



**Figure 17.5:** Diboson control region, from left to right: 2016, 2017 and 2018 datasets.



**Figure 17.6:** Scale factors derived from diboson control region, from left to right: 2016, 2017 and 2018 datasets.

## 17.4 Other Experimental Uncertainties

- Lepton reconstruction, identification, and isolation: electron/muon reconstruction uncertainties are provided by relevant POG. Uncertainties associated with lepton identification and isolation scale factors were evaluated by the authors of the mvaTOP ID citemvaTOP. The statistical components of these uncertainties are treated as uncorrelated while the other components are treated as fully correlated. For high  $p_T$  muons ( $p_T > 200\text{GeV}$ ), an additional uncertainty (denoted "muIDHighPt") is assigned and it increases linearly from 0 to 10% (200GeV-1000GeV) and is capped at 10% after 1000GeV.
- Muon scale uncertainties: uncertainties on muon momentum (denoted "MuonScale") are assigned using different method according to the muon  $p_T$ : Rochester algorithm is used for low  $p_T$  muons ( $p_T < 200\text{GeV}$ ), while GEciteGEmethod method is used for high  $p_T$  muons ( $p_T > 200\text{GeV}$ ).
- b-tagging: b-tagging efficiency and uncertainty is provided by the BTV group.
- trigger scale factor: trigger scale factors are set to 1 and a flat 2% uncertainty is assigned. This uncertainty is treated as uncorrelated.

- Luminosity: The uncertainties of 1.0%, 2.0% and 1.5% are assigned to the integrated luminosity for 2016, 2017, 2018, respectively. When processing the run 2 data the individual year uncertainties are treated as uncorrelated while there are two additional correlated luminosity uncertainties; firstly, 0.6%, 0.9% and 2.0% for each year respectively to account for the correlation between 2016, 2017 and 2018 data and secondly, 0.6%, 0.2% for 2017 and 2018 respectively to account for the correlation between 2017 and 2018 data.
- jet energy scale and resolution: The jet energy scale, JES, and jet energy resolution, JER, are centrally provided by the JET/MET POG [citeJEC](#). Variations of jet energy scale and resolution are propagated to the MET and b-tagging SF.
- pile-up reweighting: The measured minimum bias cross-section is varied up and down by 4.6%. This uncertainty is treated as correlated.
- MET unclustered missing energy is considered and treated as uncorrelated across the years.
- L1 ECAL prefiring: In the 2016 and 2017 datasets, L1 EGamma triggers fired early 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 tool [citeECALPre](#) provided by the L1 DPG is used to reweight MC events. This uncertainty is treated as correlated.
- HEM15/16 Issue: The HEM15/16 issue refers to two HCAL modules whose power supply died in the middle of the data taking (runs  $\geq 319077$ , i.e. last certified run of 2018B, and all of 2018C+D). The HEM issue is likely a very small effect but we still have to check it following the procedure in [citeHEM](#)

Uncertainties related to B-tagging scale factors are split into different sources. For b and udsg jets, we applied  $l_f$ ,  $h_f$ ,  $hfstats1/2$ , and  $lfstats1/2$  uncertainties. For c jets, we applied  $cferr1/2$  uncertainties. Correlations between different sources are specified in the Table [??](#).

**Table 17.1:** A hyphen (—) denotes that a source is not correlated between the different years.

| Source     | Correlated | Description                         |
|------------|------------|-------------------------------------|
| $l_f$      | ✓          | udsg+c jets in heavy flavor region  |
| $h_f$      | ✓          | 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    |

### 17.4.1 Jet energy scale uncertainties

Uncertainties associated with JES are split into their 27 components and properly treated the correlations of the split uncertainties by year as recommended by the JETMET POG and described in the Table 17.2.

B-tagging scale factors and MET vector are computed for each of the JES templates and treated as uncertainties that are fully correlated to the respective JES sources.

The overall systematic uncertainty on background is estimated to be about 15 % in SR.

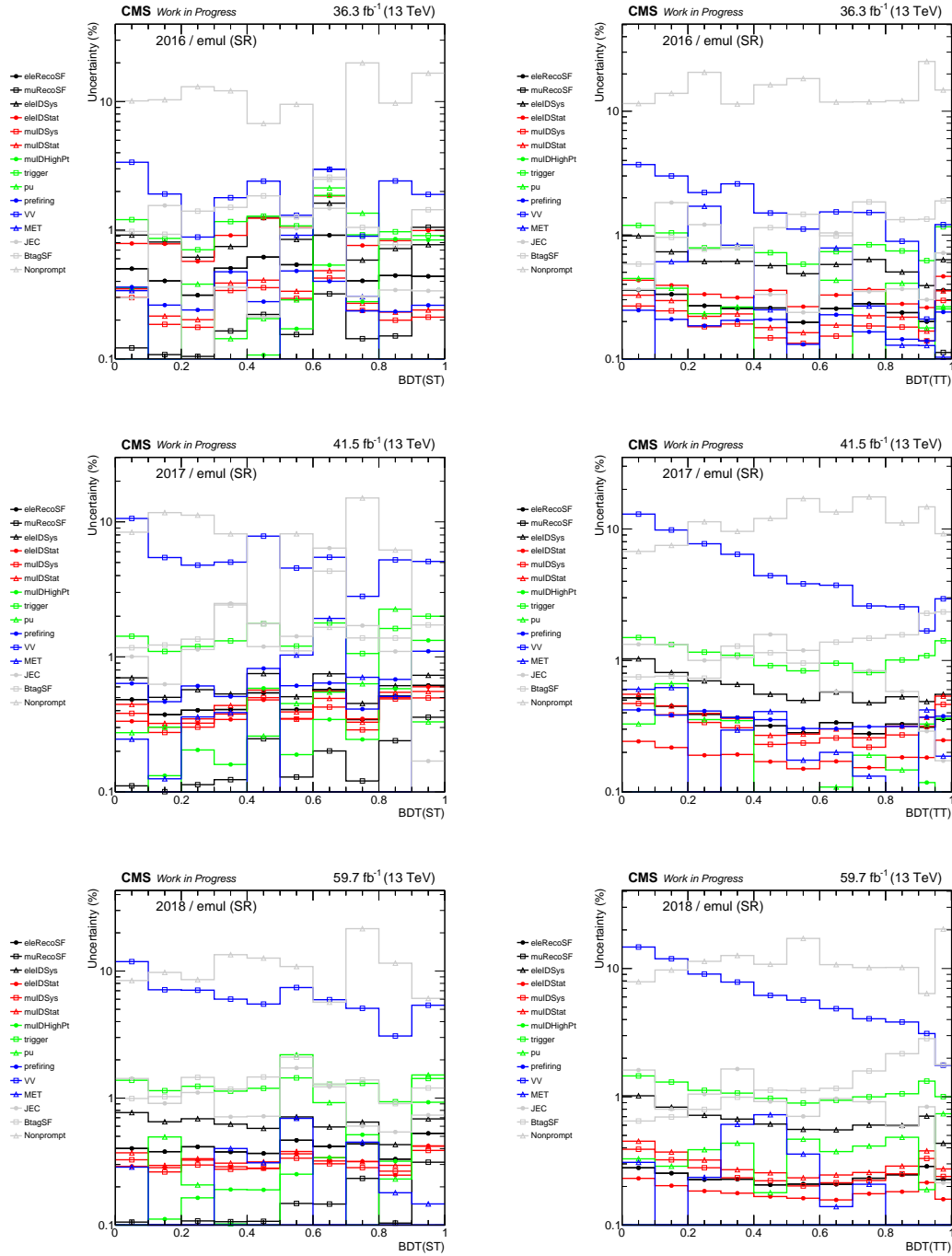
A comparison of different sources of systematic uncertainties of the background estimates in SR is shown in Figure 17.7.

