

A Search for Higgs Decay to Pseudoscalar Higgs-like Particles at  
CMS

Samuel Higginbotham

A Dissertation  
Presented to the Faculty  
of Princeton University  
in Candidacy for the Degree  
of Doctor of Philosophy

Recommended for Acceptance  
by the Department of  
Physics  
Advisor: Daniel Marlow

August, 2022

© Copyright 2022 by Samuel Higginbotham.

All rights reserved.

# Abstract

A search is conducted for exotic decays of the Standard Model Higgs Boson,  $H$ , decaying to a pair of pseudoscalars,  $a$ , which then decay to a pair of muons and tau leptons. Pseudoscalar masses between 15 and 60 GeV are investigated using the full Run II dataset, corresponding to  $137\text{ fb}^{-1}$ . Motivation for the existence of the pseudoscalar Higgs is primarily supported by Beyond Standard Model (BSM) Two Higgs Doublet Models with the extension of a Singlet (2HDM+S) - which include the Next to Minimal Super Symmetric Model (NMSSM). Upper limits on the branching fraction are set.

# Acknowledgements

thank

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The Standard Model and Beyond . . . . .	1
1.2 Gauge Principle, Yang Mills Theories, and the Weinberg-Salam Model . . . . .	2
1.3 The Higgs Mechanism . . . . .	4
1.4 Higgs Doublet Models . . . . .	5
1.5 Previous and present searches in 2HDM+S models . . . . .	8
<b>2 The Large Hadron Collider</b>	<b>10</b>
2.1 CERN and the Large Hadron Collider . . . . .	10
2.2 The Compact Muon Solenoid detector . . . . .	10
2.3 Subdetector Systems . . . . .	13
2.3.1 Muon Identification Systems . . . . .	13
2.3.2 Tau Identification Systems . . . . .	13
<b>3 Silicon Detectors and Luminosity</b>	<b>15</b>

<b>4 Data and Simulation</b>	<b>18</b>
4.1 Data and Simluation Used for Analysis . . . . .	18
4.1.1 $2\mu 2\tau$ Signal Samples . . . . .	19
4.2 Corrections to Simulations . . . . .	19
4.2.1 Muon energy scale . . . . .	20
4.2.2 Electron energy scale . . . . .	20
4.2.3 $\tau$ energy scale . . . . .	20
4.2.4 $\tau_h$ ID efficiency . . . . .	22
4.2.5 $\rightarrow \tau_h$ misidentification rate . . . . .	23
4.2.6 $\rightarrow \tau_h$ misidentification rate . . . . .	23
4.2.7 Pileup reweighting . . . . .	24
4.2.8 Electron and muon identification efficiency . . . . .	24
4.2.9 Generator event weights and luminosity . . . . .	24
4.2.10 Offline Muon Selection . . . . .	25
4.2.11 Visualizing the Corrections . . . . .	25
<b>5 Event Selection</b>	<b>27</b>
5.1 Framing an Analysis at CERN . . . . .	27
5.2 Defining Signal and Control Regions . . . . .	27
5.2.1 Optimizing Lepton Pair Selection . . . . .	28
<b>6 Background</b>	<b>35</b>
6.0.1 Brief Outline of the Fake Rate Method . . . . .	35
6.0.2 Measurement of the Fake Rate . . . . .	36
6.0.3 Application of the Fake Rate Method . . . . .	38
<b>7 Statistical Inference Modeling</b>	<b>40</b>
7.1 Statistical Inference at CERN . . . . .	40
7.2 Worked Example for Low Stat Analyses . . . . .	42
7.3 Fit model for pseudoscalar Higgs search . . . . .	43
<b>8 Results</b>	<b>68</b>

<b>9 Conclusions</b>	<b>74</b>
<b>A Data and Simulation Samples</b>	<b>75</b>
A.1 Data and Simulation Used for Analysis . . . . .	75
<b>B Fake Rate Measurements</b>	<b>80</b>
<b>C Systematic Uncertainties</b>	<b>86</b>
<b>References</b>	<b>97</b>

# List of Figures

1.1	Standard Model particles . . . . .	2
1.2	Type I 2HDM+S scenario branching fractions . . . . .	7
1.3	Type II 2HDM+S scenario branching fractions . . . . .	7
1.4	Type III 2HDM+S scenario branching fractions . . . . .	8
1.5	Type IV 2HDM+S scenario branching fractions . . . . .	8
1.6	Diagram of Higgs decay to pseudoscalar $a$ particles (Left) and branching ratios for pseudoscalar production in different $\tan\beta$ scenarios and different 2HDM+S Types (right) . . . . .	9
2.1	Overview of the Large Hadron Collider spanning Switzerland and France - Maximilien Brice, CERN . . . . .	11
2.2	Muon system involving multiple subtector systems: tracker, solenoid, and the muon gas chambers around the iron yoke (grey) - Maximilien Brice, CERN . . . . .	14
3.1	encapsulated token bit manager of a forward pixel module currently installed in CMS	16
4.1	diagram depicting the possible decays of the tau lepton: 65% of the time hadronically and 18% to tau neutrino and electron with antielectron neutrino (17% for muon) . .	21
4.2	Tau mass distributions considering the nominal $\tau_h$ energy scale in simulation (left), or the $\tau_h$ energy scale shifted by -6 (left) or +6% (right), in the $\mu\tau_h$ final state, for the 1 prong + pizero decay mode. . . . .	22
4.3	Systematic Shift in the Uncertainty Model for 2016 $\mu\mu\tau\tau$ for the muon energy scale shift down (left), nominal (mid), and up (right), no data is shown on this plot as it directly reflects the signal region without the extraction cuts . . . . .	25

5.1	Invariant Mass of the four lepton system for 2016 data in all Channels . . . . .	31
5.2	Invariant Mass of the four lepton system for 2017 data in all Channels . . . . .	32
5.3	Invariant Mass of the four lepton system for 2018 data in all Channels . . . . .	33
5.4	Several Data-MC control plots for full RunII data in all Channels . . . . .	34
6.1	Fake factors determined in the QCD multijet determination region with 0 jet in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	37
6.2	Diagram depicting the measurement and application regions this ABCD method is multiplied by the two $\tau$ leptons in the final sate . . . . .	39
6.3	Validation of the fake factor method, fake factors are applied to the same sign tight region . . . . .	39
7.1	Signal fit using a Voigtian function . . . . .	44
7.2	Signal fits using a voiigtian function for a-mass at 40GeV . . . . .	45
7.3	spline functions for 2016 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	46
7.4	spline functions for 2016 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	47
7.5	spline functions for 2016 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	48
7.6	spline functions for 2016 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	49
7.7	irreducible background fit using Bernstein polynomials . . . . .	50
7.8	reducible background fit using Bernstein polynomials . . . . .	51
7.9	2017 Signal fit using a Voigtian function . . . . .	52
7.10	spline functions for 2017 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	53

7.11	spline functions for 2017 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	54
7.12	spline functions for 2017 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	55
7.13	spline functions for 2017 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	56
7.14	2017 irreducible background fit using Bernstein polynomials . . . . .	57
7.15	2017 reducible background fit using Bernstein polynomials . . . . .	58
7.16	2018 Signal fit using a Voigtian function . . . . .	59
7.17	spline functions for 2018 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	60
7.18	spline functions for 2018 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	61
7.19	spline functions for 2018 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	62
7.20	spline functions for 2018 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean . . . . .	63
7.21	2018 irreducible background fit using Bernstein polynomials . . . . .	64
7.22	2018 reducible background fit using Bernstein polynomials . . . . .	65
7.23	Expected systematic impacts for the fit model $\mu\mu\mu\tau$ . . . . .	67
8.1	Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sections	69
8.2	Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sec- tions for 2017 . . . . .	70
8.3	Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sec- tions for 2018 . . . . .	71
8.4	Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sec- tions for the full Run II dataset $137^{-1}\text{fb}$ . . . . .	72
8.5	2HDM model specific scenarios upper 95% C.L. limits on the branch fraction of Higgs to a a times the ratio of the SM cross sections for the full Run II dataset $137^{-1}\text{fb}$ .	73

B.1	Fake factors determined in the W+jets determination region with 0 jet in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	80
B.2	Fake factors determined in the W+jets determination region with one jet in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	81
B.3	Fake factors determined in the W+jets determination region with at least two jets in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	81
B.4	Fake factors determined in the W+jets determination region with 0 jet in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	81
B.5	Fake factors determined in the W+jets determination region with one jet in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	82
B.6	Fake factors determined in the W+jets determination region with at least two jets in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	82

B.7	Fake factors determined in the QCD multijet determination region with 0 jet in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	82
B.8	Fake factors determined in the QCD multijet determination region with one jet in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	83
B.9	Fake factors determined in the QCD multijet determination region with at least two jets in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	83
B.10	Fake factors determined in the QCD multijet determination region with one jet in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	83
B.11	Fake factors determined in the QCD multijet determination region with at least two jets in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	84
B.12	Fake factors determined in the $t\bar{t}$ determination region in data in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	84

B.13	Fake factors determined in the $t\bar{t}$ simulation in the $e\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	84
B.14	Fake factors determined in the $t\bar{t}$ determination region in data in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	85
B.15	Fake factors determined in the $t\bar{t}$ simulation in the $\tau_h$ final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit. . . . .	85
C.1	Expected systematic impacts for the fit model $\mu\mu\mu\tau$ . . . . .	86
C.2	Expected systematic impacts for the fit model $\mu\mu e\tau$ . . . . .	87
C.3	Expected systematic impacts for the fit model for 2016 $\mu\mu\tau\tau$ . . . . .	88
C.4	Expected systematic impacts for the fit model for 2016 $\mu\mu e\mu$ . . . . .	89
C.5	Expected systematic impacts for the fit model $\mu\mu e\tau$ . . . . .	90
C.6	Expected systematic impacts for the fit model for 2017 $\mu\mu\tau\tau$ . . . . .	91
C.7	Expected systematic impacts for the fit model for 2017 $\mu\mu e\mu$ . . . . .	92
C.8	Expected systematic impacts for the fit model $\mu\mu\mu\tau$ . . . . .	93
C.9	Expected systematic impacts for the fit model $\mu\mu e\tau$ . . . . .	94
C.10	Expected systematic impacts for the fit model for 2018 $\mu\mu\tau\tau$ . . . . .	95
C.11	Expected systematic impacts for the fit model for 2018 $\mu\mu e\mu$ . . . . .	96

# List of Tables

4.1	Measured energy scale correction for genuine across all years. . . . .	20
4.2	Measured energy scale correction for genuine across all years. . . . .	21
4.3	Measured $\tau_h$ energy scale correction for genuine $\tau_h$ 's across all years. . . . .	22
4.4	Scale factors to correct for Offline muon selection being less than the online trigger threshold. . . . .	25
5.1	Event selection requirements for the four decay channels. . . . .	29
7.1	. . . . .	66
8.1	Expected Event Yields of Signal an Background Across All Years . . . . .	68
A.1	List of data sets included in the analysis for the 2016 data taking period. . . . .	75
A.2	List of data sets included in the analysis for the 2017 data taking period. . . . .	76
A.3	List of data sets included in the analysis for the 2018 data taking period. . . . .	76
A.4	List of data sets included in the analysis for the 2016 data taking period. . . . .	77
A.5	List of data sets included in the analysis for the 2017 data taking period. . . . .	78
A.6	List of data sets included in the analysis for the 2018 data taking period. . . . .	79

# Chapter 1

## Introduction

### 1.1 The Standard Model and Beyond

Holding 18 particles, numerous interactions between particle fields, containing many symmetries and holding a description of the fundamental interactions within nature, the Standard Model is the most precise theory with empirical evidence in the universe. This introduction's purpose is not to derive the standard model or show all the fundamental interactions; however, the aim is to frame the analysis conducted by briefly showcasing the Standard Model and it's extension of supersymmetry with two Higgs doublet models. A table listing the particles can be referenced below in figure 1.1.

Particles can be represented by fields and interact within the theory. To frame the Standard Model the typical approach is to look at the Dirac Lagrangian and then establish  $U(1)$  interactions that have a conserved quantity (charge). Emmy Noether originally proved that group global transformations implies conserved quantities which has extensive implications for particle physics and theories that predict quantum numbers ???. After analyzing the  $U(1)$  group, one typically extends the theory to include more fields and structure through higher dimensional groups. To parallel typical literature in a concise fashion, the gauge principle will be examined for  $U(1)$  and then expanding to  $SU(2)$ , before taking the external product of these groups to form the Weinberg-Salam Model.

## Standard Model of Elementary Particles

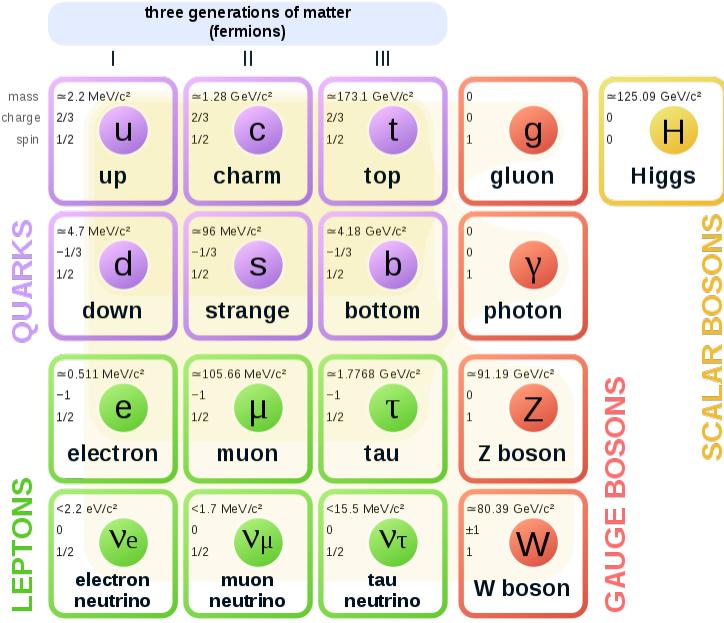


Figure 1.1: Standard Model particles

## 1.2 Gauge Principle, Yang Mills Theories, and the Weinberg-Salam Model

The gauge principle sets the stage for interactions between particle-fields in a theory. Much like in differential geometry and general relativity, there is a cost to interacting or change through the action of a transformation. Consider the covariant derivative on a vector field  $V(x)$ .

$$\mathcal{D}_\mu \equiv \lim_{\Delta x \rightarrow 0} \frac{V_{||}(x + \Delta x) - V(x)}{\Delta x^\mu}$$

How the field changes under the transformation will determine the nature of the field. The connection terms that link the field with the effect of the transformation follow from the transformation. A unitary operator can capture the parallel component in the covariant derivative. This operator carries the field and the *local* transformation law depending on the symmetries and complexity of the math structure that the unitary operator carries.

$$V_{||}(x + \Delta x) = U(x + \Delta x)V(x + \Delta x)$$

After expanding the unitary operator one can find the terms and phases that are carried under this transformation. For the  $U(1)$  group these actions give the electromagnetic field tensor and charge

conservation.

So in any theory the typical thing to do is to define the transformation and field then work out the form of the covariant derivative and examine how the gauge field transforms [1]. To put it generally

$$\Psi' = U(\vec{x})\Psi \text{ particle under local transformation} \quad (1.1)$$

$$\mathcal{D}^\mu = \partial^\mu + igB^\mu \text{ covariant derivative} \quad (1.2)$$

$$B^\mu = UB^\mu U^{-1} + \frac{i}{g}(\partial^\mu U)U^{-1} \text{ gauge field transform} \quad (1.3)$$

For  $U(1)$  with a unitary transformation of the form  $\exp \frac{i}{2}Y \cdot \beta x$

$$\Psi'(x) = (1 + i\alpha)\Psi \text{ particle field with local transformation}$$

$$\mathcal{D}_\mu = \partial_\mu + ieA_\mu \text{ covariant derivative}$$

$$B^\mu = B^\mu + \frac{1}{g}(\partial^\mu \beta) \text{ gauge field transform}$$

An interesting commutation relation is worth pointing out. The commutator between the covariant derivatives yields the practical field that it carries. For the  $U(1)$  case, it is the electromagnetic field tensor.

$$[\mathcal{D}^\mu, \mathcal{D}^\nu] \Psi = i \cdot e \cdot F^{\mu\nu} \Psi$$

These relations are important when higher dimensional groups and more complex particle fields are considered. Yang Mill theories take these theoretical components and analyzes it under groups like  $SU(2)$  or other special unitary groups. When one considers a local  $SU(2)$  transformation, the terms that show up are more rich than  $U(1)$ .

$$\partial_\mu \rightarrow \partial_\mu + \frac{i}{2}(\tau \cdot \alpha(\mathbf{x})) \partial_\mu \Psi + \frac{i}{2}(\tau \cdot \partial_\mu \alpha(\mathbf{x})) \Psi$$

implying the covariant derivative

$$\mathcal{D}_\mu = \partial_\mu - \frac{i}{2}g\tau \cdot \mathbf{W}_\mu(\mathbf{x})$$

Where  $\tau$  are the Pauli matrices. The gauge field then transforms as

$$\tau \cdot \mathbf{W}_\mu \rightarrow \tau \cdot \mathbf{W}_\mu + \frac{1}{g}\tau \cdot (\partial_\mu \alpha) - \tau \cdot (\alpha \times \mathbf{W}_\mu)$$

So to concatenate the information:

For  $SU(2)$  with a unitary transformation of the form  $\exp \frac{i}{2} \tau \cdot \alpha \mathbf{x}$

$$\Psi'(x) = (1 + \frac{i}{2} \tau \cdot \alpha) \Psi \text{ particle field with local transformation}$$

$$\mathcal{D}_\mu = \partial_\mu - \frac{i}{2} g \tau \cdot \mathbf{W}_\mu(\mathbf{x}) \text{ covariant derivative}$$

$$\tau \cdot \mathbf{W}_\mu \rightarrow \tau \cdot \mathbf{W}_\mu + \frac{1}{g} \tau \cdot (\partial_\mu \alpha) - \tau \cdot (\alpha \times \mathbf{W}_\mu) \text{ gauge field transform}$$

Now if we consider  $SU(2) \otimes U(1)$  and build states out of that with a coupling of only one component in the  $SU(2)$  interaction then add a scalar field, the structure is there for the Weinberg-Salam model.

### 1.3 The Higgs Mechanism

As mentioned in the previous section, taking the  $SU(2) \otimes U(1)$  groups with an additional scalar field and defining their local transformations indicate the interactions within the electroweak Weinberg-Salam model. The fields transform the same way as before; however, Peter Higgs and company thought of adding a scalar field to theory. In order to see the interactions with the scalar field we need to see how it transforms. The scalar field interacts with the other fields under the  $SU(2)$  and the  $U(1)$  transformations.

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_3 - i\phi_4 \end{pmatrix}$$

This scalar field transforms in the following way under  $U(1)$  and  $SU(2)$  transformations

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow \begin{pmatrix} e^{i\beta/2} & 0 \\ 0 & e^{i\beta/2} \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow e^{\frac{i}{2}\tau \cdot \alpha} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

$$\mathcal{D}_\mu \phi = \partial_\mu \phi - \frac{i}{2} g \tau \cdot \mathbf{W}_\mu \phi - \frac{i}{2} g' \beta_\mu \phi$$

yielding the Higgs field lagrangian component

$$\mathcal{L} = (\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi) + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2$$

The Higgs potential contains the typical "Mexican hat" shape and the kinetic term holds interesting interactions and implications for the gauge bosons in the theory. All together the Weinberg-Salam model with the Higgs field is [2]

$$\mathcal{L} = \bar{L} i \gamma^\mu \mathcal{D}_\mu L + \bar{R} i \gamma^\mu \mathcal{D}_\mu R + (\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi) \quad (1.4)$$

$$\begin{aligned} &+ \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2 - G_e (\bar{L} \phi R + \bar{R} \phi^\dagger L) \\ &- \frac{1}{4} G_{\mu\nu}^{(W)} \cdot G^{(W)\mu\nu} - \frac{1}{4} F_{\mu\nu}^{(B)} F^{(B)\mu\nu} \end{aligned} \quad (1.5)$$

The Weinberg-Salam model has major implications, the terms  $L$  and  $R$  along with their hermitian conjugates form an interaction with the Higgs field giving electrons mass. The kinetic terms  $(\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi)$  imply interactions that show a massless photon and massive gauge bosons.