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

Representative range of systematic uncertainties on background and signal MC samples are summarized in citeSysError. These uncertainties are extracted from the last bins of the BDT distribution.

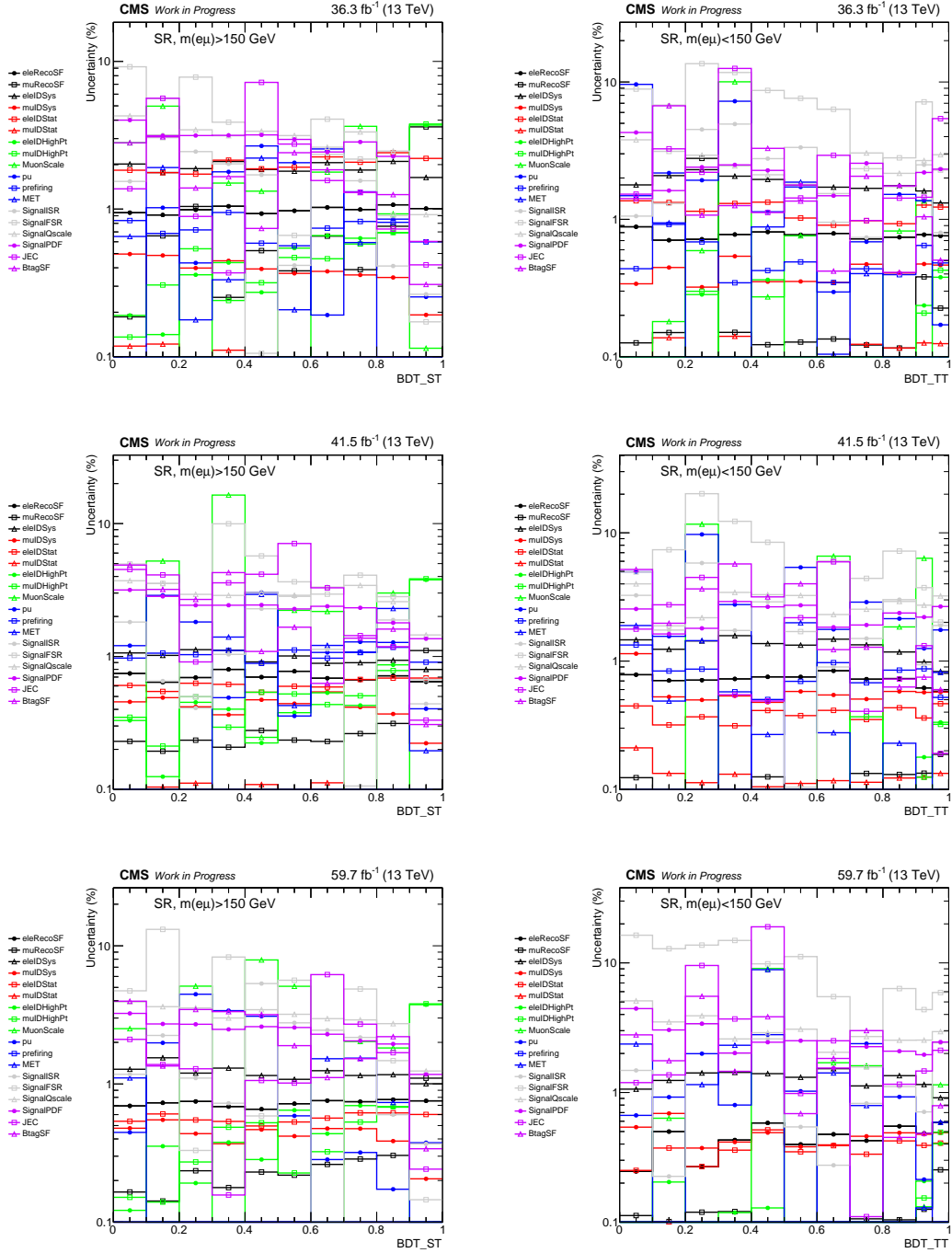
**Table 17.2:** Summary of the sources of uncertainty related to the JECs. A hyphen (–) denotes that a source is not correlated between the different years.

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



**Figure 17.7:** Distributions of relative uncertainties on total expected backgrounds as a function of BDT output in ST enriched SR (left), TT enriched SR (right). Luminosity and normalization uncertainties are not included in these plots. JES, JER and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table ?? are combined into "BtagSF". From top to bottom: 2016, 2017 and 2018 datasets.





**Figure 17.8:** Distributions of relative uncertainties on signal ( $C_{e\mu\tau\tau}^{vector}$  is used as an example) as a function of BDT output in ST enriched SR (left), TT enriched SR (right). Luminosity and normalization uncertainties are not included in these plots. JES, JER and HEM are combined into "JEC". Sources of b-tagging uncertainties listed in Table ?? are combined into "BtagSF". From top to bottom: 2016, 2017 and 2018 datasets.

# CHAPTER 18

## Statistical Analysis

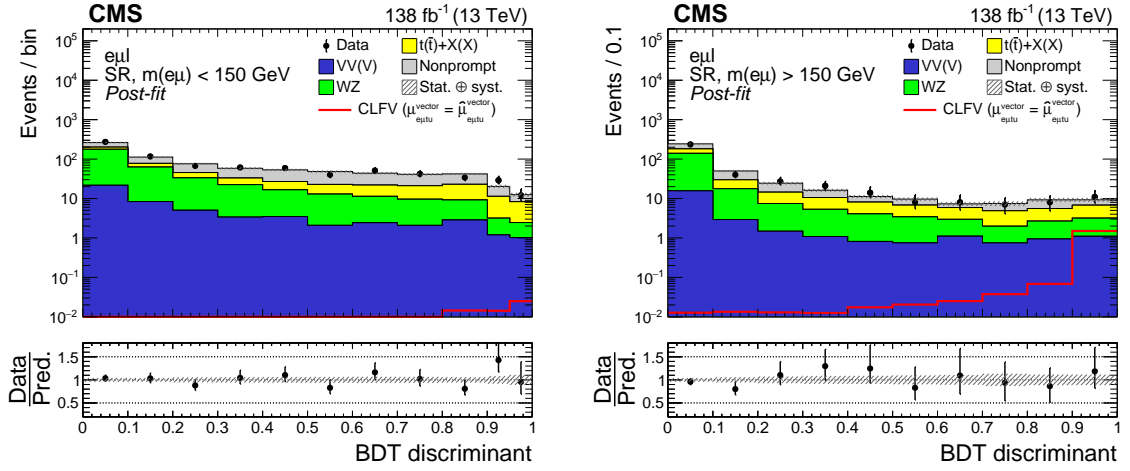
### 18.1 Profile Likelihood Fit

A binned likelihood function  $\mathcal{L}(\mu, \theta)$  is constructed to perform the statistical analysis on the BDT discriminator distributions. The parameter of interest (POI), denoted by  $\mu$ , is the signal strength that governs the cross section of the ST and TT signals simultaneously. All the uncertainties are incorporated into the likelihood function as nuisance parameters, denoted by  $\theta$ . The uncertainties that affect the shape of the BDT discriminator distributions are considered with Gaussian distributions while other uncertainties that only affect the normalizations are considered with log-normal distributions.

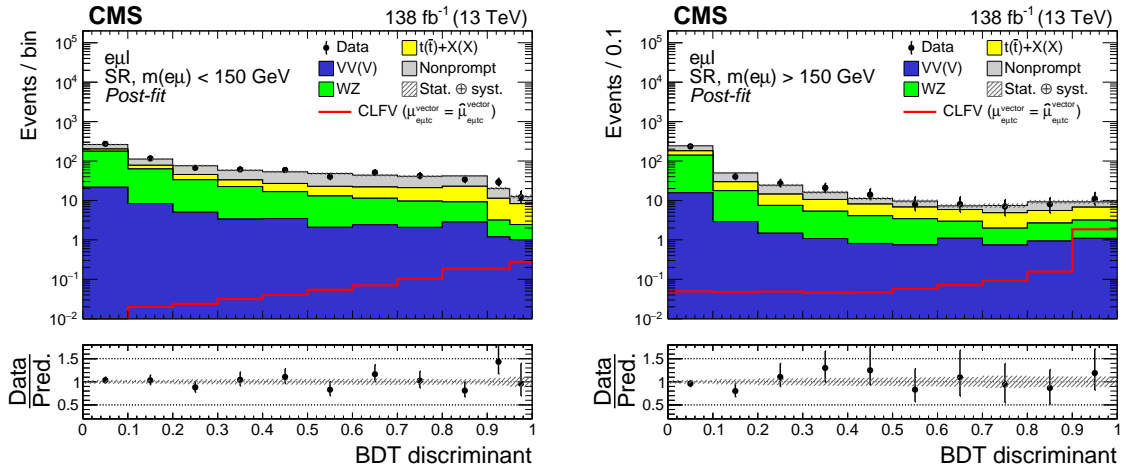
To control the systematic uncertainties, a profile likelihood fit is performed simultaneously in six regions (three data-taking years and two signal regions) by maximizing the likelihood function  $\mathcal{L}(\hat{\mu}, \hat{\theta})$ . Both  $\hat{\mu}$  and  $\hat{\theta}$  are the maximum likelihood estimators for the signal strength and nuisance parameters, respectively. The scaling of the signal for the purpose of a better visualization is optimized based on "VecU" signal, and it is kept the same for the other signals to compare their cross sections relative to each other. The post-fit distributions of the BDT discriminator are shown in Fig. 18.1-18.6. The most prominent uncertainties affecting the likelihood fit are the statistical uncertainties on MC samples.