## 1.4 Higgs Doublet Models

In order to extend the model beyond the standard model, a natural place to start is giving more complexity to the fields. Suppose that there were multiple scalar fields instead of just the single Higgs field  $\phi$ , then one can consider adding another component, thus making it a doublet. Adding the doublet, and working out the relations for the field, the most general two Higgs doublet model (2HDM) Higgs potential is found in equation 1.6 [3].

$$\begin{aligned} V = & m_1^2 |H_1|^2 + m_2^2 |H_2|^2 + \frac{\lambda_1}{2} |H_1|^2 + \frac{\lambda_2}{2} |H_2|^2 \\ &+ \lambda_3 |H_1|^2 |H_2|^2 + \lambda_4 |H_1^\dagger H_2|^2 \\ &+ \frac{\lambda_5}{2} \left( (H_1 H_2)^2 + c.c. \right) + m_{12}^2 (H_1 H_2 + c.c.) \\ &+ \left( \lambda_6 |H_1|^2 (H_1 H_2) + c.c. \right) \\ &+ \left( \lambda_7 |H_2|^2 (H_1 H_2) + c.c. \right) \end{aligned} \quad (1.6)$$

Expanding around the minima of the potential yields two doublets with vacuum expectation values, which is usually mixed under a rotation parameter  $\tan \beta = \frac{v_1}{v_2}$ . After one carries out the

interactions with the SM gauge bosons which consume the complex field components and a neutral pseudoscalar combination of scalar Higgs field components, the surviving three real degrees of freedom yield one neutral pseudoscalar mass eigenstate along with two, neutral scalar mass eigenstates, which will be denoted  $A$  for the pseudoscalar,  $h$  for the lighter neutral SM like Higgs, and  $H^0$  for the remaining scalar. As with notation for fields with interact with the potential, the rotation matrix that mixes these scalars into the components that interact under the potential is parametrized by the continuous parameter  $\alpha$ . In literature this parameter is important because  $\tan \beta$  and  $\alpha$  set the possible couplings to SM particles. Next, if a complex scalar singlet is added, couplings to SM fermions and bosons are supported [4].

$$S = \frac{1}{\sqrt{2}}(S_R + iS_I)$$

It doesn't have Yukawa couplings and couples to  $H_{1,2}$  - and correspondingly to Standard Model (SM) fermions. Preserving the SM Higgs  $h$ , the mixing needs to be small which then allows for decays like  $h \rightarrow aa \rightarrow X\bar{X}Y\bar{Y}$ , where  $X$  and  $Y$  are SM fermions or bosons. Looking into the phase space where the mixing is light then frames the Higgs pseudoscalar analysis in a region with little SM resonance - making it also a general search for any beyond SM phenomena.

$$a \equiv \cos \theta_a S_i + \sin \theta_a A, \quad \theta_a \ll 1$$

There are terms in the effective lagrangian that support the  $h \rightarrow aa$  decays:

$$\begin{aligned} \mathcal{L} &\subset g_{hAA} hAA + \lambda_S |S|^2 \\ &\subset g_{hAA} \sin^2 \theta_a haa + 4\lambda_S v_s \sin \zeta_1 \cos^2 \theta_a haa \end{aligned} \tag{1.7}$$

here  $\zeta$  is just the angle that mixes the singlet with the SM Higgs (needed because of the added state). These interactions give rise to different scenarios that favor certain SM fermions and bosons. According the literature four distinct scenarios are typically entertained, enumerated below. The scenarios, supported by the effective lagrangian 1.7, yield branching fractions as a function of a-mass and  $\tan \beta$ .

- Type I: Fermions couple only to the  $H_2$  field and are independent of  $\tan \beta$ , pseudoscalar coupling are proportional to the SM Higgs (final state of the fermions) represented in figure 1.2.

- Type II: Down-type fermions are particularly favored supporting more NMSSM models, and is dependent on  $\tan \beta$  represented in figure 1.3.
- Type III: Branching ratios are directly dependent on  $\tan \beta$  and are emphasized when more than one lepton is considered.  $\tau^+ \tau^-$  can dominate in this scenario represented in figure 1.4.
- Type IV: Dependent on  $\tan \beta$ , for  $\tan \beta < 1$  branching ratios for  $b\bar{b}$ ,  $c\bar{c}$ , and  $\tau^+ \tau^-$  are similar (supports  $2b2\tau$ ) represented in figure 1.5.

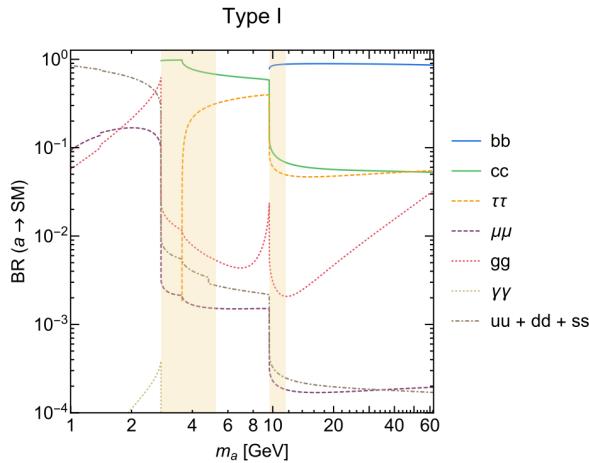


Figure 1.2: Type I 2HDM+S scenario branching fractions

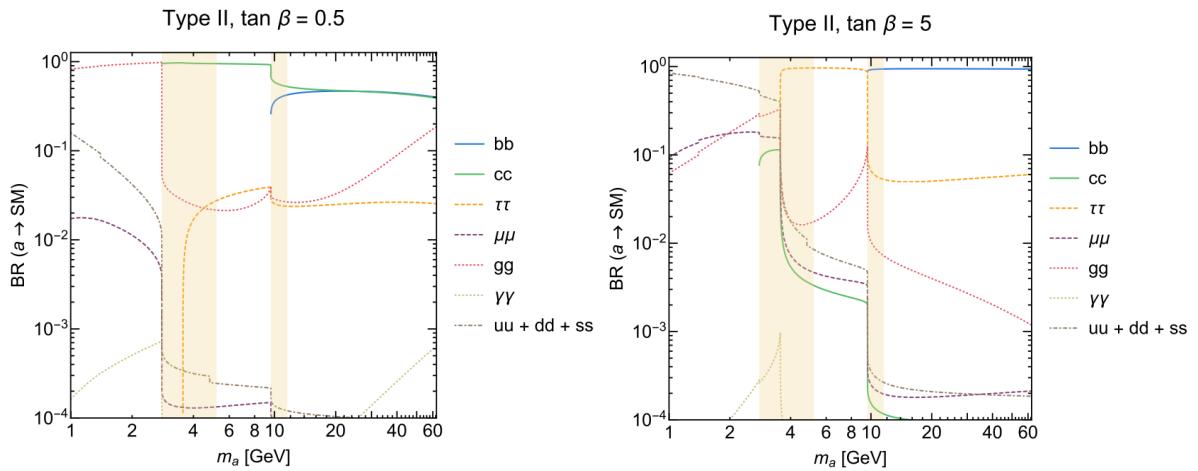


Figure 1.3: Type II 2HDM+S scenario branching fractions

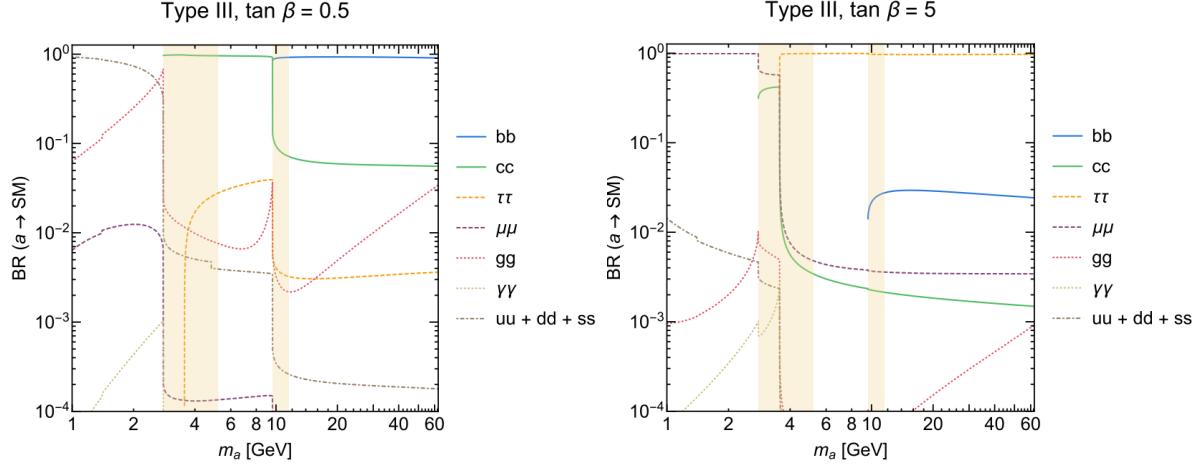


Figure 1.4: Type III 2HDM+S scenario branching fractions

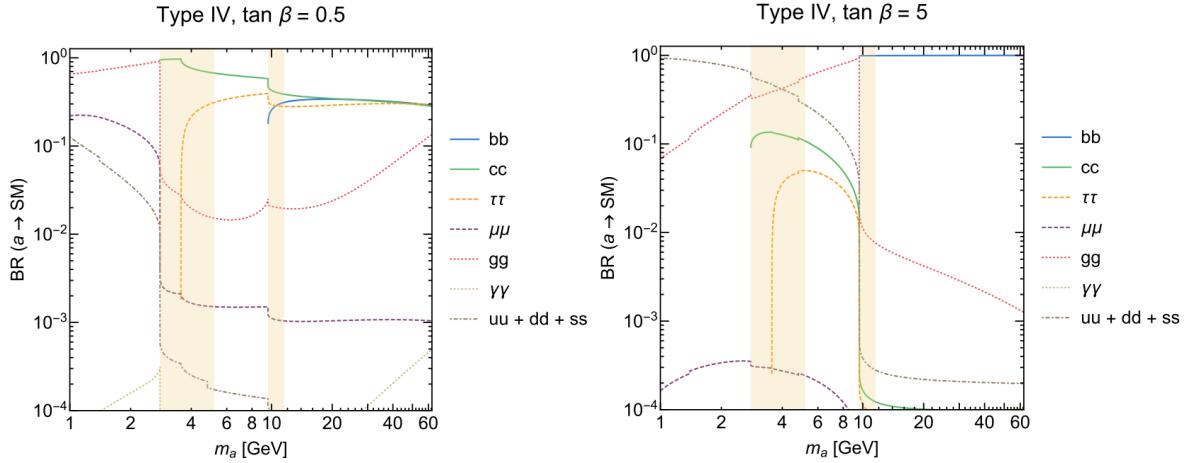


Figure 1.5: Type IV 2HDM+S scenario branching fractions

## 1.5 Previous and present searches in 2HDM+S models

As shown in the previous section, the Higgs couples to all SM particles and also to new particles provided that the new particle has mass. Beyond Standard Model (BSM) theories contain ample room for the Higgs to couple to particles beyond the SM, making the Higgs an excellent window to investigate any physics beyond the SM. Notably, the two Higgs doublet model (2HDM) with its extension of a scalar singlet is considered (2HDM+S). These types of BSM theories can solve the  $\mu$  coupling problem in Super Symmetry (SUSY), while maintaining general support of SUSY (Holomorphy), Axion-like Models (Peccei-Quinn), electroweak baryogenesis and several Grand Unified Theories (GUTs) [3].

A representative diagram showing the physics process and the Branching ratio as a function of  $\tan\beta$  is shown in figure 1.6. This pseudoscalar Higgs search for "resolved"  $a$  particles in the range of 20 to 60 GeV is a good search for new physics. In 2016 this general search was carried out with  $35.9 \text{ fb}^{-1}$  of data and new competitive limits were set in article [5]. Given the improvement in the limits for different 2HDM+S types, this search has garnered attention for the full Run II dataset of CMS.

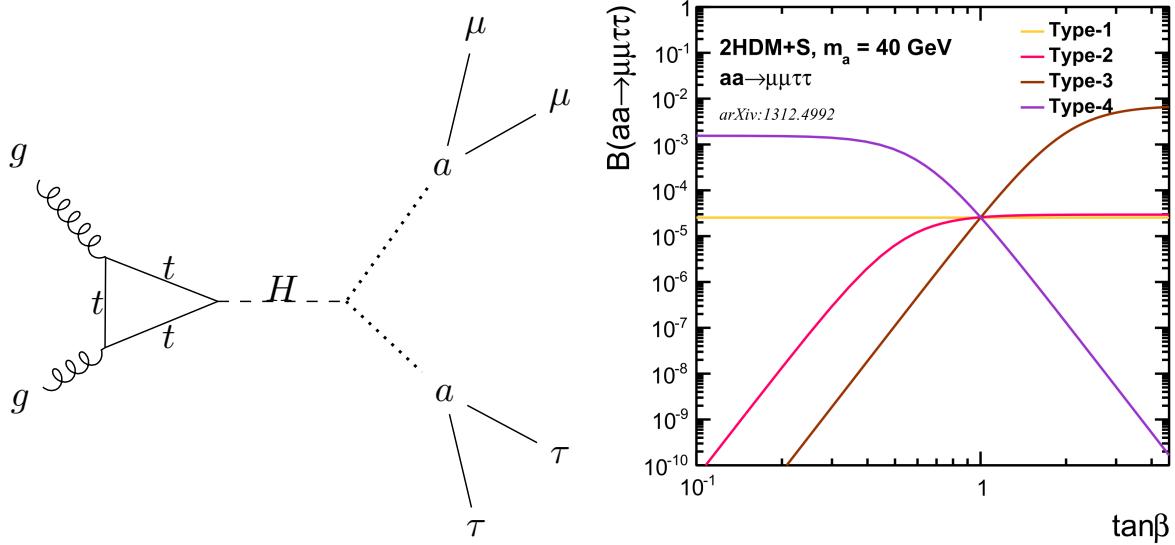


Figure 1.6: Diagram of Higgs decay to pseudoscalar  $a$  particles (Left) and branching ratios for pseudoscalar production in different  $\tan\beta$  scenarios and different 2HDM+S Types (right)

Certain scenarios are entertained using yields provided by different 2HDM+S models. The branching ratios change based on the function of  $\tan\beta$  depending on the type of model under investigation. In particular, we look at types 1-4 which were showcased in the previous section and explained in detail here [?]. Type III is expected to be most sensitive as it maintains a larger branching ratio compared to other decay modes over the range of the pseudoscalar masses when focusing on the final state of two muons and two tau leptons.

Using the full Run II dataset, more constraints on these models are expected; however, the search is looking for any deviation of the standard model with the effective mass range making the search somewhat model independent.

# Chapter 2

## The Large Hadron Collider

### 2.1 CERN and the Large Hadron Collider

The organisation européenne pour la recherche nucléaire or CERN is the organization that conducting the world's frontier particle physics experiments. It's a coalition of scientists representing numerous countries around the world, who work together for a greater understanding of the universe. CERN hosts the Large Hadron Collider, currently the largest particle collider in the world. It is 27 km in circumference and hosts eight experimental caverns 150 meters underneath the earth [1]. Accelerator physicists and engineers strive to provide high energy collisions along these eight sites. Monumental effort is made to guide the beam, utilizing more than 2000 superconducting magnets to steer, press, and shape the accelerating protons. Crossing points facilitated by crab-like magnets force these particles to cross, collide, and interact at the caverns where the experiments take data. Multiple campaigns for particle physics are done such as low intensity fills, lead-lead collisions, and proton collision. The Compact Muon Solenoid (CMS) is the general purpose proton-proton detector, located at P5 in Cessy, France, that took the data that will be analyzed in this dissertation.

### 2.2 The Compact Muon Solenoid detector

At 14,000 tonnes the CMS detector may not seem like it would be compact; however, it is quite dense. As shown in the diagram, the detector contains many sub-detector systems and also features

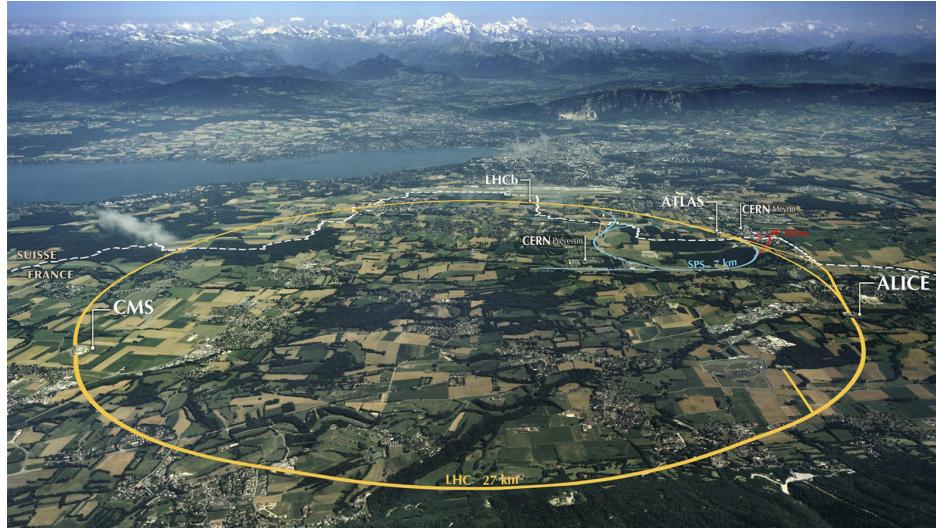
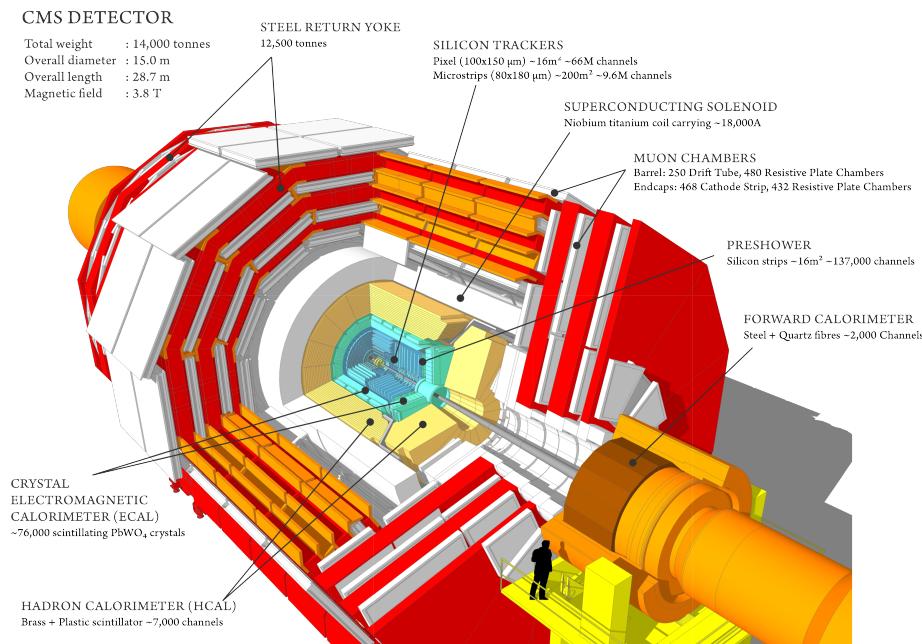


Figure 2.1: Overview of the Large Hadron Collider spanning Switzerland and France - Maximilien Brice, CERN

the most powerful solenoid magnet ever made for it's size. At 6 meters inner diameter, and a combined weight of the solenoid and the steel return yolk of 12,500 tonnes it is the central feature of the Compact Muon Solenoid (CMS) 2.2. Within the solenoid volume, there is the



silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap

sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid.

From the central interaction point, the CMS detector hosts the silicon pixel and strip tracker, the lead tungstate crystal electromagnetic calorimeter (ECAL), and the brass-scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. The silicon pixel and tracking systems as well as the calorimeters are contained within the solenoid volume.

The nominal bunch crossing rate at the LHC is 40 MHz. This rate would be too extreme for any data taking system. In order to reduce the rate of events that are recorded for offline analysis, events of interest are selected using a two-step trigger system [?]. The first level (L1) is composed of custom built electronics which makes use of high speed optical links and large Field Programmable Gate Arrays (FPGAs). L1 reduces the event rate from the nominal bunch crossing to a rate of around 100 kHz within a time interval of less than  $3.5\ \mu\text{s}$ . The second level, known as the High Level Trigger (HLT), consists of a farm of generic processors running a version of the full event reconstruction software that has been optimized for fast processing. The HLT reduces the event rate to about 1 kHz before data storage.