### 18.2 Upper Limits

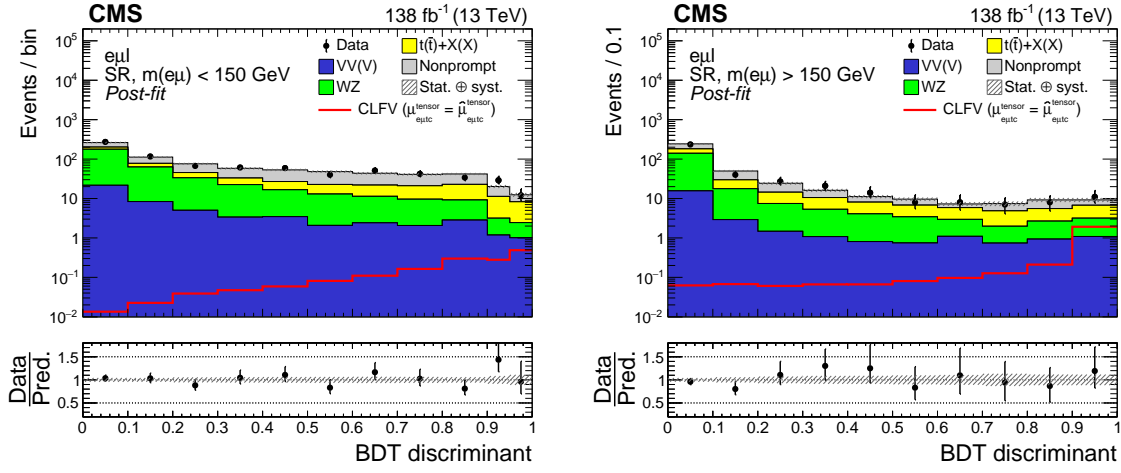
The results for the one-dimensional limits are summarized in Table 18.1. Assuming a linear relationship between  $\mathcal{B}(t \rightarrow e\mu u)$  and  $\mathcal{B}(t \rightarrow e\mu c)$  in the case of nonvanishing signals, the two-dimensional limits can also be obtained through interpolation (see Fig. 18.7). This analysis constitutes the most stringent limits on these processes to date.



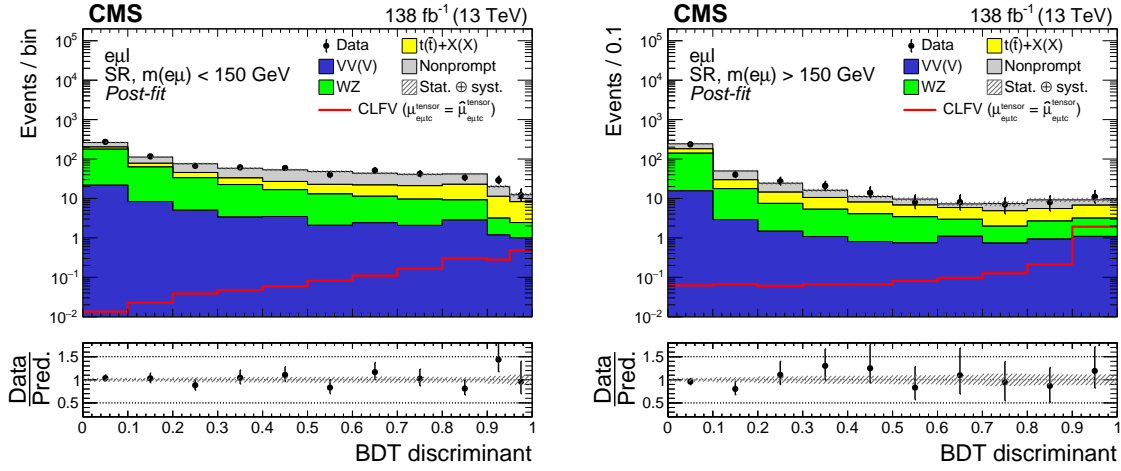
**Figure 18.1:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with vector-like operator involving an Up quark. The three data-taking years are aggregated for better visualization.



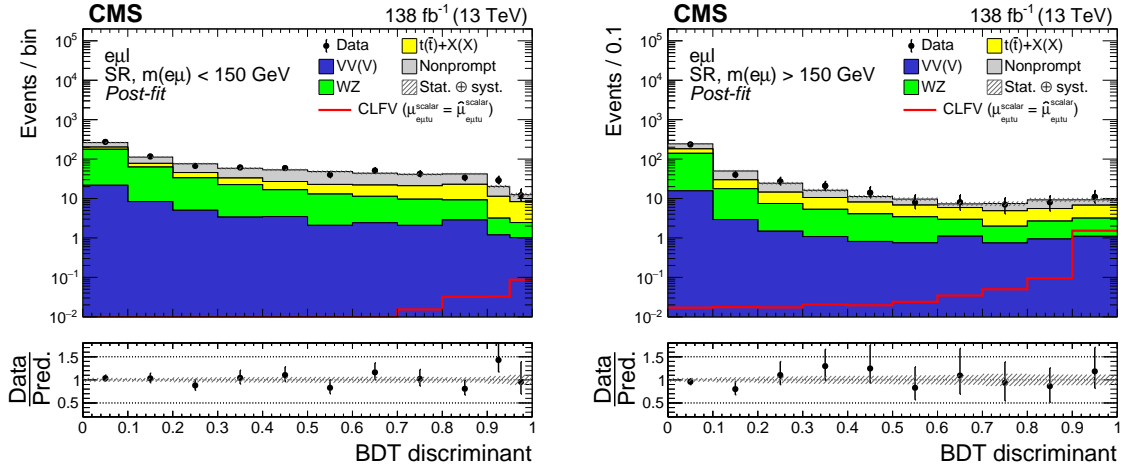
**Figure 18.2:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with vector-like operator involving a Charm quark.