Since the 2012 data taking, significant upgrades of the L1 trigger have benefited this analysis, especially in the final state with two semi-hadronically decaying leptons, denoted as  $\tau_h$ . These upgrades improved the  $\tau_h$  identification at L1 by giving more flexibility to object isolation, allowing new techniques to suppress the contribution from additional interactions per bunch crossing, and to reconstruct the L1  $\tau_h$  object in a fiducial region that matches more closely that of a true  $\tau_h$  decay.

A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [6].

## 2.3 Subdetector Systems

### 2.3.1 Muon Identification Systems

While all subdetector systems are important to physics at CMS due to the reconstruction, there are a couple detectors that contribute to the particles that are identified in pseudoscalar search. These are the tracker system, muon system, and the hadronic calorimeter. With two muons and two tau leptons in the final state, muon and tau identification are paramount to the search.

In order to identify good muons, several subdetector systems work together in order to reconstruct muons. Looking at muons that come from the interaction vertex - prompt muons - the tracker plays an important role in indentifying charged particle tracks.

As CMS implies in it's name, muons are certainly a focal point in particle detection. Muon reconstruction is possible for the whole kinematic range of the LHC. The tracker system works in tandem with the gas chambers to reconstruct muons, and the solenoid accentuates charge particle's momentum allowing for the most accurate mass resolution possible in a hadronic based physics detector. The gas chambers comprise of Drift Tubes which cover a pseudorapidity region ( $|\eta| < 1.2$ ) split into 4 stations in the flux return plates.

In the higher  $|\eta|$  regions, cathode strip chambers (CSCs) are used and preferred for their fast response time, fine segmentation, and radiation resistance [7]. For a visual representation of a cross section of the muon system please refer to figure 2.2. Notably, the track of the muon is bent inside the solenoid by the Lorentz force, but outside of that the muon isn't curved in the typical helical fashion.

During reconstruction, the particle flow algorithm [8] can identify tracker or global muons which are ultimately divided into four sub types depending on the  $\chi^2$  of the track reconstruction and momentum of the candidate muon [9]

### 2.3.2 Tau Identification Systems

Tau leptons are the probably the most interesting SM particle. They are leptons, that directly physics of tau decay types of tau decay pions and intermediate mesons HPS

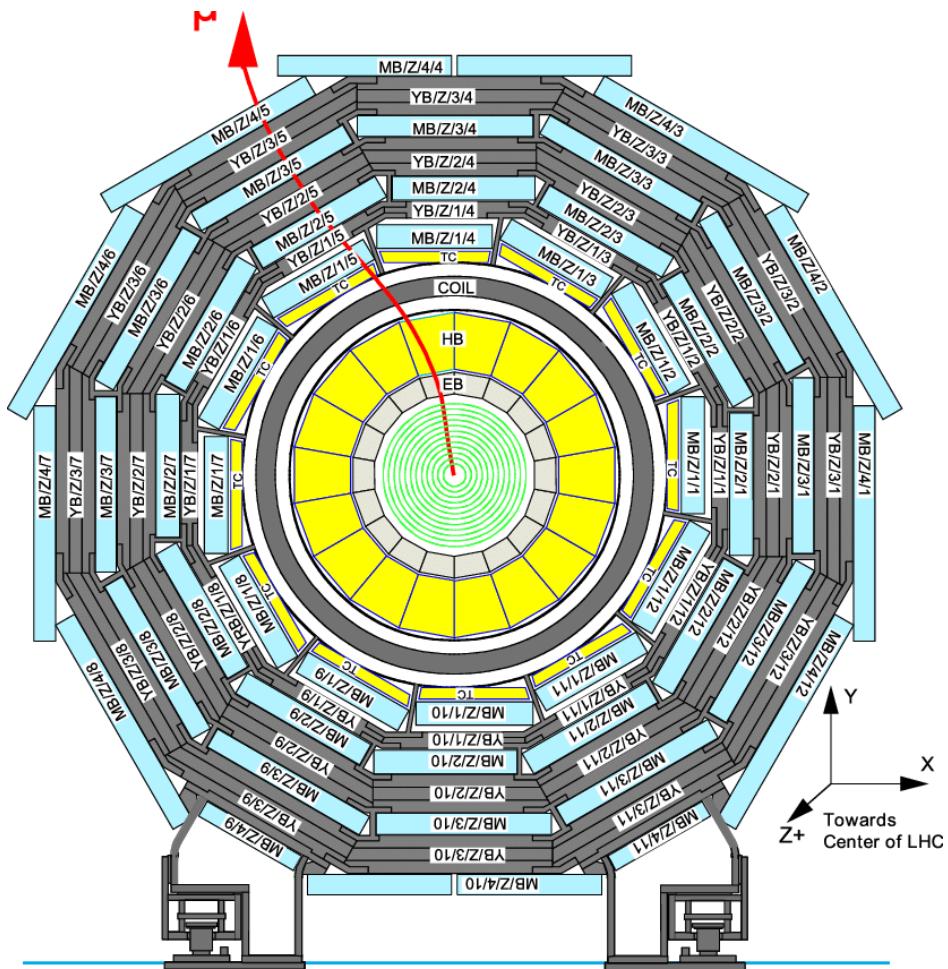


Figure 2.2: Muon system involving multiple subdetector systems: tracker, solenoid, and the muon gas chambers around the iron yoke (grey) - Maximilien Brice, CERN

## Chapter 3

# Silicon Detectors and Luminosity

Luminosity sets the scale for physics at the LHC. It is how bright the beam is and correspondingly, dictates how many interactions can be expected over a data taking period. Therefore, it is important for all physics analyses to use the correct luminosity - and its measured error - to obtain an accurate result.

Several subsystems are used to measure luminosity at CMS. Noteably, the Pixel Luminosity Telescope, Hadronic Calorimeter (HF) with summed transverse energy(ET) and zero occupancy (0C), BCM1F, and Tracker based or Pixel based luminosity. In this section, the tracker based luminosity will be the focus. The pixel system has been part of my research with CMS for almost ten years and is integral in the reconstruction of events for physics analysis and in measuring the luminosity.

In 2016, the phase 1 Forward Pixel System (FPIX) was constructed and tested. At Purdue University, an Aerotech robotic gantry control system was used to join a hybrid flex circuit to a bump-bonded silicon pixel module. Then after wirebonding, the gantry system encapsulated the wirebonds for protection from corrosion and magnetic field resonance.

Using LabVIEW, we developed a state machine to assemble and encapsulate these pixel modules. Pattern recognition and a linear algebra suite were developed to perform precise operations at a 50 micron resolution. An example of a post encapsulated token bit manager - which resides on top of the high density interconnect of a completely assembled module - is shown in figure 3.1.

In 2017, this system was installed in CMS. Due to the experimental design of CMS, the inner sub-detector systems may be taken out of the inner body - the solenoid - and serviced.

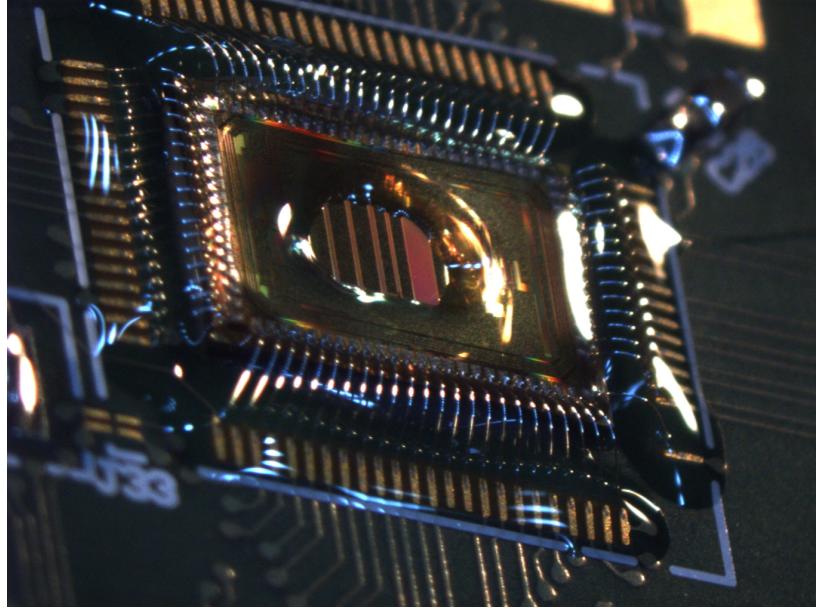


Figure 3.1: encapsulated token bit manager of a forward pixel module currently installed in CMS

The pixel system plays an important role in measuring the luminosity. There was validation of the new pixel detector in early 2017 and within the Lumi-POG (Luminosity physics object group), automation of measuring the luminosity using the clusters from the pixel detector was completed. We measure the luminosity by counting pixel clusters in a low channel occupancy setting and scaling it by a visual cross Section determined in a separate analysis.

$$\langle N_{\text{cluster}} \rangle \equiv \frac{\sigma_{\text{cluster}}}{f_{\text{LHC}}} \mathcal{L}_{\text{SBIL}} \quad (3.1)$$

Where  $f_{\text{LHC}}$  is the revolution frequency of the LHC, and  $\mathcal{L}_{\text{SBIL}}$  is the instantaneous luminosity of a single bunch crossing - the aggregate collection of protons that are in the beam typically 3564 total bunches during standard pp-collisions.  $\sigma_{\text{cluster}}$ , is the cross Section that is measured in a separate analysis, more details can be found here [?].

For the pixel luminosity, a two component correction is applied on the fly to correct for radiative effects on the pixel modules and for inefficiency. One detail that is important in estimating the luminosity from the pixel detector is ensuring that the data that is taken and analyzed has consistent performance. Several times in a year, the Lumi-POG and Beam Radiation Instrumentation Luminosity (BRIL) groups analyze the performance of each sub-detector used to measure luminosity and certify the data once the analysis is complete. In 2017 and 2018 data taking campaigns,

the luminosity from the pixel detector is vetted by looking at relative module performance over the runs of data taking for the year. If the modules don't have consistent performance they are removed from the final result.

Even though up to half of the modules are vetoed after this procedure, there are plenty of statistics (lots of pixel clusters per event) to keep the statistical uncertainties low. This was made by taking the total clusters in each module and scaling them so that the overall total clusters are 1, then per module the ratio is taken of that fraction of the total for each run. This helps the analyzer look at consistent performance and manage a list of passing modules to be used in a final luminosity measurement using the pixel detector.

For Run III data taking (Summer 2022 to Fall 2025), integration of the cluster counting procedure was made into the High Level Trigger at CMS. A data compression of  $10^3$  was made by taking the raw level data from the silicon pixels and storing them in a simple data container, saving a peta-byte of data. Online luminosity using these data containers and methods are being investigated as more CMS-physicists are interested in using the central tracking system for luminosity measurements.

In addition to measurements, upgrade studies were also investigated at scenarios like  $3000 \text{ fb}^{-1}$ , which will be the target integrated luminosity of the High Luminosity Large Hadron Collider (HL-LHC). At the HL-LHC many upgrades to the detector will be made. In particular, the MIP timing detector upgrade [?] was investigated for performance in the  $\rightarrow \tau\tau$  analysis and results in luminosity measurements at the increased luminosity scenario. The timing upgrade will reproduce mean interactions - pile up - conditions like those that exist during Run II (2015-2018), making low occupancy methods of measuring luminosity possible at high pile up scenarios. Which will result in vast improvements to many physics analyses.

# Chapter 4

## Data and Simulation

### 4.1 Data and Simulation Used for Analysis

Data gathered at CMS is organized primarily on the trigger implemented to identify events. For this analysis, requiring muons is paramount and the search should be inclusive. Therefore datasets published by CMS that are considered are single muon, double muon, and electron plus photon datasets depending on the year. These datasets contain the triggers that are the most important for event selection. Single muon triggers that contain isolated muons at 24 and 27 GeV thresholds are implemented, along with double muon triggers with good reconstructed muons that have 17 GeV threshold. More information on triggers and selection will be given in the event selection section.

The simulation used to compare to data typically use MadGraph5@NLO along with a PYTHIA 8 hadronizer. These CMS-centrally generated samples are then digitized to the same format as real data events collected and processed at CMS high level trigger. These raw data formats are then reconstructed to physics objects - like tracks and higher level objects like muons and  $\tau$  leptons. A direct comparison between data and simulation can then be made after calibrating simulation in control regions.

Data taken from CMS during the entire Run II period was examined. This corresponds to 137  $\text{fb}^{-1}$  of integrated luminosity. The list for data and simulation Monte Carlo(MC) is exhaustive and listed in the appendix A.

For the MC production of the signal samples, to reflect the 2HDM modeling, events were

generated at tree level for a pseudoscalar Higgs like boson between the masses of 15 and 60 in intervals of 5 GeV with gluon fusion production. These masses are sufficient for the parametric modeling described in the fit to obtain a more precise peak resolution. The signal samples were privately produced for 2017 and 2018, because CMS MC generator contacts weren't sufficient in providing the samples in an acceptable manner. Sampled existed from the 2016 analysis. MadGraph5@NLO v2.6.5 was used to generate these events with a PYTHIA 8 hadronizer. Privately produced samples were used for 2017 and 2018. However, the scripts and conditions used are located here: <https://github.com/samhiggle/iDM-analysis-AODproducer/tree/haa>. The NMSSMHEt model was used to simulate the events. Parameters and information can be seen in the package: <https://cms-project-generators.web.cern.ch/cms-project-generators/>.

#### 4.1.1 $2\mu 2\tau$ Signal Samples

For the Monte Carlo production of the signal samples, to reflect the 2HDM modeling, events were generated at tree level for a pseudoscalar Higgs like boson between the masses of 15 and 60 in intervals of 5 GeV with gluon fusion production. These masses are sufficient for the parametric modeling described in the fit to obtain a more precise peak resolution. MadGraph5@NLO v2.6.5 was used to generate these events with a PYTHIA 8 hadronizer. Privately produced samples were used for 2017 and 2018. However, the scripts and conditions used are located here: <https://github.com/samhiggle/iDM-analysis-AODproducer/tree/haa>. The NMSSMHEt model was used to simulate the events. Parameters and information can be seen in the package: <https://cms-project-generators.web.cern.ch/cms-project-generators/>.

## 4.2 Corrections to Simulations

For accurate results that reflect true experimental data, many corrections to MC samples are made. In general, in compliance with CMS's Physics Object Groups (POGs), standard techniques are applied to ensure proper simulation. Of critical importance is the corrections to energy scales for the leptons in the analysis. These corrections will affect the nominal energy recorded for the event as well as the rates in which objects are identified. In order to protect against bias and to investigate systematic errors, corrections that could effect the results are considered in the overall

error in the statistical inference model.

### 4.2.1 Muon energy scale

Corrections to the muon's energy scale are computed for the muons that pass selection for the analysis. Medium muon ID with track based isolation that pass the isolated single muon triggers at 24 GeV and 27GeV are then rescaled for pileup, efficiency, di-lepton  $P_T$  and EWK reweighting of based on accurate gauge boson measurements. After selection, the scale factors for energy corrections are measured and prarmetrized in  $\phi$  and  $\eta$  in multiplicative and additive corrections

$$\rho^{\text{cor}} = \kappa(\eta, \phi)\rho + Q\lambda(\eta, \phi)$$

In practice these are just scale factors applied to the energy scale in certain eta phi regions.

Table 4.1: Measured energy scale correction for genuine across all years.

Correction (%)	
$\eta$ region	scale factor
0 – 1.2	0.4
1.2 – 2.1	0.9
> 2.1	2.7

### 4.2.2 Electron energy scale

Genuine electron energy scale and resolution requires corrections to be applied to MC in order to match data [?]. These corrections are provided directly by the E/Gamma POG, and applied to genuine electrons coming from tau lepton decays for the channels  $\mu\mu e\mu$  and  $\mu\mu e\tau$ .

The energy shift is split depending on the  $\eta$  of the electron4.2.

### 4.2.3 $\tau$ energy scale

There is a central energy shift based on the type of tau decay. Due to the electroweak interactions  $\tau$  leptons decay hadronically and leptonically. Figure 4.1 shows the typical tau decay leptonically

Table 4.2: Measured energy scale correction for genuine across all years.

Correction (%)	
$\eta$ region	scale factor
0 – 1.2	1.0
1.2 – 2.1	1.0
> 2.1	2.0

or hadronically.

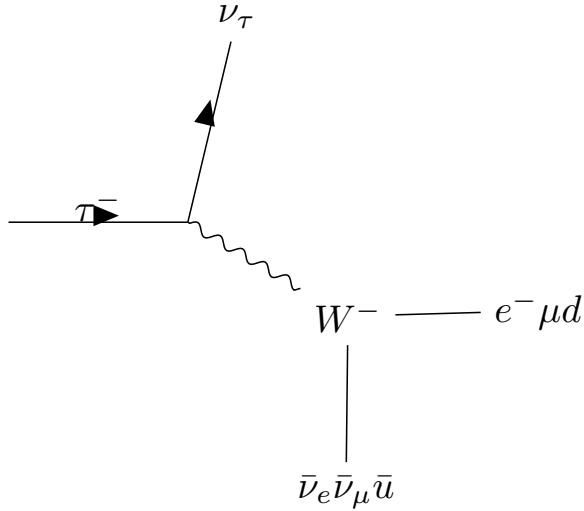


Figure 4.1: diagram depicting the possible decays of the tau lepton: 65% of the time hadronically and 18% to tau neutrino and electron with antineutrino (17% for muon)

When the tau decays hadronically, many different intermediate mesons are produced and each type of decay has different kinds of signatures particularly when they hadronize and deposit the energy within the hadronic calorimeter of CMS. The Tau-POG has measured the central value systematic deviation as a form of scalar factor that should be applied for an accurate result of measuring the tau's energy. This is split by the prongs (charged hadrons) and  $\pi^0$  s.

The deviation in the model is measured by between the difference in data and MC for different values of hadronic tau energy. The uncertainty is measured for each decay mode considered in the analysis. Figures 4.2 shows the differences in data and MC for the 1 prong +  $\pi_0$  decay mode as a function of tau energy, all other tau decay modes are also measured.

Table 4.3: Measured  $\tau_h$  energy scale correction for genuine  $\tau_h$ 's across all years.

Decay mode	Correction (%)		
	2016	2017	2018
$h^\pm$	-0.6	0.7	-1.3
$h^\pm\pi^0$	-0.5	-0.2	-0.5
$h^\pm h^\pm h^\pm$	0.0	0.1	-1.2

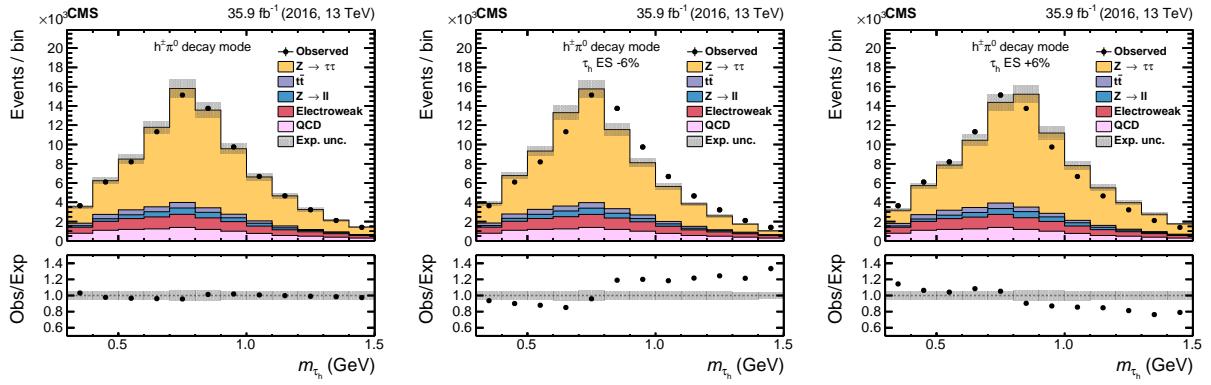


Figure 4.2: Tau mass distributions considering the nominal  $\tau_h$  energy scale in simulation (left), or the  $\tau_h$  energy scale shifted by -6 (left) or +6% (right), in the  $\mu\tau_h$  final state, for the 1 prong + pizero decay mode.