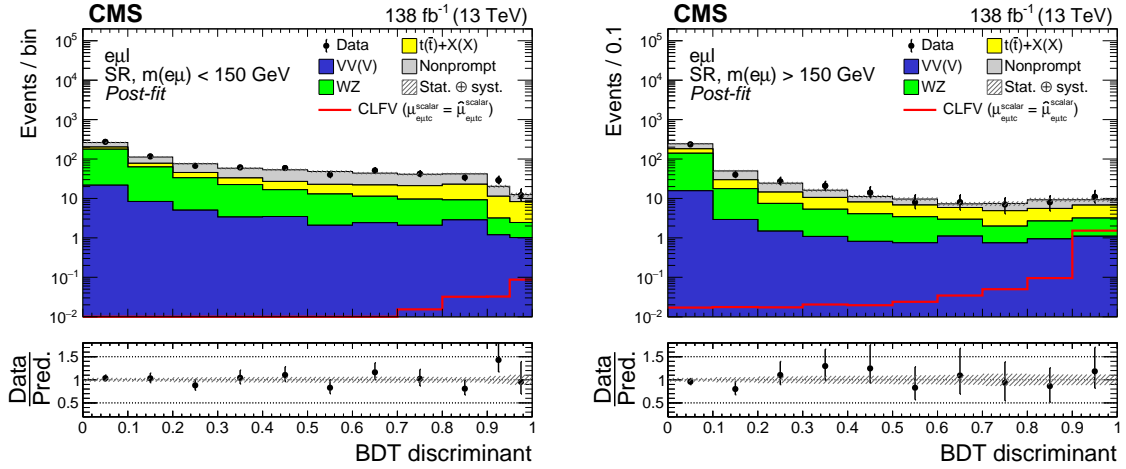
**Figure 18.3:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with tensor-like operator involving an Up quark.



**Figure 18.4:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with tensor-like operator involving a Charm quark.



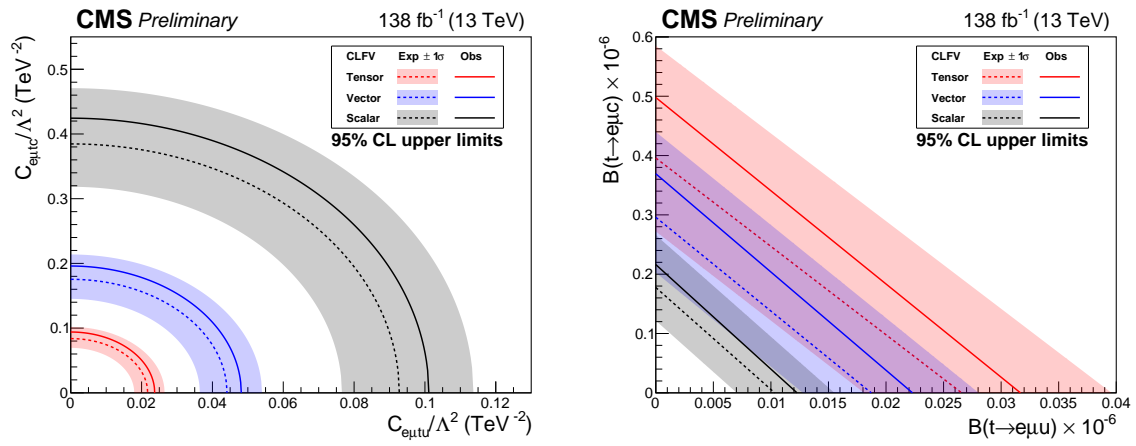
**Figure 18.5:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with scalar-like operator involving an Up quark.



**Figure 18.6:** Distributions of the posterior BDT discriminator distributions for the TT-enriched SR (left) and the ST-enriched SR (right). Signals are generated with scalar-like operator involving a Charm quark.

| LFV coupling | Lorentz structure | $\mu_{e\mu tq}$           |                | $C_{e\mu tq}/\Lambda^2 (TeV^{-2})$ |              | $B(t \rightarrow \mu^\pm e^\mp q) \times 10^{-6}$ |              |
|--------------|-------------------|---------------------------|----------------|------------------------------------|--------------|---|--------------|
|              |                   | exp (68% range)           | obs            | exp (68% range)                    | obs          | exp (68% range)                                   | obs          |
| $e\mu tu$    | tensor            | 0.00047 (0.00032–0.0007)  | <b>0.00056</b> | 0.022 (0.018–0.026)                | <b>0.024</b> | 0.027 (0.018–0.04)                                | <b>0.032</b> |
|              | vector            | 0.00195 (0.00133–0.00291) | <b>0.00232</b> | 0.044 (0.036–0.054)                | <b>0.048</b> | 0.019 (0.013–0.028)                               | <b>0.022</b> |
|              | scalar            | 0.00861 (0.00586–0.01293) | <b>0.01022</b> | 0.093 (0.077–0.114)                | <b>0.101</b> | 0.01 (0.007–0.016)                                | <b>0.012</b> |
| $e\mu tc$    | tensor            | 0.00702 (0.00483–0.01038) | <b>0.00883</b> | 0.084 (0.069–0.102)                | <b>0.094</b> | 0.396 (0.272–0.585)                               | <b>0.498</b> |
|              | vector            | 0.03076 (0.02107–0.04572) | <b>0.03839</b> | 0.175 (0.145–0.214)                | <b>0.196</b> | 0.296 (0.203–0.44)                                | <b>0.369</b> |
|              | scalar            | 0.14795 (0.10135–0.22167) | <b>0.18013</b> | 0.385 (0.318–0.471)                | <b>0.424</b> | 0.178 (0.122–0.266)                               | <b>0.216</b> |

**Table 18.1:** Upper limit on the LFV signal using the full Run 2 data set.



**Figure 18.7:** Two-dimensional upper limits on the Wilson coefficients (left) and the branching ratios (right).

## **Part IV**

# **Outlook for CLFV Searches Using Top Quarks**

# CHAPTER 19

## **Inclusive Signal Generation**

### **19.1 UFO Model**

### **19.2 EFT Reweighting**



# CHAPTER 20

## **Object and Event Selection**

### **20.1 Object Selection**

### **20.2 Event Selection**

# CHAPTER 21

## **Fake Tau Estimate**

### **21.1 ABCD Method**

### **21.2 Validation**

# CHAPTER 22

## **Expected Sensitivity**

### **22.1 Asimov Fit**

### **22.2 Expected Upper Limits**

# CHAPTER 23

## **Run-2 Combination**

### **23.1 Correlation Scheme**

### **23.2 Minimal Flavor Violation Interpretation**

# CHAPTER 24

## Conclusion

## References

- [1] J. M. Campbell, R. K. Ellis, and C. Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011. URL: [https://link.springer.com/article/10.1007/JHEP07\(2011\)018](https://link.springer.com/article/10.1007/JHEP07(2011)018). (Cited on page 18.)
- [2] M. Czakon and A. Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014. URL: <https://www.sciencedirect.com/science/article/pii/S0010465514002264>. (Cited on page 18.)
- [3] R. Frederix and I. Tsinikos. On improving NLO merging for  $t\bar{t}W$  production. *JHEP*, 11:029, 2021. URL: [https://link.springer.com/article/10.1007/JHEP11\(2021\)029](https://link.springer.com/article/10.1007/JHEP11(2021)029). (Cited on page 18.)
- [4] A. Kulesza, L. Motyka, D. Schwartländer, T. Stebel, and V. 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. URL: <https://link.springer.com/article/10.1140/epjc/s10052-020-7987-6>. (Cited on page 18.)
- [5] Y. Li and F. Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev. D*, 86:094034, 2012. URL: <https://journals.aps.org/prd/abstract/10.1103/PhysRevD.86.094034>. (Cited on page 18.)

# 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_CaloldVT_GsfTrkIdT                     |
|                | HLT_Ele115_CaloldVT_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_CaloldL_TrackIdL_IsoVL_DZ         |
|                | HLT_Ele16_Ele12_Ele8_CaloldL_TrackIdL             |
|                | HLT_DoubleEle33_CaloldL_MW                        |
|                | HLT_DoubleEle33_CaloldL_GsfTrkIdVL                |
|                | HLT_DoubleEle33_CaloldL_GsfTrkIdVL_MW             |
| MuonEG         | HLT_Mu23_TrkIsoVVL_Ele8_CaloldL_TrackIdL_IsoVL    |
|                | HLT_Mu23_TrkIsoVVL_Ele8_CaloldL_TrackIdL_IsoVL_DZ |
|                | HLT_Mu8_TrkIsoVVL_Ele23_CaloldL_TrackIdL_IsoVL    |
|                | HLT_Mu30_Ele30_CaloldL_GsfTrkIdVL                 |
|                | HLT_Mu33_Ele33_CaloldL_GsfTrkIdVL                 |
|                | HLT_DiMu9_Ele9_CaloldL_TrackIdL                   |
|                | HLT_Mu8_DiEle12_CaloldL_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_CaloldVT_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_CaloldL_TrackIdL_IsoVL             |
|                | HLT_Ele16_Ele12_Ele8_CaloldL_TrackIdL              |
|                | HLT_DoubleEle25_CaloldL_MW                         |
|                | HLT_DoubleEle33_CaloldL_MW                         |
|                | HLT_DiEle27_WPTightCaloOnly_L1DoubleEG             |
| MuonEG         | HLT_Mu23_TrkIsoVVL_Ele12_CaloldL_TrackIdL_IsoVL    |
|                | HLT_Mu23_TrkIsoVVL_Ele12_CaloldL_TrackIdL_IsoVL_DZ |
|                | HLT_Mu8_TrkIsoVVL_Ele23_CaloldL_TrackIdL_IsoVL_DZ  |
|                | HLT_Mu12_TrkIsoVVL_Ele23_CaloldL_TrackIdL_IsoVL_DZ |
|                | HLT_Mu27_Ele37_CaloldL_MV                          |
|                | HLT_Mu37_Ele27_CaloldL_MV                          |
|                | HLT_DiMu9_Ele9_CaloldL_TrackIdL                    |
|                | HLT_DiMu9_Ele9_CaloldL_TrackIdL_DZ                 |
|                | HLT_Mu8_DiEle12_CaloldL_TrackIdL                   |