#### 4.2.4 $\tau_h$ ID efficiency

Genuine  $\tau_h$  identification efficiency can be different in Data and MC [?]. To correct for this difference, measurements are made in an inclusive  $\mu\tau$  channel, using genuine Drell-Yan to  $\tau_h$  as a signal and using the invariant mass of the  $\tau_h$  as an observable. Naturally, this region has far more statistics than the control and signal regions in the pseudoscalar analysis. To measure the identification efficiency precisely, it is done in the inclusive regions with an emphasis of simulation containing real taus. This measurement is done by the Tau POG, and the scale factors are provided to CMS.

For the  $\mu\mu e\tau_h$  and  $\mu\mu\mu\tau_h$  channels, scale factors are binned in 3 different  $p_T$  bins: 30–35, 35–40, and 40+ GeV. In the  $\tau_h\tau_h$  final state, the efficiencies are also binned by decay mode.

These efficiencies, while used in the primary event and the parameter of interest in the fit, they are not considered in the systematic model as they are expected to have very little impact on fit and limits based on the 2016 result.

#### 4.2.5 $\rightarrow \tau_h$ misidentification rate

In accordance to the TauPOG, the efficiencies associated with  $\tau$  identification through the Deep Neural Net approach is considered. The efficiency of the discriminators against electrons misidentified as  $\tau_h$  candidates can be mismodeled in simulation. These data/MC scale factors are binned by barrel/endcap region of the measured  $\eta_{\tau_h}$ , and by  $\tau_h$  decay mode. They depend on the anti-electron discriminator used. Scale factors are measured to correct this and are applied to  $\rightarrow \tau_h$  in MC.

The scale factors are obtained by separating events with a reconstructed electron and a reconstructed  $\tau_h$  passing the deep tau ID into a category where the  $\tau_h$  passes or fails the discriminator against electrons. All the processes in these regions are taken from simulation, except the QCD multijet background, which is estimated from SS data, and the W+jets background, which is normalized in a region with high  $m_T$ . The visible mass distributions in the pass and fail regions are fitted using the  $\rightarrow \tau_h$  scale factor as parameter of interest. An increase of the POI causes an increase of the signal yield in the pass region and simultaneously a decrease of the signal in the fail region, such that the signal yield in the two regions combined is unchanged. Additionally, the normalization of the  $\rightarrow$  background is freely floating to model a scale factor related to applying the discriminator against jets to electrons. The scale factor applied to electrons faking taus in the analysis is the product of the scale factor of the discriminator against jets and scale factor of the discriminator against electrons. The expected impact on the systematic model from these anti-lepton discriminators are expected to be very small so they are not included in the uncertainty model .

#### 4.2.6 $\rightarrow \tau_h$ misidentification rate

In accordance to the TauPOG, the efficiencies associated with  $\tau$  identification through the Deep Neural Net approach is considered similarly for muons. Efficiency at rejection of  $\rightarrow \tau_h$  can be different between data and MC[?]. Scale factors are measured to correct this and are applied to  $\rightarrow \tau_h$  in MC. They are binned in pseudorapidity and decay mode of the fake  $\tau_h$ .

The scale factors are obtained by separating events with a reconstructed muon and a reconstructed  $\tau_h$  passing the deep tau ID into a category where the  $\tau_h$  passes or fails the discriminator against muons. All the processes in these regions are taken from simulation, except the QCD multijet background, which is estimated from SS data, and the W+jets background, which is normalized

in a region with high  $m_T$ . The visible mass distributions in the pass and fail regions are fitted using the  $\rightarrow \tau_h$  scale factor as parameter of interest. An increase of the POI causes an increase of the signal yield in the pass region and simultaneously a decrease of the signal in the fail region, such that the signal yield in the two regions combined is unchanged. Additionally, the normalization of the  $\rightarrow$  background is freely floating to model the scale factor related to applying the discriminator against jets to muons. The scale factor applied to muons faking taus in the analysis is the product of the scale factor of the discriminator against jets and scale factor of the discriminator against muons. The expected impact on the systematic model from these anti-lepton discriminators are expected to be very small so they are not included in the uncertainty model.

#### 4.2.7 Pileup reweighting

MC events are reweighted using a minimum bias cross section equal to the luminosity for the corresponding year.

#### 4.2.8 Electron and muon identification efficiency

Scale factors derived within the HTT group are applied for muons [10], and the EGamma POG scale factors are applied for electrons [?]. The scale factors for muons with  $5 < p_T < 9$  GeV and  $9 < p_T < 10$  GeV, are computed privately as there are no official numbers, and are approved by the MUON POG.

#### 4.2.9 Generator event weights and luminosity

Generator weights are applied on an event-by-event basis. Samples produced with the aMCatNLO generator contain both positive and negative event weights. The presence of negative event weights reduces the effective statistics of the samples. The event weights for simulation are scaled to the expected yields for each sample. The number of generated events in each sample is used, however in the aMCatNLO sample this sum of generated events is weighted by the generator weights, effectively making the aMCatNLO samples smaller when weighting for luminosity and cross section. Additionally, K-factors are considered for W+Jets and Drell-Yan samples in order to correct between purely generated events and reconstructed event yeilds for Drell-Yan a factor of

1.1637 is used and for W+Jets a factor of 1.221 is used. These are applied to each exclusive and inclusive samples during the combination of inclusive and exclusive sample processing.

#### 4.2.10 Offline Muon Selection

Due to the selection of muons at 5 GeV, which is below the trigger threshold, scale factors were measured in the barrel and endcap using the tag and probe technique in the 2016 analysis. These factors are considered, in addition to the Muon POG's recommendation, to support correct simulation of data.

Table 4.4: Scale factors to correct for Offline muon selection being less than the online trigger threshold.

	Barrel	Endcap
Muons with $5 < pT < 9$ GeV	0.956	0.930
Muons with $9 < pT < 10$ GeV	0.916	0.897

#### 4.2.11 Visualizing the Corrections

The energy scales for the various leptons used in the analysis are not only changed in the nominal case, but their uncertainty is measured and then propagated to the fit model via changes in the normalization for the distribution. To visualize this and provide a cross check distributions for the  $\tau \mu$  and e energy scales are cross checked in the parameter of interest (the mass of the di-muon system from the leading pt-pseudoscalar "a" particle). These systematics for  $\mu\mu\tau\tau$  channel are visualized here in figure 4.3.

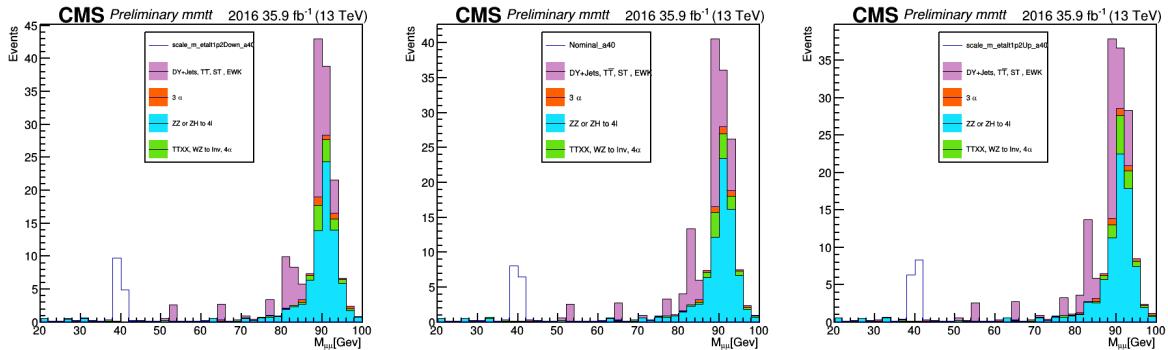


Figure 4.3: Systematic Shift in the Uncertainty Model for 2016  $\mu\mu\tau\tau$  for the muon energy scale shift down (left), nominal (mid), and up (right), no data is shown on this plot as it directly reflects the signal region without the extraction cuts

In the end, the uncertainty considered in the fit model would then be the percent yield up and down as a log-normal that affects the normalization of the template. Because the model is so statistically limited the log-normal is sufficient in capturing the changes in the distributions over the fit range.

# Chapter 5

## Event Selection

### 5.1 Framing an Analysis at CERN

In order to make a definitive statement and concrete hypothesis test, two perspectives are made. There is the null hypothesis, ergo the hypothesis of the accepted standard within High Energy Physics - the Standard Model - which will be comprised of all possible events that are categorized as *background*. Then the alternative hypothesis is taken to be the model that encompasses the Standard Model and adds additional physics that will contain events from the theory under consideration in the analysis *signal*. For this analysis, it's events pertaining to the pseudoscalar Higgs-like particle "a", with two muons and two tau leptons in the final state.

### 5.2 Defining Signal and Control Regions

After obtaining sufficient motivation to conduct a search or measure particular processes, the next thing to consider is what data should be used to test that hypothesis. Working with CMS, central Standard Model simulation samples are provided and should be used along with the recommendation from the sub-groups of analyzers to ensure that the simulation can be used to compare to data. Then to make the analysis competitive, regions of the full data and simulation are cut away in order to increase the number of signal events relative to background events. This process is known as "making cuts". In order to not bias the result apriori, typically regions are setup to investigate the agreement of simulation to data (the control region) and to conduct the statistical hypothesis

test (signal region). When the data and MC simulation agreement are reasonable in the control region then the statistical test can be made in the signal region.

### 5.2.1 Optimizing Lepton Pair Selection

In order to maximize the signal event yield - as it is expected that the analysis is statistically limited - a simple selection algorithm was used in order to identify good lepton pairs that come from the pseudo scalar  $a$ . Standard working point cuts are made - in addition to delta R cleaning - two prompt-like muons with opposite charge and leading scalar summed  $P_t$  are chosen to form the first decay products of the  $a$  and two opposite charged leading scalar summed  $P_t \tau$  leptons are chosen to for the second  $a$ . This choice increased the signal acceptance compared to choosing mass window cuts to form the  $a$  pairs. The follow table reflects the efficiency done on the preliminary dataset from 2016 NanoAODv6:

$a$ - Mass	Efficiency
M15	0.87
M20	0.82
M25	0.79
M30	0.79
M35	0.79
M40	0.80
M45	0.80
M50	0.83
M55	0.85
M60	0.87

All selections follow the Physics Object Group (POG) recommendations for object selection. Since the search in 2016, a new identifier for  $\tau$  leptons was considered using a Deep Neural Network (DNN). To identify the dimuon pair which come from one pseudoscalar  $a$ , several cuts were applied across all channels:

- each  $\mu_{pt} > 5.0$  GeV,
- charge dimuon  $< 0$

- all  $\mu_{iso} <= 0.2$
- medium ID of the  $\mu$ , require global and track requirements (good  $\mu$ )
- number of b-quark tag jets  $< 1$
- signal extraction cuts (not shown in data MC control plots)
  - invariant mass of the 4 lepton system (AMass)  $< 120\text{GeV}$
  - $M_{\mu\mu} > M_{\tau\tau}$

The trigger requirements are inclusive, selecting events that pass single muon and double muon triggers. Single muon triggers contain muons that are isolated with 22, 24, and 27 GeV muons. Double muon triggers have a 17 GeV threshold for the leading muon and 8 GeV for the subleading. Triple muon triggers are used for the  $\mu\mu\mu\tau$  and  $\mu\mu e\mu$  channels and have a descending threshold of 12, 10, and 5 GeV. These triggers and other minimal selection requirements are listed in table 5.1.

Table 5.1: Event selection requirements for the four decay channels.

Channel	Trigger requirement	Minimal lepton selection		
		$p_T$ (GeV)	$\eta$	Isolation
$\mu\mu\tau_h\tau_h$	$\mu[22] \mu[24] \mu[27]$ or $\mu\mu[17, 8]$	$p_T^{\tau_h} > 18.5$	$ \eta^{\tau_h}  < 2.3$	Med. DNN $\tau_h$
$\mu\mu\mu\tau_h$	$\mu[22] \mu[24] \mu[27]$ or $\mu\mu[17, 8]$ or $\mu\mu\mu[12, 10, 5]$	all $p_T^\mu > 5$ $p_T^{\tau_h} > 18.5$	$ \eta^\mu  < 2.3$	$I^\mu < 0.2$
$\mu\mu e\tau_h$	$\mu[22] \mu[24] \mu[27]$ or $\mu\mu[17, 8]$	dimuon $p_T^\mu > 5$ $p_T^e > 7$ $p_T^{\tau_h} > 18.5$	$ \eta^e  < 2.5$ $ \eta^{\tau_h}  < 2.4$	$I^e < 0.15$ Med. DNN $\tau_h$
$\mu\mu e\mu$	$\mu[22] \mu[24] \mu[27]$ or $\mu\mu[17, 8]$ or $\mu\mu\mu[12, 10, 5]$	$p_T^e > 7$ $p_T^\mu > 5$	$ \eta^e  < 2.4$	$I^e < 0.15$

As mentioned in the instrumentation and detector section, the muon identification system uses the tracker to identify charged tracks and the muon chambers identify the particles later in their trajectory after they exit the solenoid. Typically "good" muons are those that are both associated with a track and their subsequent decay in the drift tubes or the CSC chambers. The physics object group's recommendations are followed, which select muons with  $p_T > 15\text{ GeV}$  and  $|\eta| < 2.4$  in addition to selecting only "good" muons.

Electrons originating from the tau decay are reconstructed in by track association in tandem with energy deposition in the Electronic Calorimeter (ECAL). Events are not favored or vetoed for candidate electrons that also show a substantial energy deposition in the HCAL. The hits and

quality of the tracks from two separate algorithms, along with the geometrical and energy matching from the ECAL are used in a Multivariate Analysis (MVA) technique like Boosted Decision Trees (BDTs) to qualify good electrons for analysis [11].

Of critical importance in event selection is identification of unique particles, particularly differentiating between lepton candidates that come from process at the interaction vertex (prompt) and those that appear from decays down the line (nonprompt). Relative isolation is typically defined in order to ensure no overlap in prompt and non-prompt candidates.

$$I^\ell \equiv \frac{\sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutral}} p_T - \frac{1}{2} \sum_{\text{charged}} p_T\right)}{p_T^\ell}. \quad (5.1)$$

$\sum_{\text{charged}} p_T$  is the scalar sum of the transverse momenta of the charged particles originating from the primary vertex and contained in a cone of size  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$  (0.3) centered on the muon (electron) direction. The sum  $\sum_{\text{neutral}} p_T$  is a similar quantity for neutral particles.

To identify  $\tau_h$  candidates, the hadron-plus-strips (HPS) algorithm is used to identify the major modes of the hadronic tau decay. Typically events with hadronic prong are considered in combination with a number of neutral pions and missing transverse energy from the neutrinos.  $\pi^0$  almost always decay to photons, so an algorithm is considered that joins the identification of charged hadrons and neutral pions. The HPS algorithm combines inner track information, hits in the HCAL, and pion association in the ECAL by deposits in a  $\eta, \phi$  strip region. The  $\tau_h$  is matched to  $h^\pm, h^\pm\pi^0, h^\pm h^\mp h^\pm$ , or  $h^\pm h^\mp h^\pm\pi^0$  depending on the overall charge vs neutral constituents [12]. In addition to HPS algorithms, a Deep Neural Network (DNN) was constructed to further aid in identification by discriminate between genuine tau leptons and those that originate from quark or gluon jets, electrons, or muons. In the DNN, the tau four-momentum and charge, the number of charged and neutral particles used to reconstruct the tau candidate, the isolation variables, the compatibility of the leading tau track with coming from the primary vertex, the properties of a secondary vertex in case of a multiprong tau decay, observables related to the  $\eta$  and  $\phi$  distributions of energy reconstructed in the ECAL strips, observables related to the compatibility of the tau candidate with being an electron, and the estimated pileup density in the event are all used. In total, 47 high-level input variables are incorporated [13]. In practice, the DNN has discriminators for muons, electrons, and jets that fake genuine taus and have performances that go from 40% to 90% in a 10% granularity. The tau POG medium working point is used for each of these discriminators.

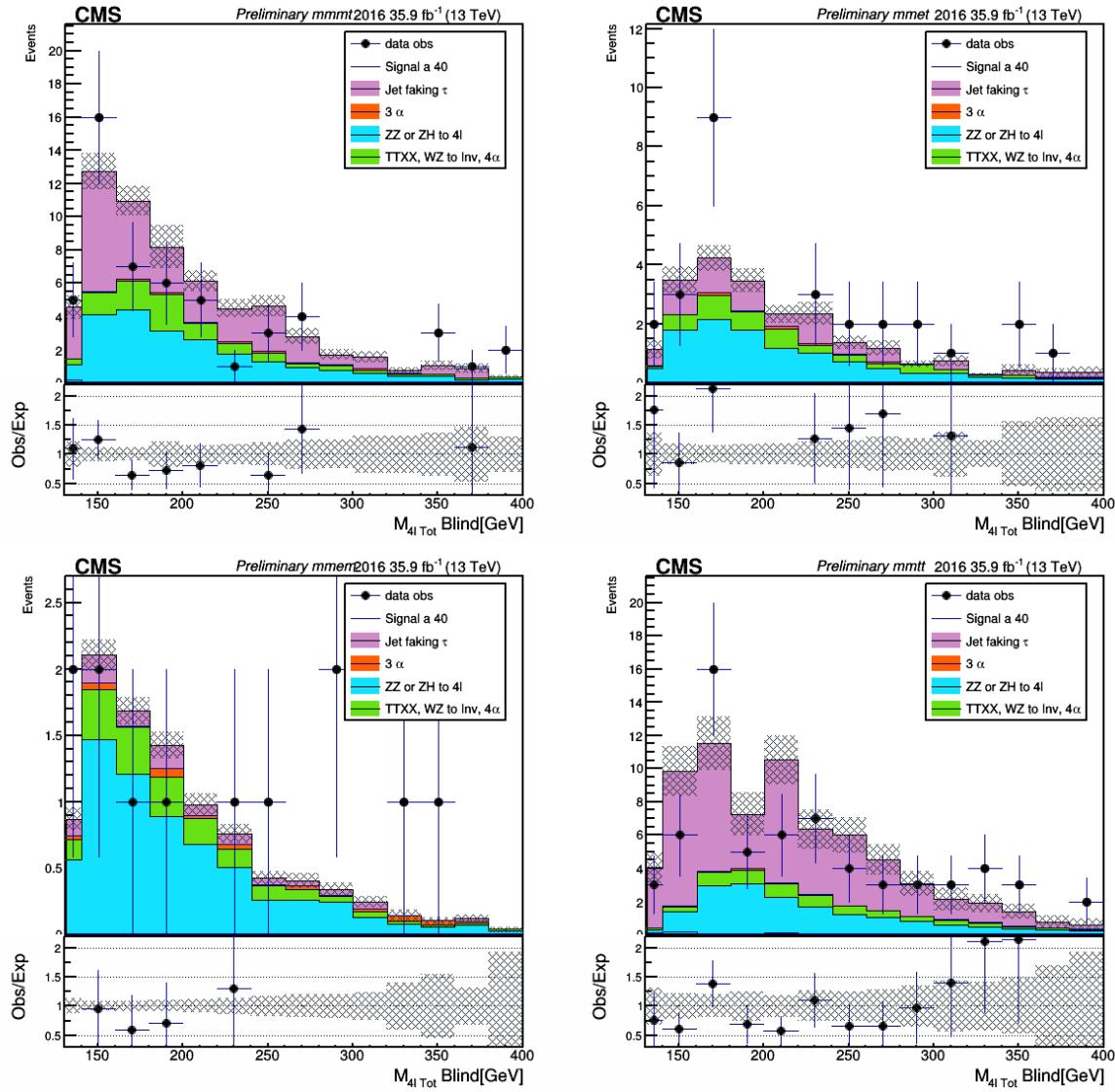


Figure 5.1: Invariant Mass of the four lepton system for 2016 data in all Channels