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

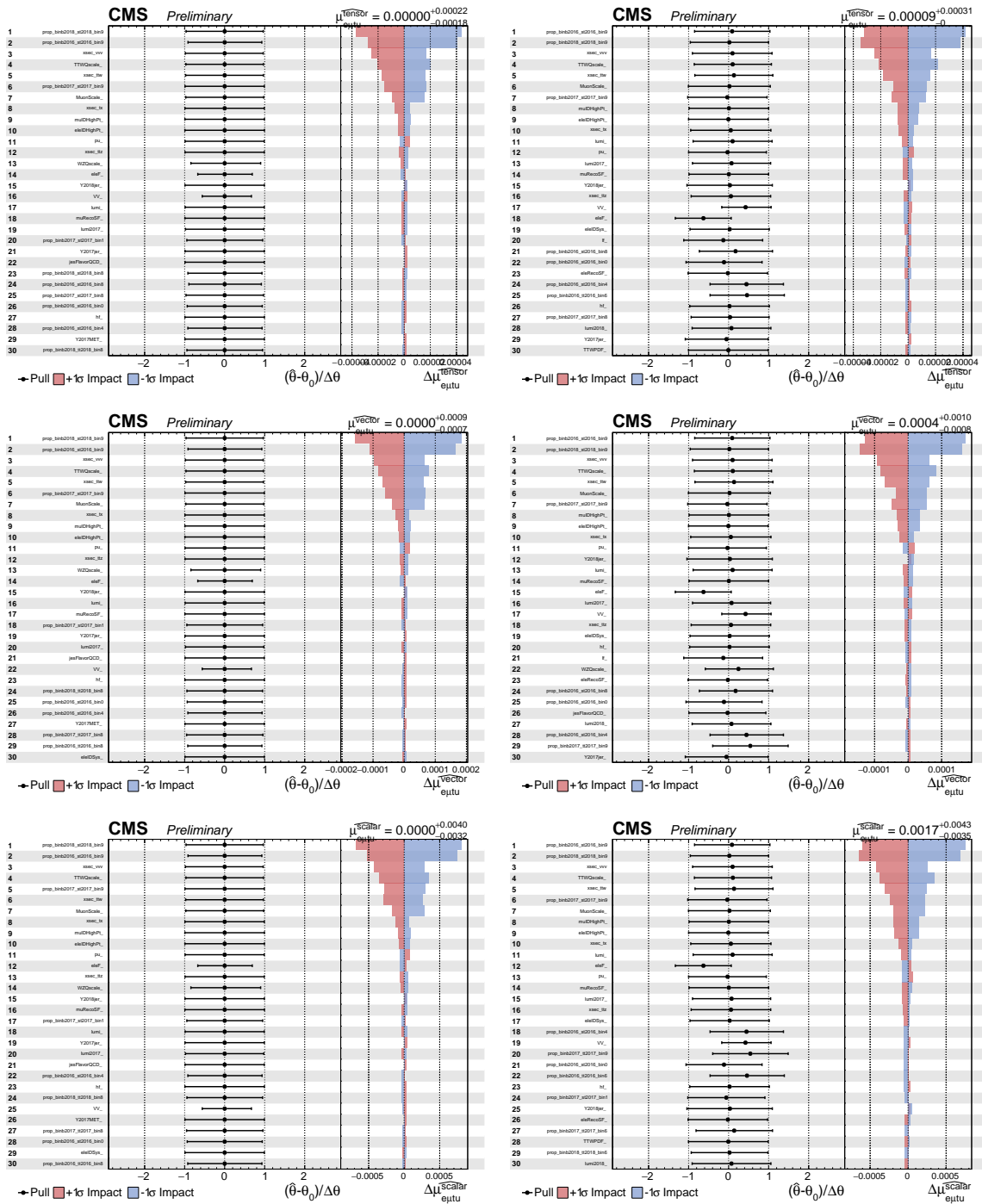
| Dataset    | Trigger Path                                       |
|------------|--|
| SingleMuon | HLT_IsoMu24  |
|            | HLT_IsoMu27  |
|            | HLT_Mu50   |
|            | HLT_OldMu100                                       |
|            | HLT_TkMu100  |
| EGamma     | HLT_Ele32_WPTight_Gsf                              |
|            | HLT_Ele115_CaloldVT_GsfTrkIdT                      |
|            | HLT_Ele23_Ele12_CaloldL_TrackIdL_IsoVL             |
|            | HLT_Ele16_Ele12_Ele8_CaloldL_TrackIdL              |
|            | HLT_DoubleEle25_CaloldL_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_CaloldL_TrackIdL_IsoVL    |
|            | HLT_Mu23_TrkIsoVVL_Ele12_CaloldL_TrackIdL_IsoVL_DZ |
|            | HLT_Mu8_TrkIsoVVL_Ele23_CaloldL_TrackIdL_IsoVL_DZ  |
|            | HLT_Mu12_TrkIsoVVL_Ele23_CaloldL_TrackIdL_IsoVL_DZ |
|            | HLT_Mu27_Ele37_CaloldL_MV                          |
|            | HLT_Mu37_Ele27_CaloldL_MV                          |
|            | HLT_DiMu9_Ele9_CaloldL_TrackIdL_DZ                 |
|            | HLT_Mu8_DiEle12_CaloldL_TrackIdL                   |

# APPENDIX B

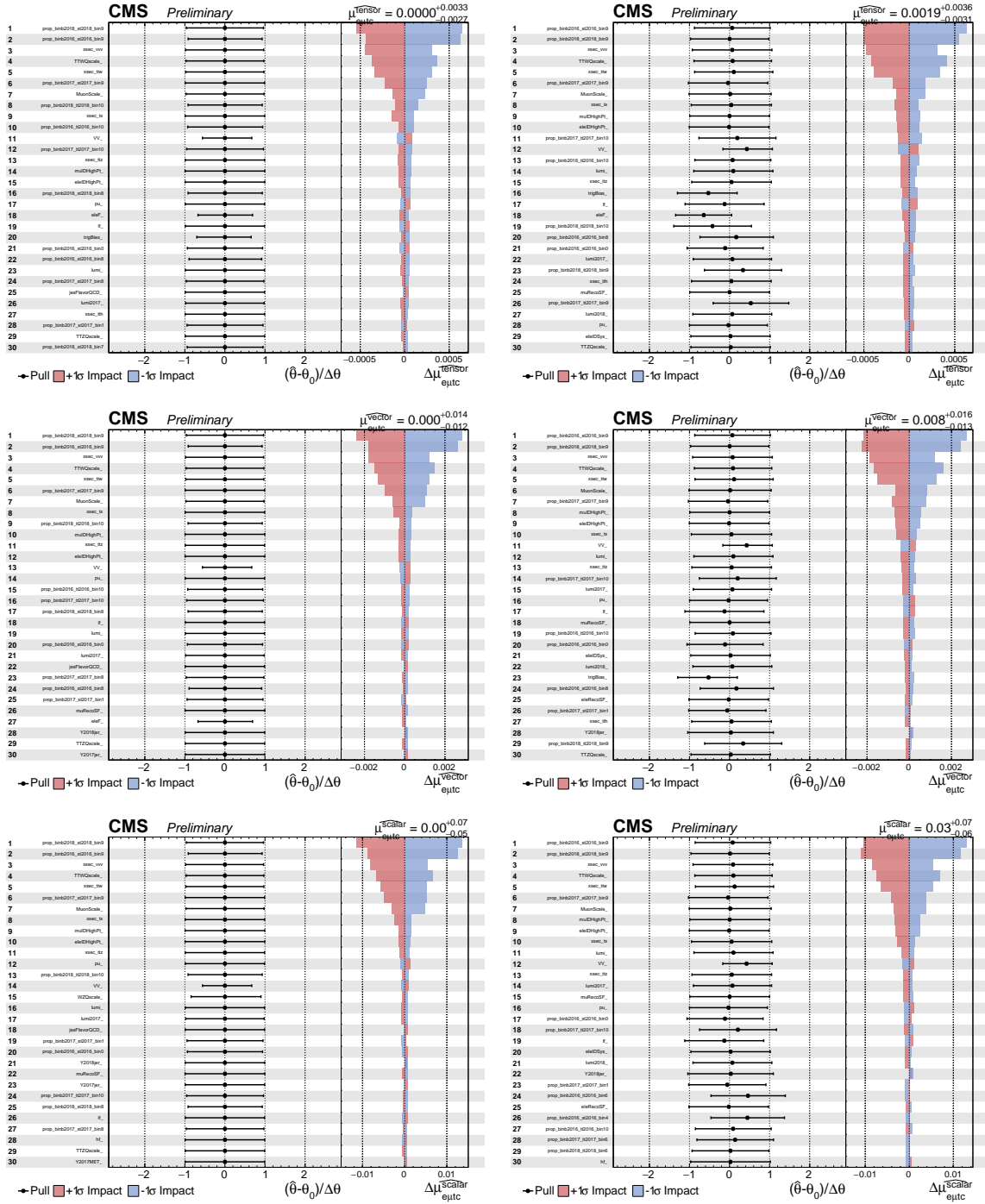
## Nuisance Parameter Impact

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

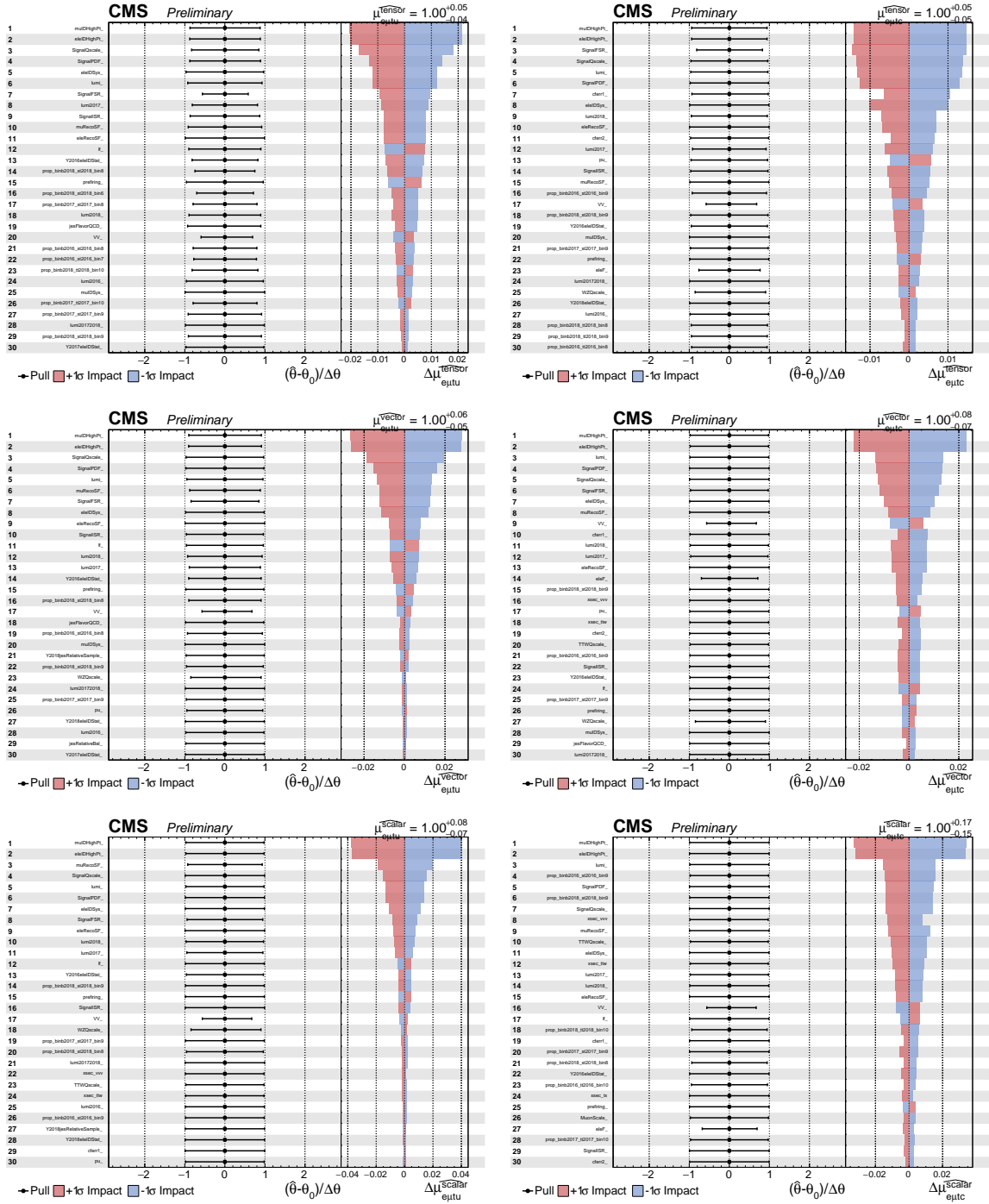
The expected ( $\mu_{exp}=1$ ) impacts of the nuisance parameters on the profile likelihood fit are shown in Figure B.3.



**Figure B.1:** Impacts of nuisance 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 B.2:** Impacts of nuisance 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 B.3:** Expected impact with an expected signal strength at 1. From top to bottom: tensor, vector, scalar. From left to right:  $e\mu tu$ ,  $e\mu tc$ .

# APPENDIX C

## More materials