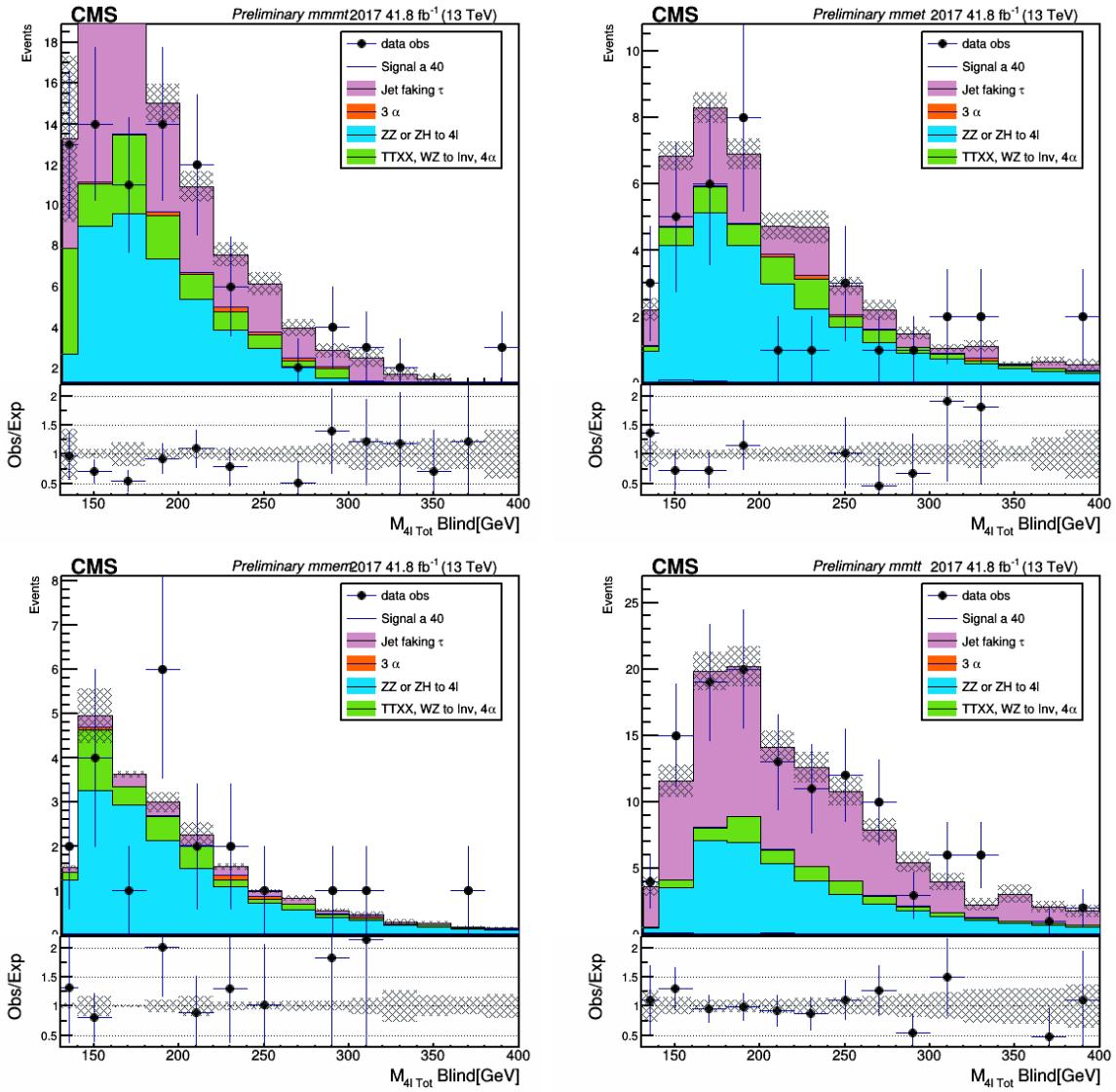


Figure 5.2: Invariant Mass of the four lepton system for 2017 data in all Channels

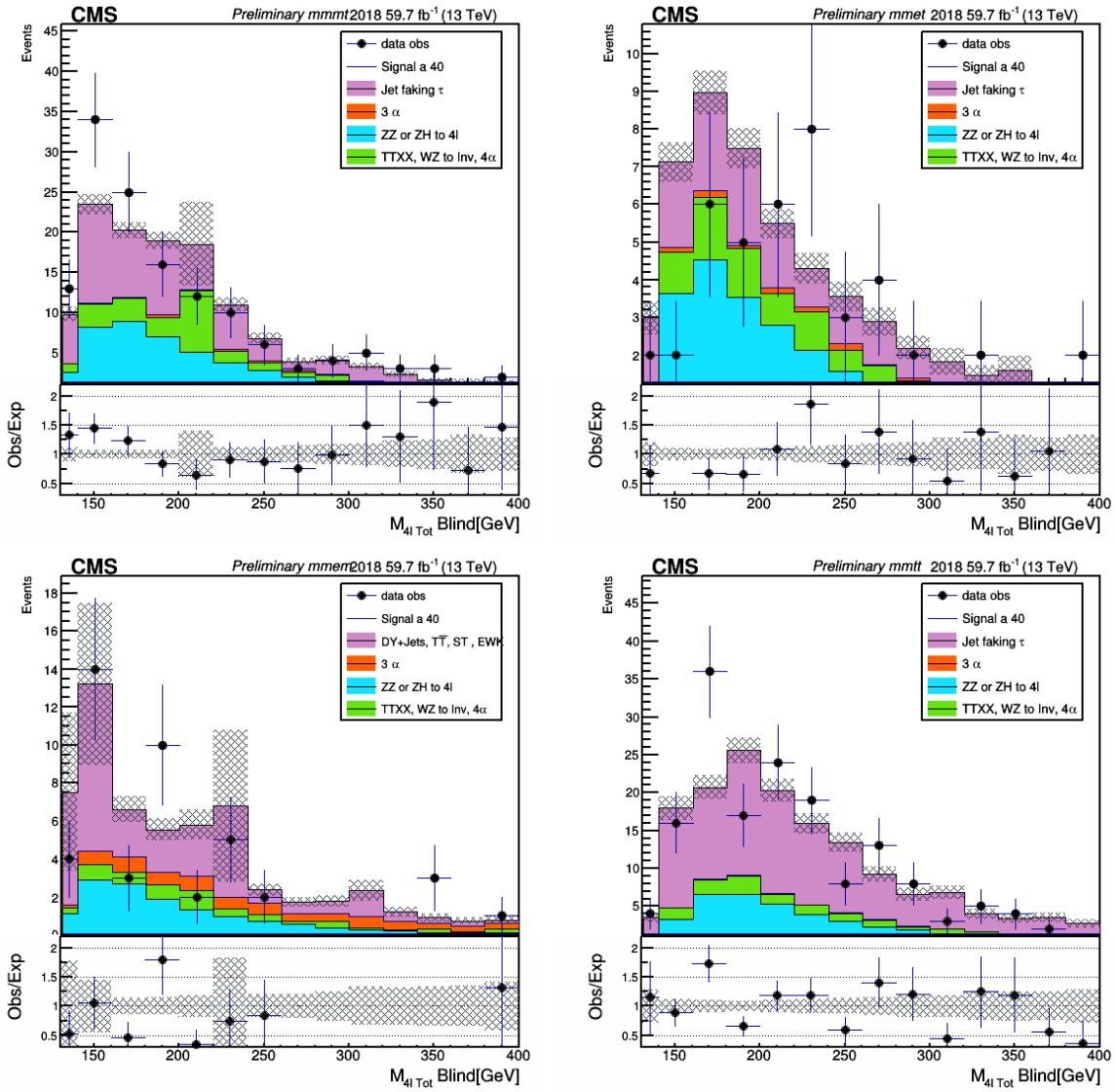


Figure 5.3: Invariant Mass of the four lepton system for 2018 data in all Channels

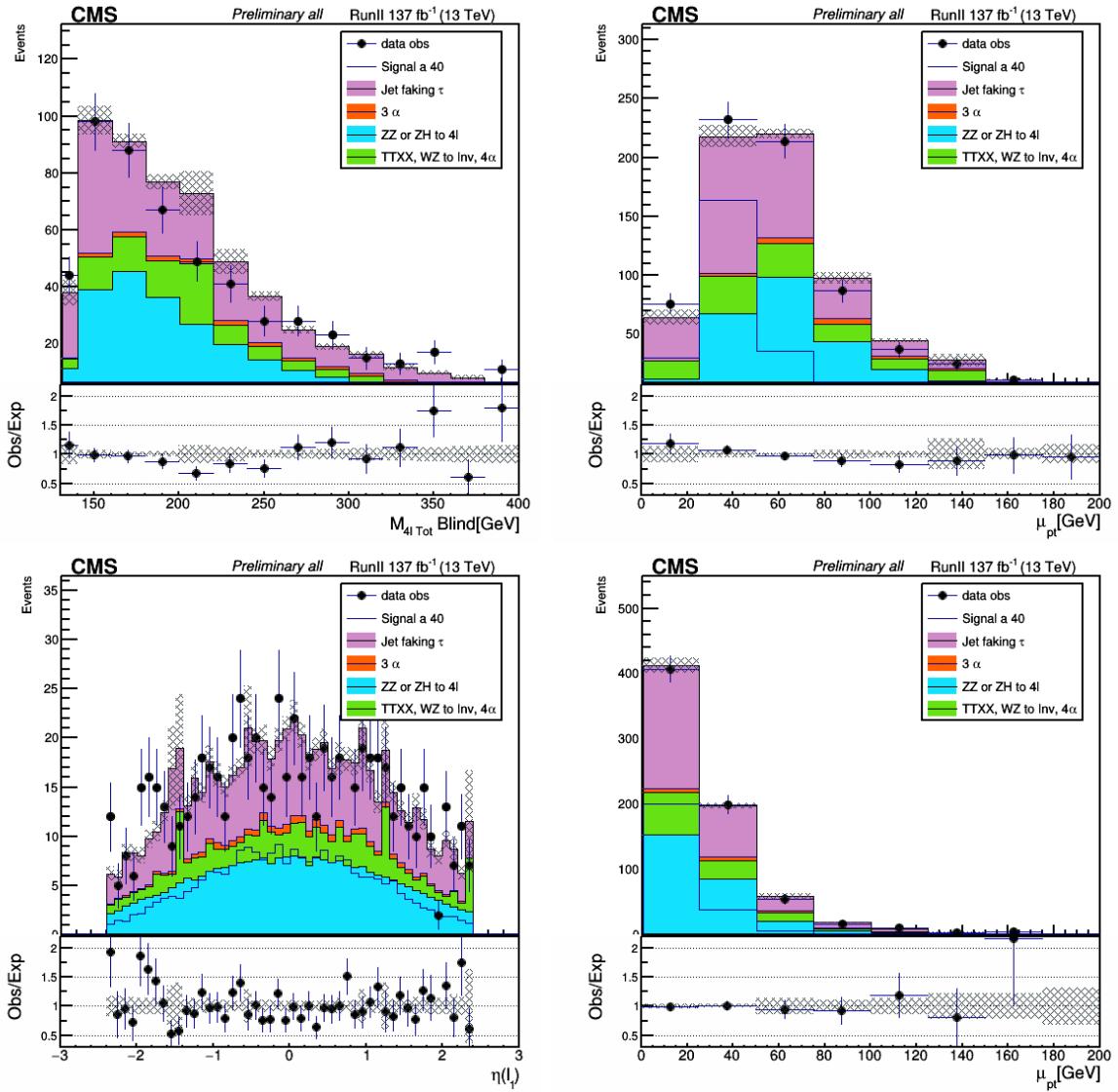


Figure 5.4: Several Data-MC control plots for full RunII data in all Channels

# Chapter 6

## Background

In addition to the Monte Carlo simulation, a datadriven method is used to estimate a significant portion of background that simulation alone is not sufficient to measure.  $\tau$  leptons decay hadronically about 65% of the time forming intermediate mesons. The jets that these decays produce are very similar to Quantum Chromodynamic (QCD) processes. These QCD events effectively fake the hadronic  $\tau$  signature for the  $a$  decay. In order to conduct the datadriven method, an "ABCD" region proportionality method is implored.

### 6.0.1 Brief Outline of the Fake Rate Method

The general outline to measure the fake rate is as follows:

- FakeRate Function in Same Sign region is *known*
- Events passing loose identification in Opposite Sign region is *known*
- Events passing signal region is *unknown*
- For channels with enough statistics in the O.S. Tight region, data – prompt MC is conducted. This is motivated by estimating the true jet faking taus background (non-prompt taus). If MC is matched to prompt then it is unlikely a jet faking tau – so they are removed

$$f_r(pt) = \frac{\text{Data Events S.S. Tight - Prompt MC Background}}{\text{Data Events S.S. Loose - Prompt MC Background}}$$

- After the measurement is made for each leg, then apply that as weight to the Opposite Sign "tight" region (the signal region) for one leg at a time weight

- The Fake Rate function is parametrized in lepton candidate transverse momentum

### 6.0.2 Measurement of the Fake Rate

To measure the fake rate multiple categories are entertained. As outlined in the SM Higgs decays to tau leptons analysis and it's supporting document on fake rate measurements [10], several regions are used to determine the fake rate. The separate "enriched" background regions are considered

- QCD multijet (large majority of jet $\rightarrow \tau_h$  fake events in the  $\tau_h \tau_h$  final state)
- W+jets (mostly in the e $\tau_h$  and  $\mu\tau_h$  final states)
- tt events with fully-hadronic or semi-leptonic decays (mostly in the VBF category)
- diboson events with fully-hadronic or semi-leptonic decays.

These are then measured as a function of  $P_T$  of the object and split into sub categories depending on the decay mode.

In the  $t\bar{t}$  isolated region, the  $t\bar{t}$  background with a fake  $\tau_h$  is only a small fraction of the total number of events, which leads to large uncertainties when other processes are subtracted from data. Therefore we choose to instead use the  $t\bar{t}$  simulation to estimate the fake rates. We will see later that this gives generally compatible results with what would be obtained from data, and the  $t\bar{t}$  background is anyway a small fraction of the total fake background in the signal region of the analysis.

In the QCD multijet region, all backgrounds estimated from MC simulation are subtracted from the data in both regions to obtain the effective QCD contribution. In the W+jets region, all backgrounds estimated from MC simulation except for the W+jets simulated process are subtracted from the data in both regions to obtain the effective W+jets contribution. In the  $t\bar{t}$  region, all backgrounds estimated from MC simulation except for the  $t\bar{t}$  events with a jet misidentified as a  $\tau_h$  candidate are subtracted from the data in both regions to obtain the effective  $t\bar{t}$  with fakes contribution.

The fake rates are fitted as a function of the  $\tau_h p_T$  with a linear function. The fits are shown in Fig. B.1 (Fig. B.4) for the W+jets region with 0 jet, in Fig. B.2 (Fig. B.5) for the W+jets region with one jet, in Fig. B.3 (Fig. B.6) for the W+jets region with more than one jet, in Fig. B.7

(Fig. 6.1) for the QCD multijet region with 0 jet, in Fig. B.8 (Fig. B.10) for the QCD multijet region with 1 jet, in Fig. B.9 (Fig. B.11) for the QCD multijet region with more than 1 jet, and in Fig. B.12 (Fig. B.14) for the  $t\bar{t}$  region in data, in Fig. B.13 (Fig. B.15) for the  $t\bar{t}$  simulation in the  $\tau_h$  ( $\tau_h$ ) final state. As shown in the last two figures, the fake factors for  $t\bar{t}$  events are compatible when measured in the simulation and in data, even if the last suffers from large uncertainties.

Because of the lack of statistics at high  $\tau_h p_T$ , the fact that a decreasing linear function inevitably leads to negative fake factors at high enough  $\tau_h p_T$ , and the fact that our signal typically does not lie at very high  $\tau_h p_T$ , we take the fake factors evaluated at a  $\tau_h p_T$  of 100 GeV for any  $\tau_h$  with  $p_T \geq 100$  GeV. Fake factor measurements for the  $\mu\mu\mu\tau$  channel in the QCD 0jet region is included in figure 6.1, the rest of all the regions are included in the appendix B.

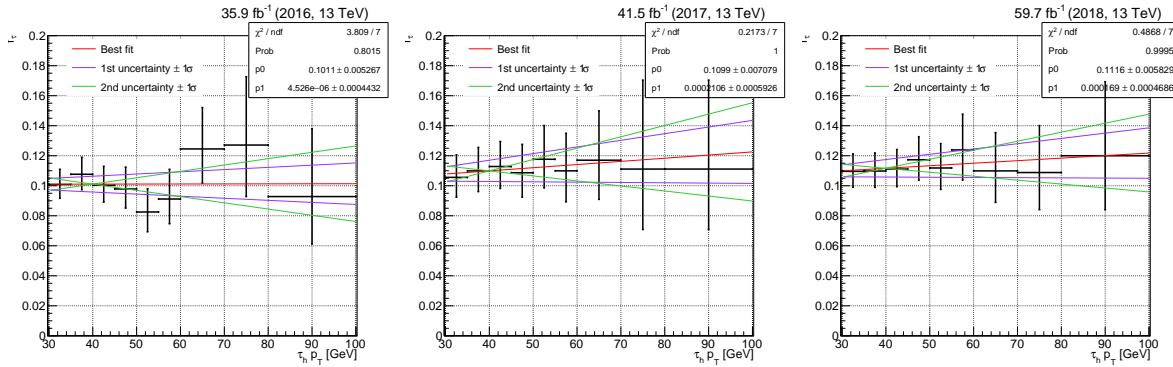


Figure 6.1: Fake factors determined in the QCD multijet determination region with 0 jet in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

### 6.0.3 Application of the Fake Rate Method

After the jet faking tau rate is measured, it is then applied to events that identify as loose. Due to the final state involving two tau leptons, this procedure is applied to each tau lepton in the final state thus requiring application of the fake rate to four different possibilities. Each lepton "leg" may identify as loose or failing the tight identification. Alternately they could pass. The fake rate is then applied depending on the identification for each leg and in the case the event fails both legs then a minus sign is included to avoid the case of double counting.

- The final weight is then applied depending on the pass and fail criteria of each lepton candidate
- If event fails id for leg 1:

$$f_1(pt) = \frac{f_{r_1}(pt)}{1 - f_{r_1}(pt)}$$

- If event fails id for leg 2:

$$f_2(pt) = \frac{f_{r_2}(pt)}{1 - f_{r_2}(pt)}$$

- If event fails id for both:

$$f_{12}(pt) = -\frac{f_{r_1}(pt)}{1 - f_{r_1}(pt)} \cdot \frac{f_{r_2}(pt)}{1 - f_{r_2}(pt)}$$

To illustrate the different regions please consider

This fake factor methodology has been supported by other analyses such as the standard model Higgs measurement with an associated Z boson [14].

For a closure test, the same application criteria is applied to the selection of the tight same sign region. The vast majority of the background should be jets faking taus in that case. Indeed it is shown in figure 6.3.

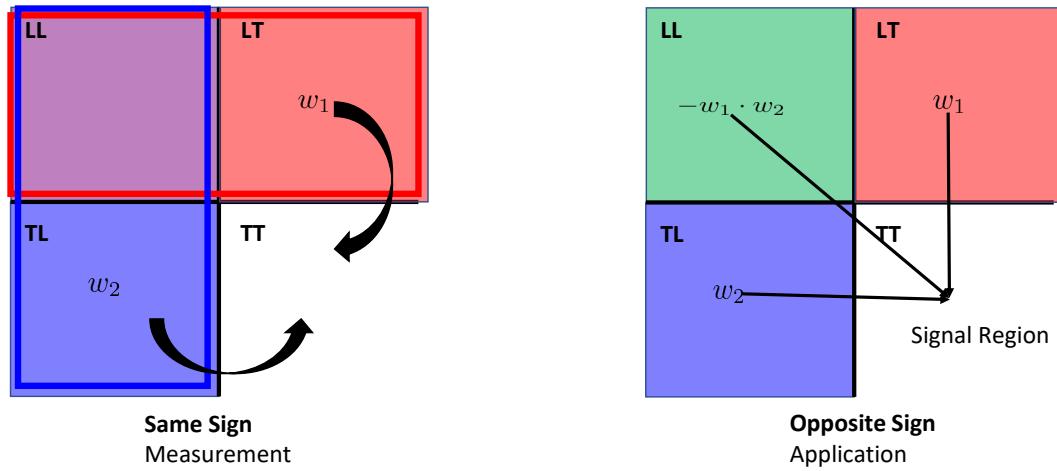


Figure 6.2: Diagram depicting the measurement and application regions this ABCD method is multiplied by the two  $\tau$  leptons in the final state

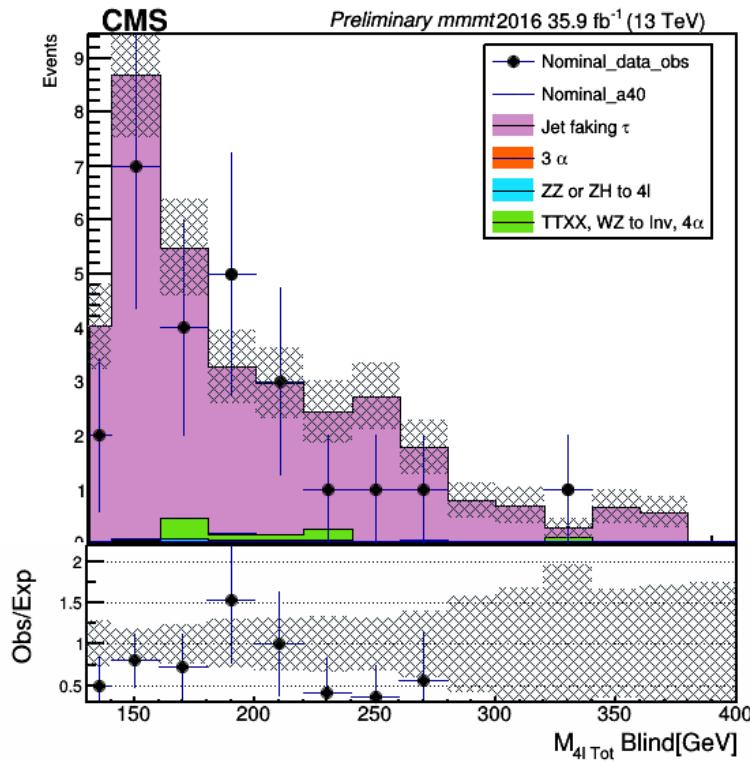


Figure 6.3: Validation of the fake factor method, fake factors are applied to the same sign tight region

# Chapter 7

## Statistical Inference Modeling

### 7.1 Statistical Inference at CERN

In order to conduct a hypothesis test, a test statistic is needed. This is constructed through the *profile likelihood ratio* and then a confidence level is set using the test statistic [15]. So construction of a likelihood is required, the test statistic from the likelihood, and a way to calculate a confidence level to complete the hypothesis test. To construct the likelihood, the typical approach is to assume a Poisson distribution for the events in the  $i$ -th bin and then "smear" it by multiplying it with a Gaussian that is also dependent on the events. The events are split, into signal and background. Because an alternative hypothesis is explored and a-priori the amount of signal events isn't known, it is floated by a coefficient denoted as  $\mu$  - the signal strength. For an upper limit this value is changed until the cumulative distribution function of the likelihood reaches the desired confidence or p-value.

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\text{data}, \mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta)$$

For binned  $n_i$  events in bin label  $i$ :

$$\text{Poisson}(\text{data}, \mu \cdot s(\theta) + b(\theta)) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i} \quad (7.1)$$

In the case of the pseudoscalar analysis discussed in this paper a slightly different likelihood is considered. A shape-based approach is considered, which instead of a product of many simple Poisson probability density functions, probability distribution functions are used over the entirety

of the fit variable (or parameter of interest POI). Multiple PDFs can be combined in this scenario, but the important difference is the absence of any binning. Therefore a good fit is required to have a good description of the events.

Unbinned or Parametric likelihood functions over  $k$  events,  $S$  and  $B$  the total event rate,  $f_s(x_i)$  and  $f_b(x_i)$  the *pdfs*

$$\text{Poisson}(\text{data}, \mu \cdot s(\theta) + b(\theta)) = k^{-1} \prod_i (\mu S f_s(x_i) + B f_b(x_i)) e^{-\mu S - B}$$

Specifically for the pseudoscalar analysis, a product of Bernstein polynomials for  $f_b(x_i)$  and Voigtian functions for  $f_s(x_i)$  are used. The Bernstein captures the slow changing background that is basically constant in a 2GeV mass window and the Voigtian captures the sharp peaking dimuon mass signal. Details in the analysis will be saved for after the overview of the stat inference model.

To outline the typical approach in the profile likelihood method, the follow steps are done in an effort to set the limit

1. To form the test statistic, the *profile likelihood ratio* is used

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}_{\max}(\text{data}|\hat{\mu}, \hat{\theta})} \quad 0 \leq \hat{\mu} \leq \mu$$

Where  $\hat{\mu}$  and  $\hat{\theta}$  are the maximum likelihood estimators.

2. Find the *observed* value of the test statistic  $\tilde{q}_\mu^{\text{obs}}$  for given signal strength  $\mu$ .
3. Find values of the nuisance parameters that best describe the experimentally observed data by maximizing the likelihood.
4. Generate toy MC pseudo data to construct probability density functions for signal and background  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu)$  and  $f(\tilde{q}_\mu|0, \hat{\theta}_\mu)$  for background only hypothesis. These pdfs are the pdfs of the test statistic under the assumption of a signal strength.
5. Generate toy MC pseudo data to construct probability density functions for signal and background  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu)$  and  $f(\tilde{q}_\mu|0, \hat{\theta}_\mu)$  for background only hypothesis.
6. Define p-values to be associated with the actual observation for both s+b and b only hypotheses

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | s+b) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_b^{obs} | b) = \int_{\tilde{q}_b^{obs}}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}_b^{obs}) d\tilde{q}_\mu$$

Then take the ratio to form the confidence levels

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b}$$

7. Let  $\alpha$  be the measure of confidence, then for  $CL_s \leq \alpha$  then the signal hypothesis is rejected in favor of the background only hypothesis
8. Further to quote the 95% confidence level on  $\mu$ , we need to adjust the signal strength ( $\mu$ ) , until  $CL_s = 0.05$

## 7.2 Worked Example for Low Stat Analyses

Order of magnitude estimates for low stat analyses like the Higgs decay to pseudoscalars can be obtained by considering a very simple statistical inference model.

Suppose there are  $N$  events

$$N = B \cdot \sigma \cdot A \cdot \mathcal{L} \quad (7.2)$$

where  $B$  is the branching fraction for the physics process,  $A$  is the signal acceptance,  $\sigma$  is the cross section, and  $\mathcal{L}$  is the luminosity.

Thus given a model that has the dependent variables, the upper 95% limit on the branching fraction can be found.

In general 7.2 works; however, a likelihood model is needed to fully get the estimate of the number of events expected.

As outlined in binned-likelihood models in CMS [15], suppose that the number of expected events follows a Poisson distribution. Then in that case if the background model would predict 0 average events, then the upper 95% bound in that case is 3.7 events.

Inverting the relation

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}}$$

Selecting the signal acceptance from the pseudoscalar analysis (2016  $\mu\mu\mu\tau$ )

$$A = \frac{\text{events pass all cuts}}{\text{starting events}} = \frac{1293}{250000} \approx 0.005$$

and taking the cross section of the gluon gluon fusion for the parent particle in the decay chain  $\sigma = 48\text{pb}$  along with the luminosity for 2016  $\mathcal{L} = 35,900\text{pb}^{-1}$

Then the upper 95% limit on the branching fraction is

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}} = 0.00043 = 4.3 \times 10^{-4}$$

### 7.3 Fit model for pseudoscalar Higgs search

After the signal extraction cuts are applied, an unbinned parametric likelihood fit was done with various shapes depending on the background and signal categorization. There is one signal distribution depending on the hypothesized  $a$  mass, and two background distributions that are considered in the final fit. The two contributing background distributions originate from "Irreducible" events coming from two Z bosons (ZZ) and from "Reducible" events coming from jets faking tau leptons (FF).

For the signal, a Voigtian function is used to fit the pseudoscalar  $a$  mass spectrum in a small window - 2GeV - of the hypothesized  $a$  mass for the sample as in figure 7.2. The Voigtian shape was chosen to reflect the narrow simulated peak that is statistically smeared by experimental measurement. The Voigtian shape has one extra degree of freedom compared to the Gaussian. This parameter controls the Lorentzian factor in the Voigtian. For the signal MC the distributions tend to increase in standard deviation as the mass approaches 60 GeV. To compare the signal MC distributions, they are all plotted in figure 7.1.

In order to obtain precise signal resolution, shapes from the signal samples in intervals of 5 GeV across the whole fit range 20-60 GeV are interpolated using spline functions. Thus precise limits can be obtained at the 1 GeV granularity. Predicated upon the fit model, using the spline in a refit would produce results that are close to the original fit to signal ensuring that in the statistical inference model that the signal is well modeled for all mass points. A spline function is necessary for each degree of freedom in the model. An example of such functions are shown in figure 7.3. The bands that envelope the spline indicate the spread and the accepted error on the spline in the statistical inference model, more details are discussed in the systematics section.

For the irreducible background coming from  $ZZ \rightarrow 4l$ , a Bernstein polynomial is used to fit the

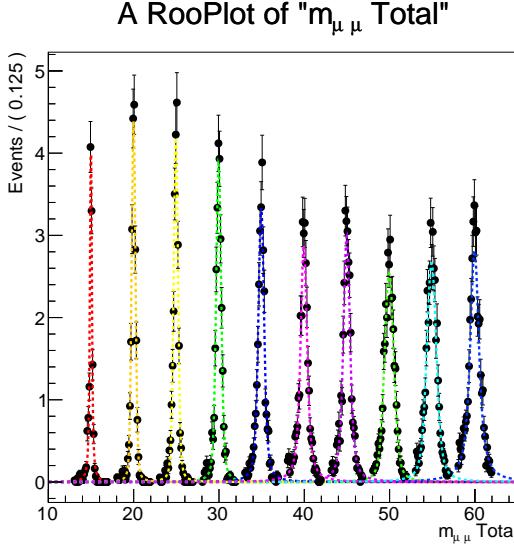


Figure 7.1: Signal fit using a Voigtian function

shape over the entire  $a$  mass range in figure 7.7. Depending on the final state and shape the degree is chosen to reflect accurate errors in the model. A Fischer F-test was conducted and there doesn't seem to be enough statistics in the bins to provide an accurate difference in the log-likelihood in order to recommend a leading degree. Thus for  $\mu\mu\tau\tau$  and  $\mu\mu e\tau$ , a linear fit is used. For the channels that do have more events like  $\mu\mu\mu\tau$  and  $\mu\mu e\mu$  - albeit even for a lower number of integrated events - a parabolic order is used. Plotted error bands reflect a simulation of 315 events in a RooFit simulation and is only meant to visualize possible 1 sigma differences in the shape. The true value of the error estimation on the parameters are take from the fit itself and can be seen in the plots, and are later shown in the impacts which demonstrate how the fit parameters effect the overall statistical inference model C.1.

For the jet faking tau background, a Bernstein polynomial is also used to fit the shape over the entire  $a$  mass range in figure 7.8. As with the ZZ - irreducible background - the jet faking tau background polynomial degree is chosen to reflect accurate errors in the model.

The rest of the channels and years are shown below.

Before a presentation of the expected limits on the branching fractions is given, a view of the uncertainty model is discussed here. In order to measure the systematic effects on the final distributions and the fits, changes in the fit templates are done and propigated to the fit model in the form of rate parameters. This is the accepted method to measure systematic impacts with

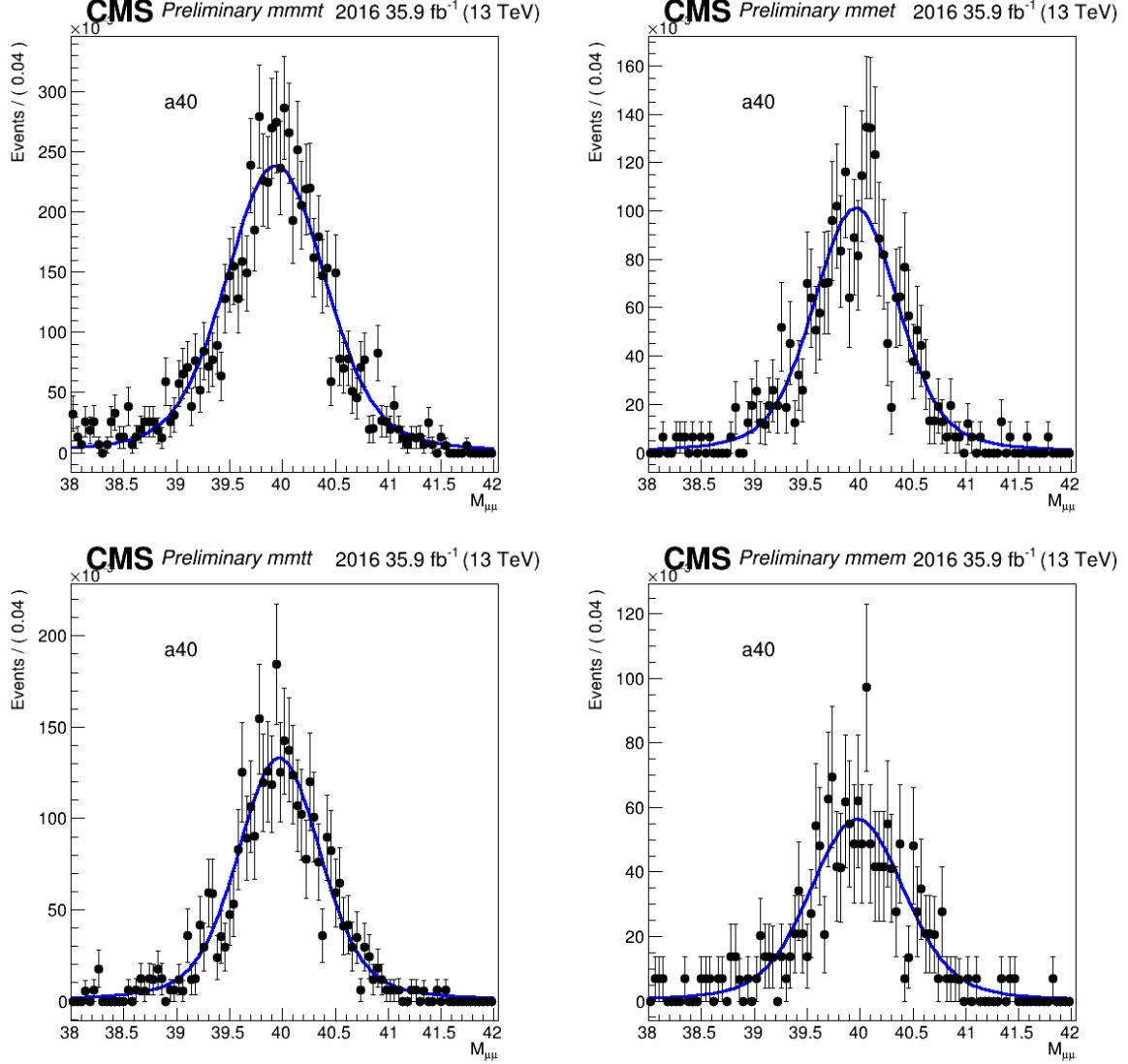


Figure 7.2: Signal fits using a voiigtian function for a-mass at 40GeV

a parametric fit model. These rate parameters differ slightly between the signal and background distributions. For background the error in the fit parameters are directly included in the uncertainty model.

For the signal, the uncertainty on the spline function is considered. As mentioned in the fit model section and shown in figure 7.3, the magnitude of this uncertainty is estimated from the fit of the parameters for the spline. Overall a 10% uncertainty is used for the lorentzian (alpha) and 20% for the standard deviation (sigma) and 0.5% for the mean (mean). Although the mean is measured very precisely, the energy scale shifts from the leptons are included in this figure. Please

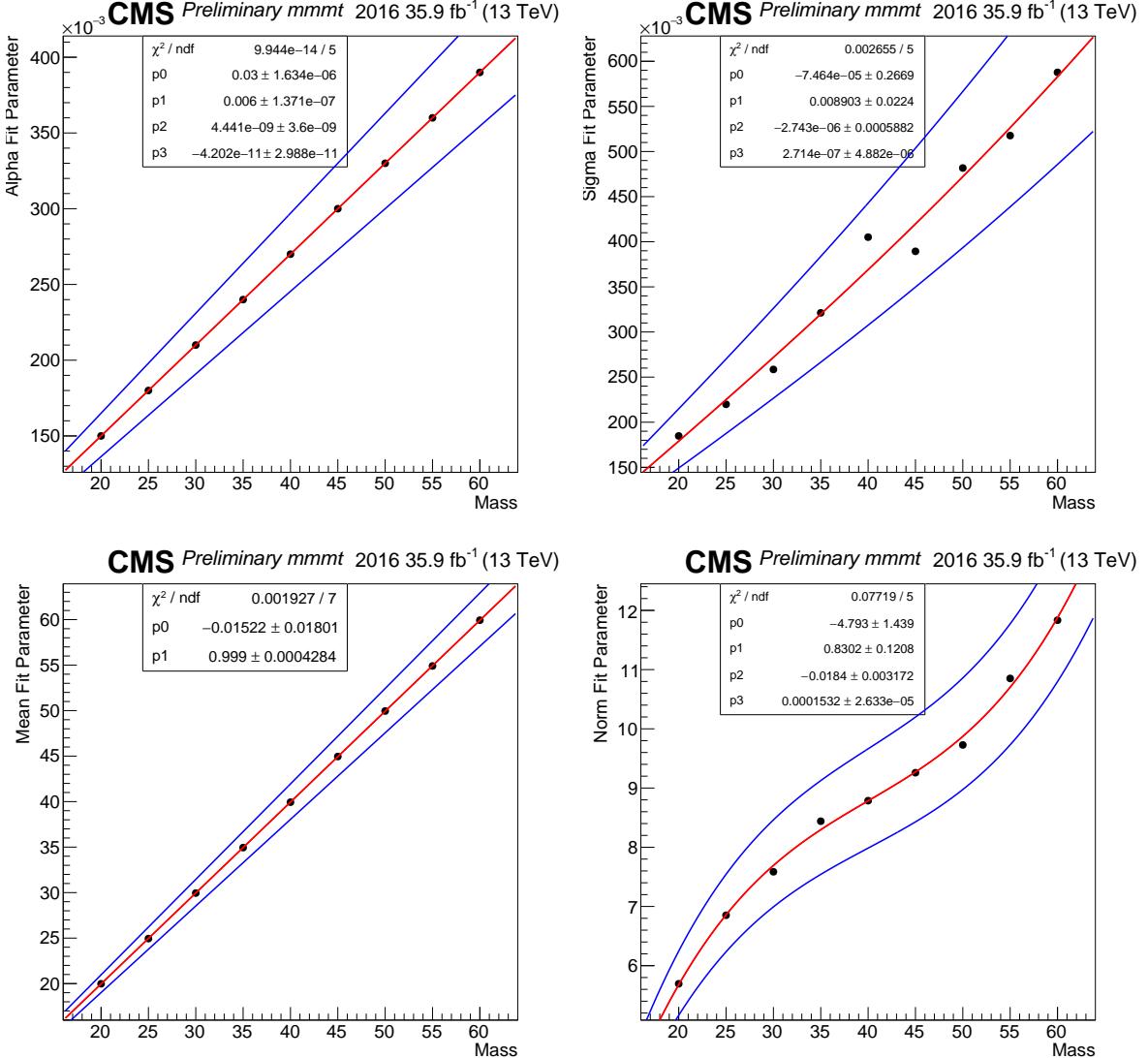


Figure 7.3: spline functions for 2016 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

look back at the section highlighting the uncertainties to see the bin-shift from the energy scale. The bin-shift indicates the amount the mean of the distribution is effected from the energy scale shift.

For the other systematic uncertainties that are not based on parametric shapes, like the energy scale of the leptons, a log-normal deviation to the normalization is considered. In order to read the impact plots, in order to understand the impacts place look at table 7.1:

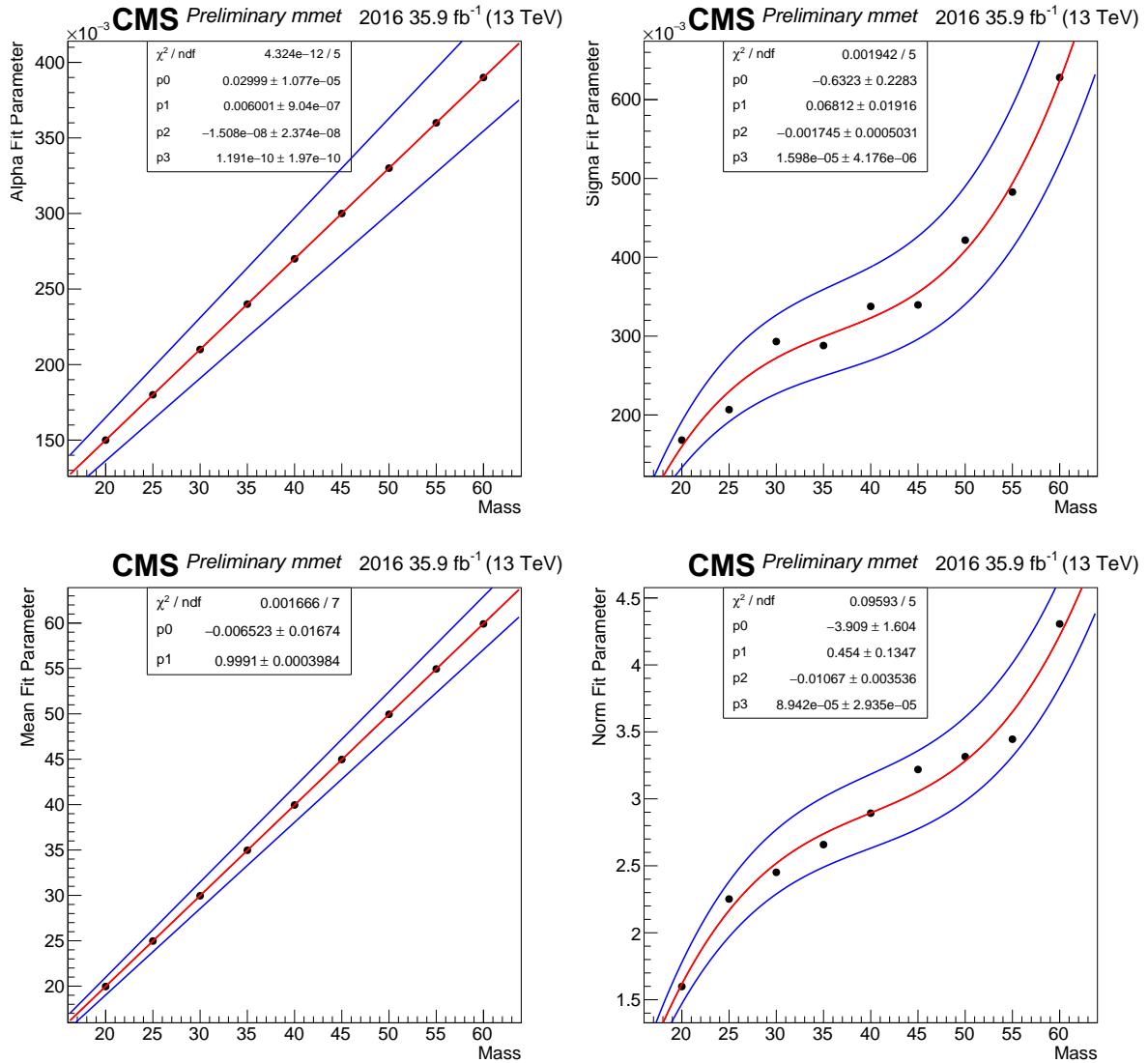


Figure 7.4: spline functions for 2016 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

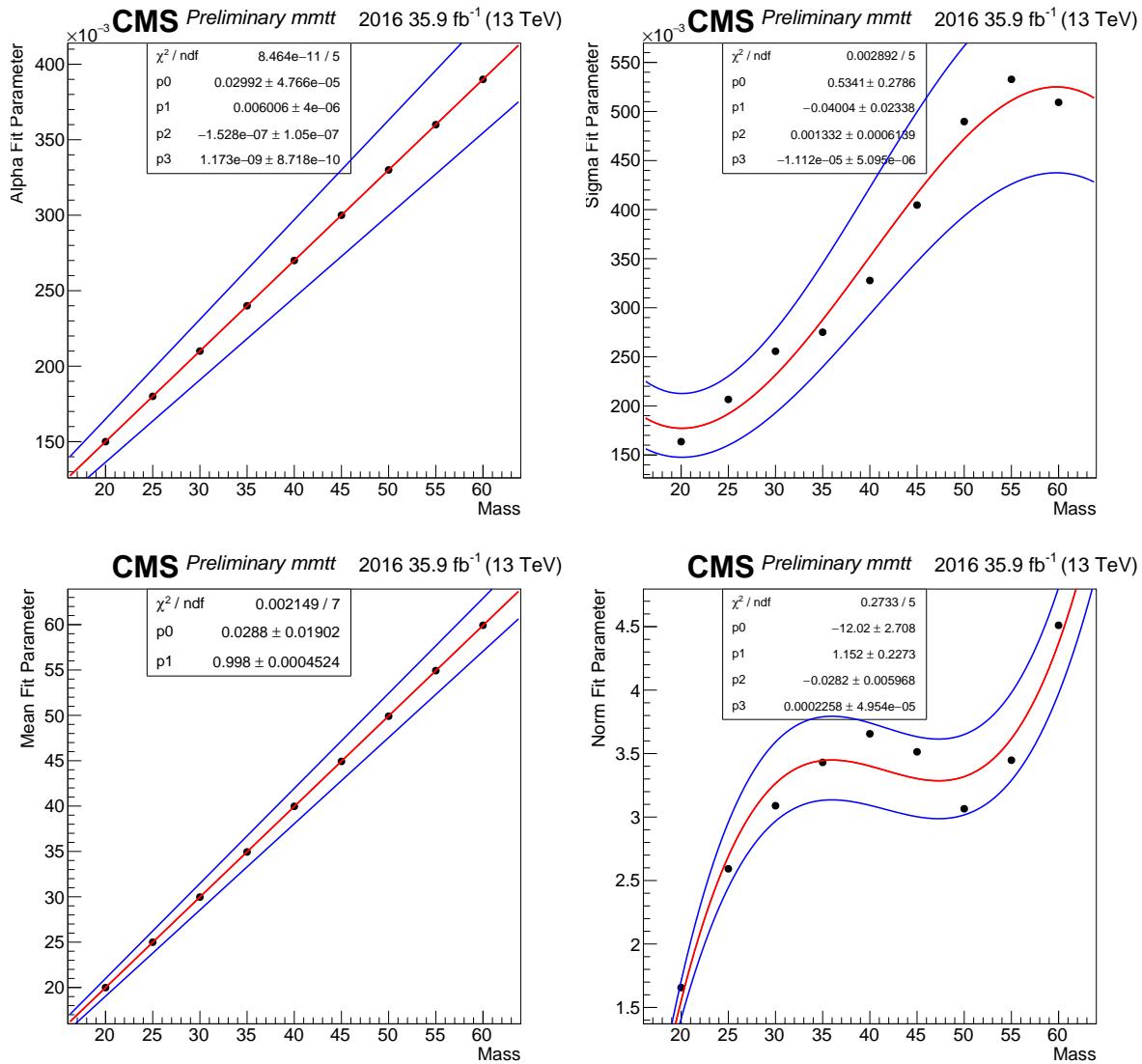


Figure 7.5: spline functions for 2016 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

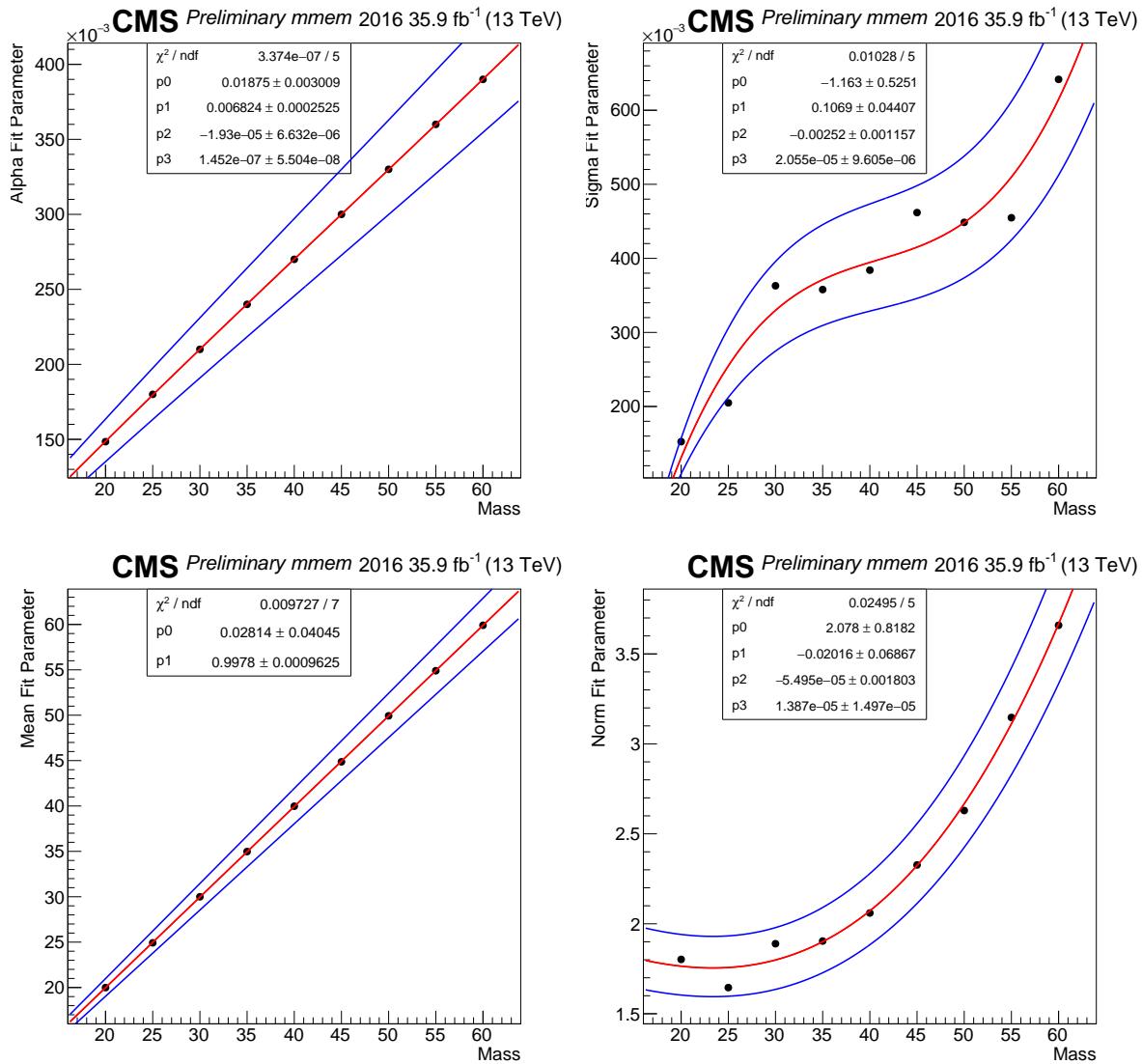


Figure 7.6: spline functions for 2016 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

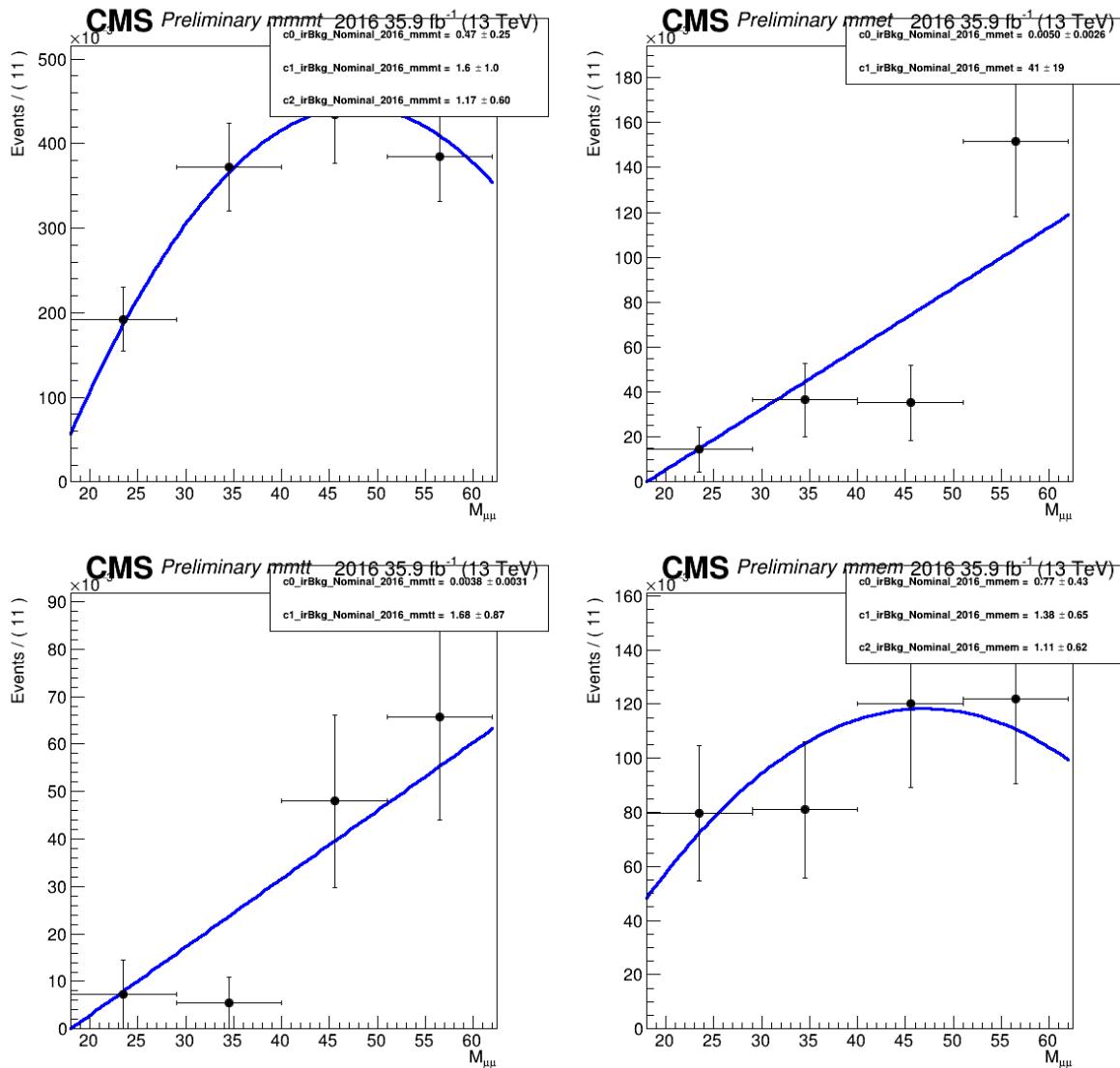


Figure 7.7: irreducible background fit using Bernstein polynomials

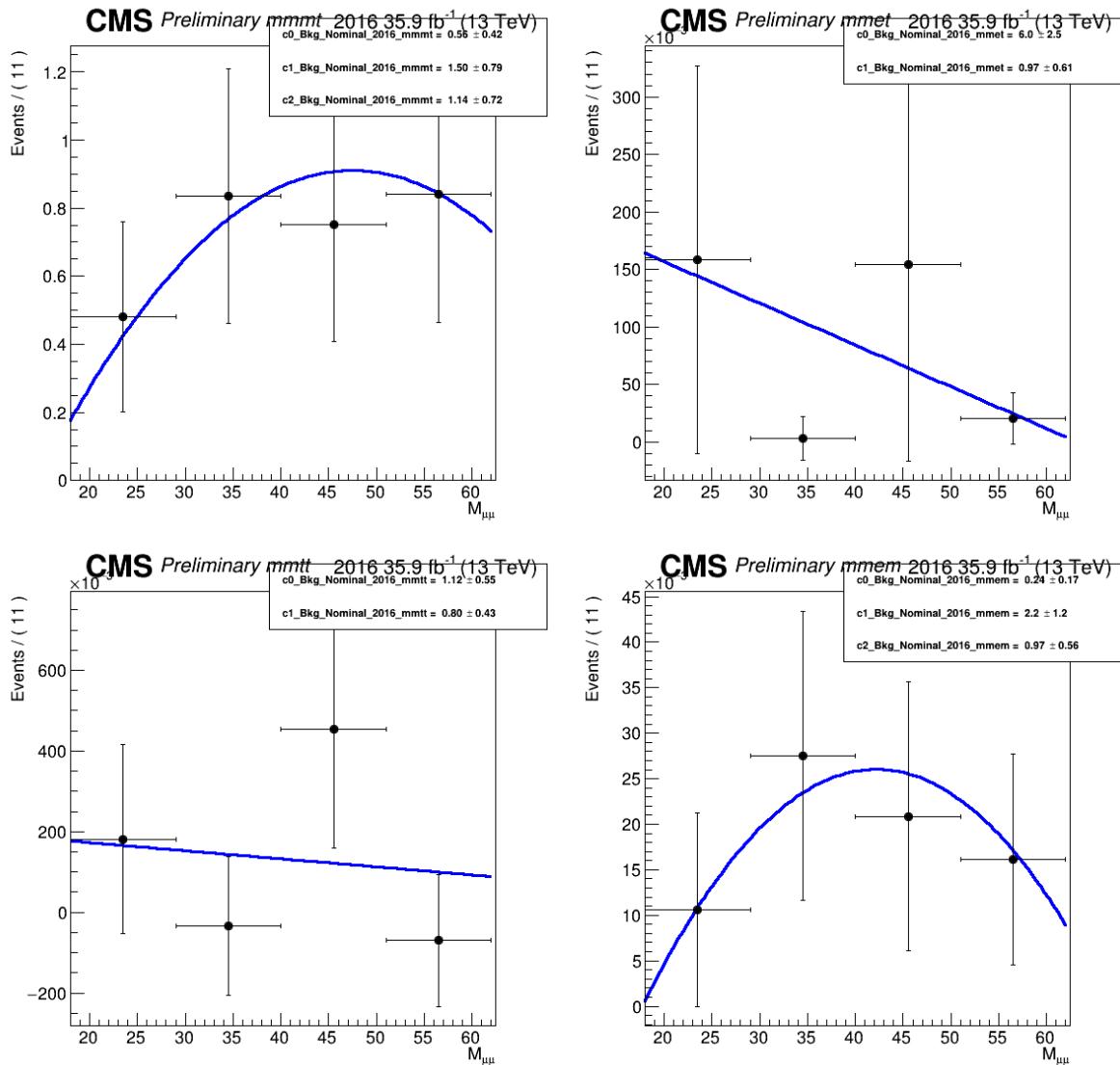


Figure 7.8: reducible background fit using Bernstein polynomials

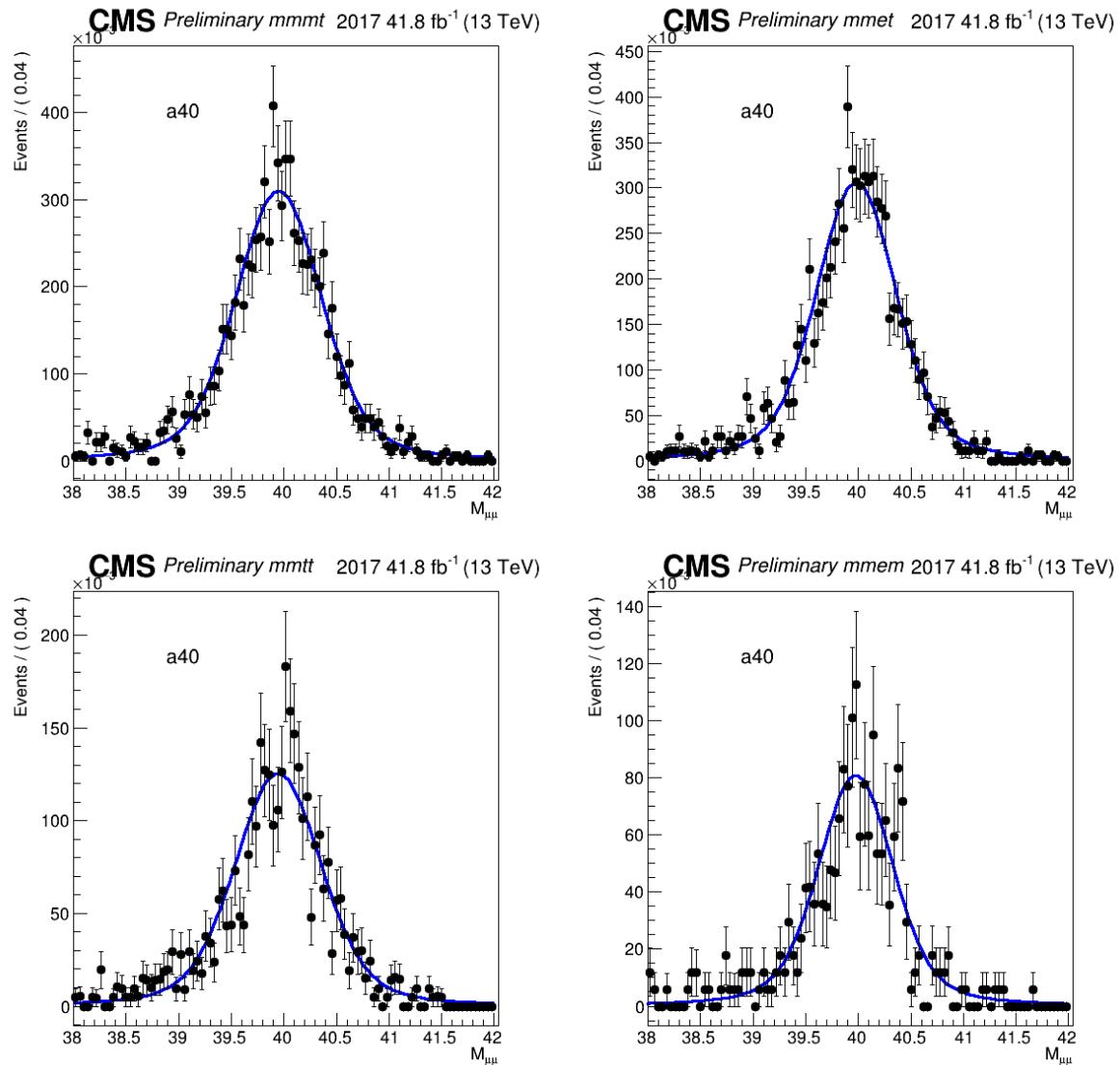


Figure 7.9: 2017 Signal fit using a Voigtian function

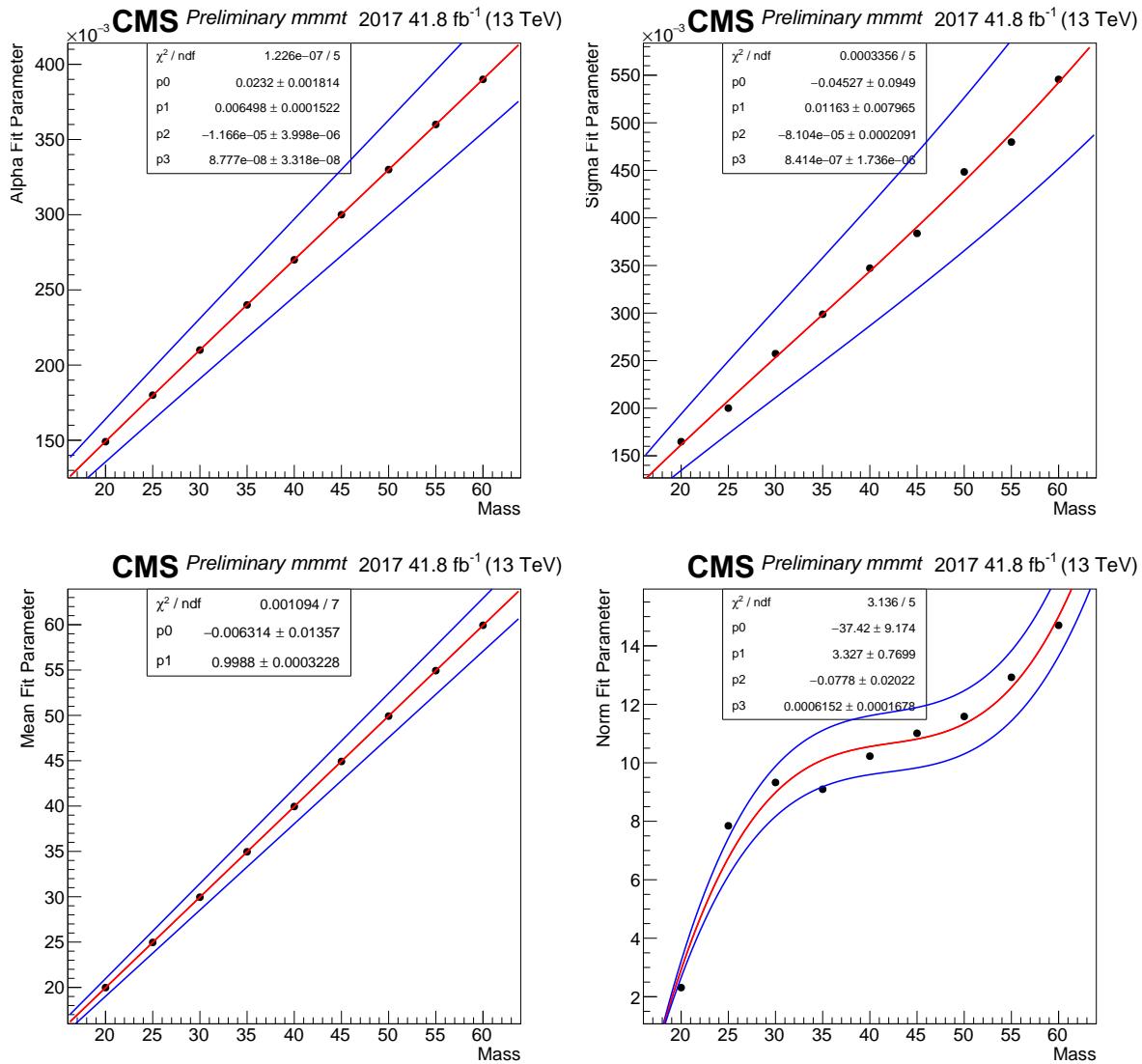


Figure 7.10: spline functions for 2017 mmm mt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

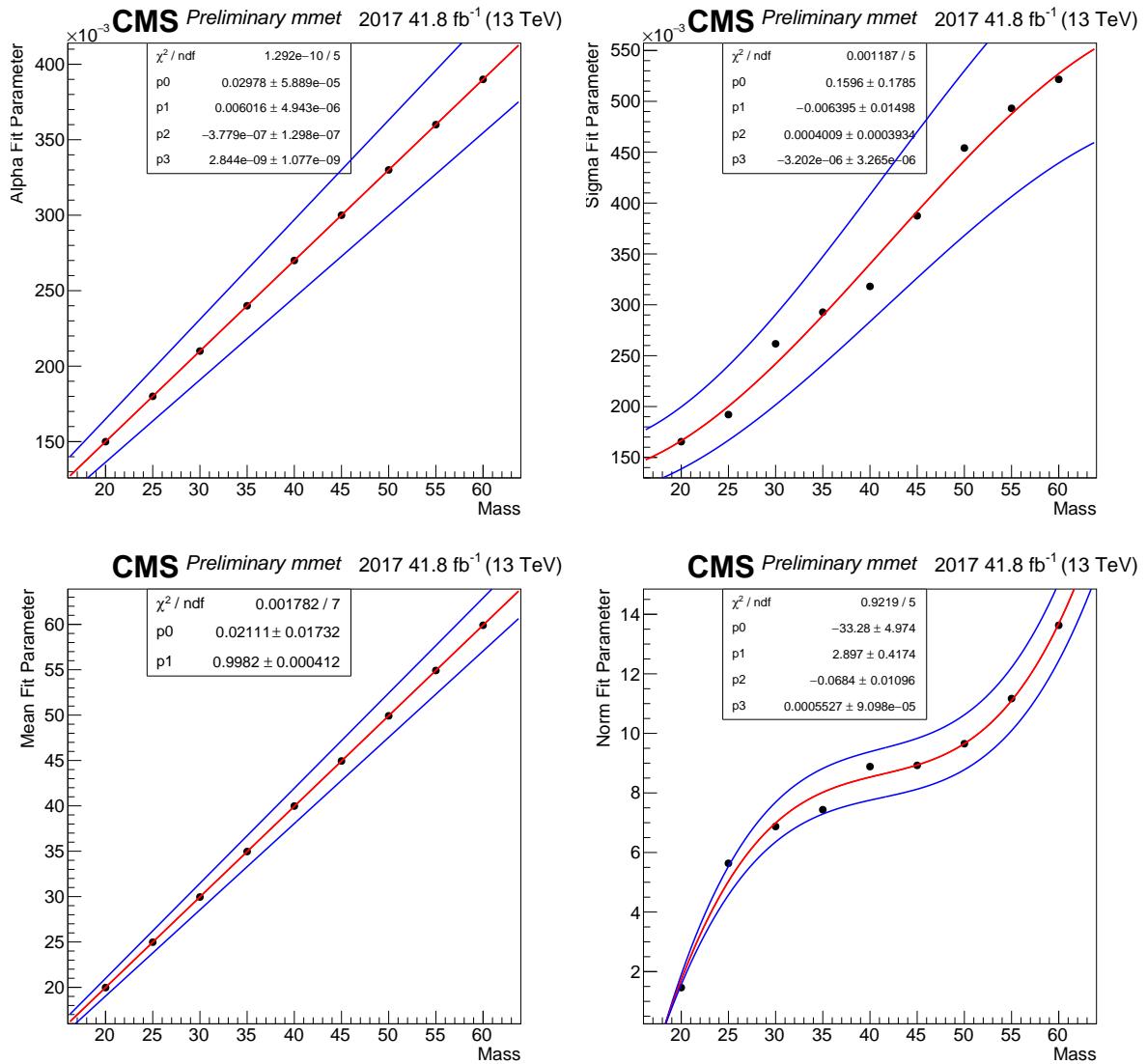


Figure 7.11: spline functions for 2017 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

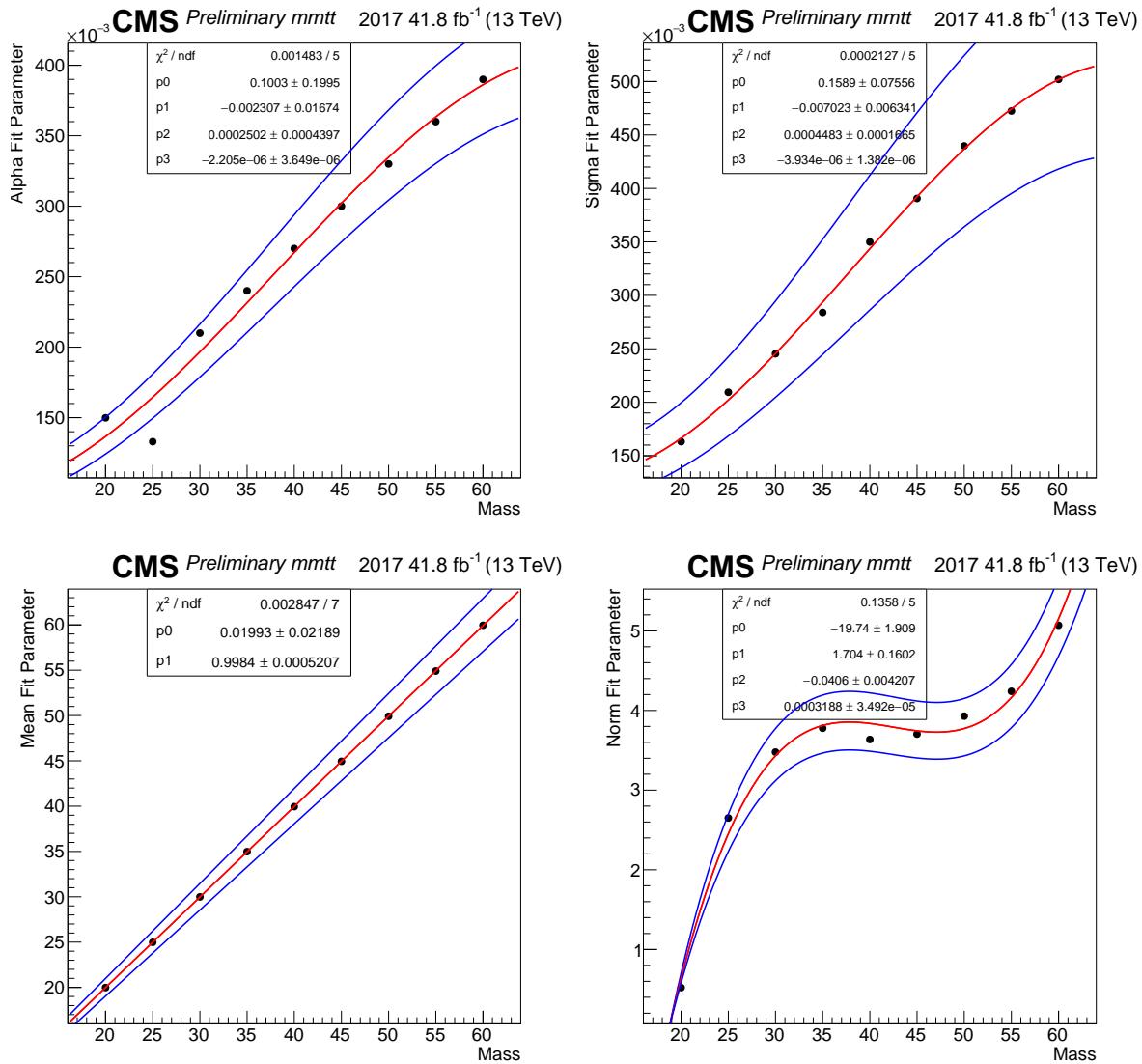


Figure 7.12: spline functions for 2017 mmtt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

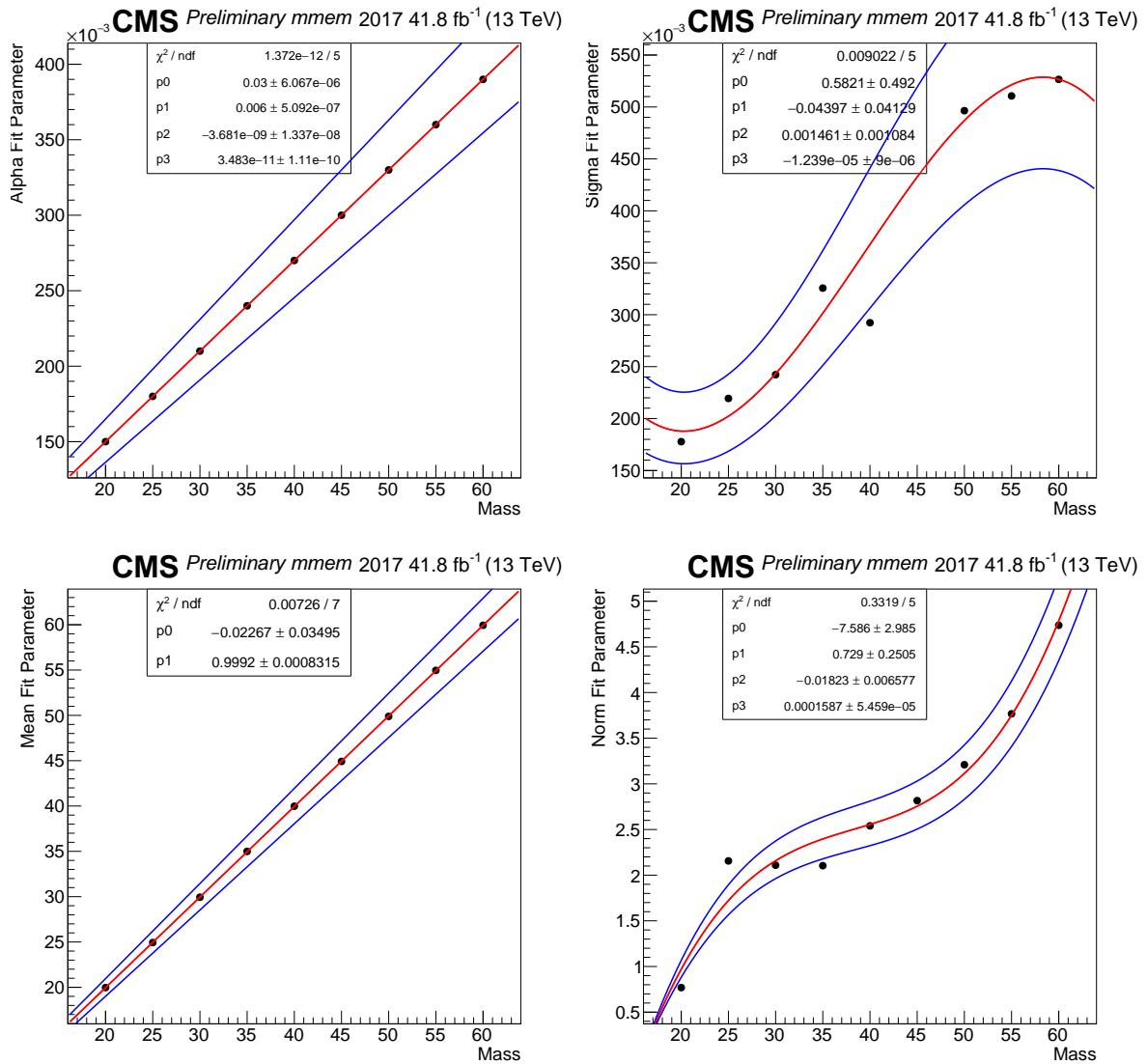


Figure 7.13: spline functions for 2017 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

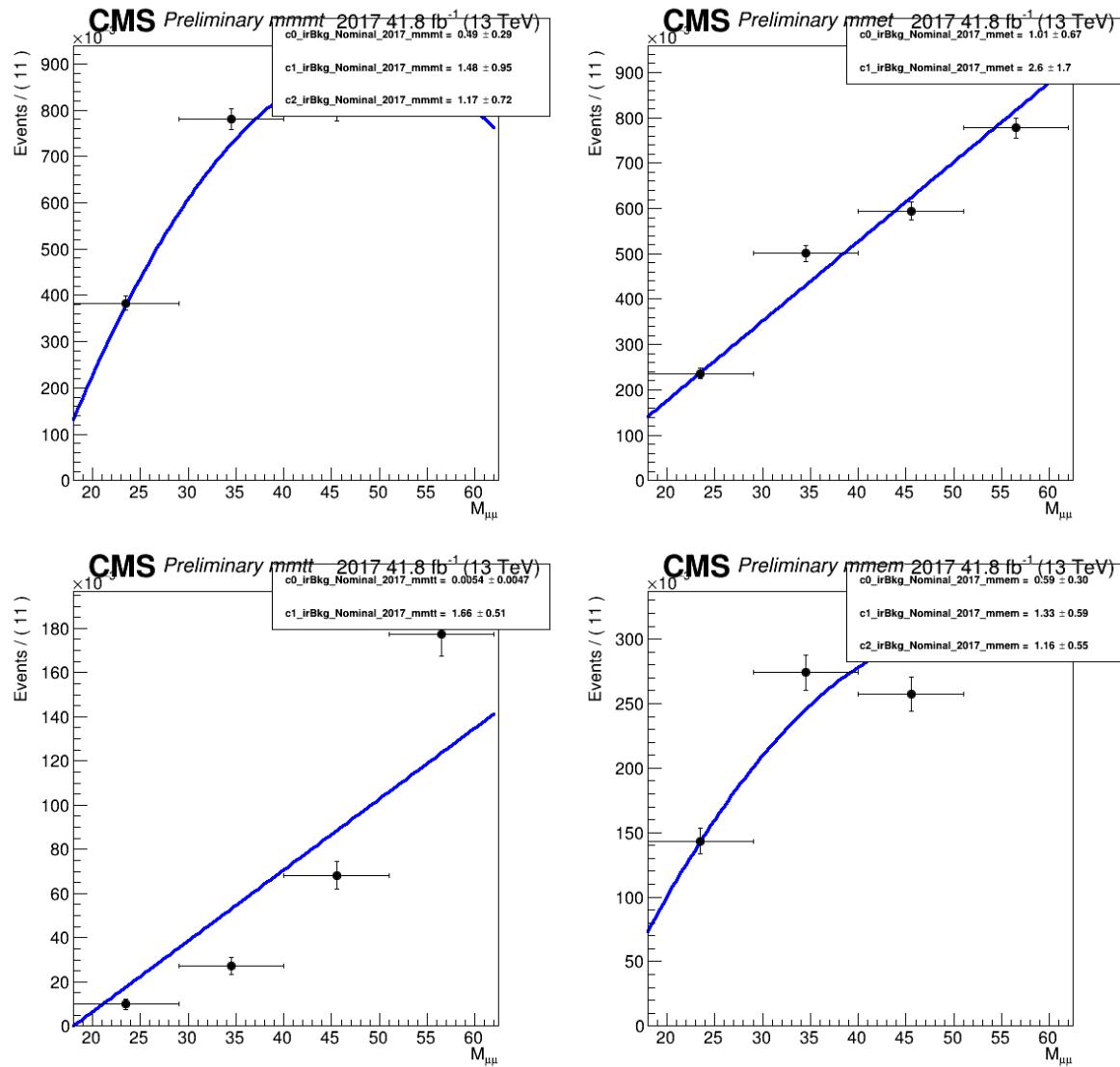


Figure 7.14: 2017 irreducible background fit using Bernstein polynomials

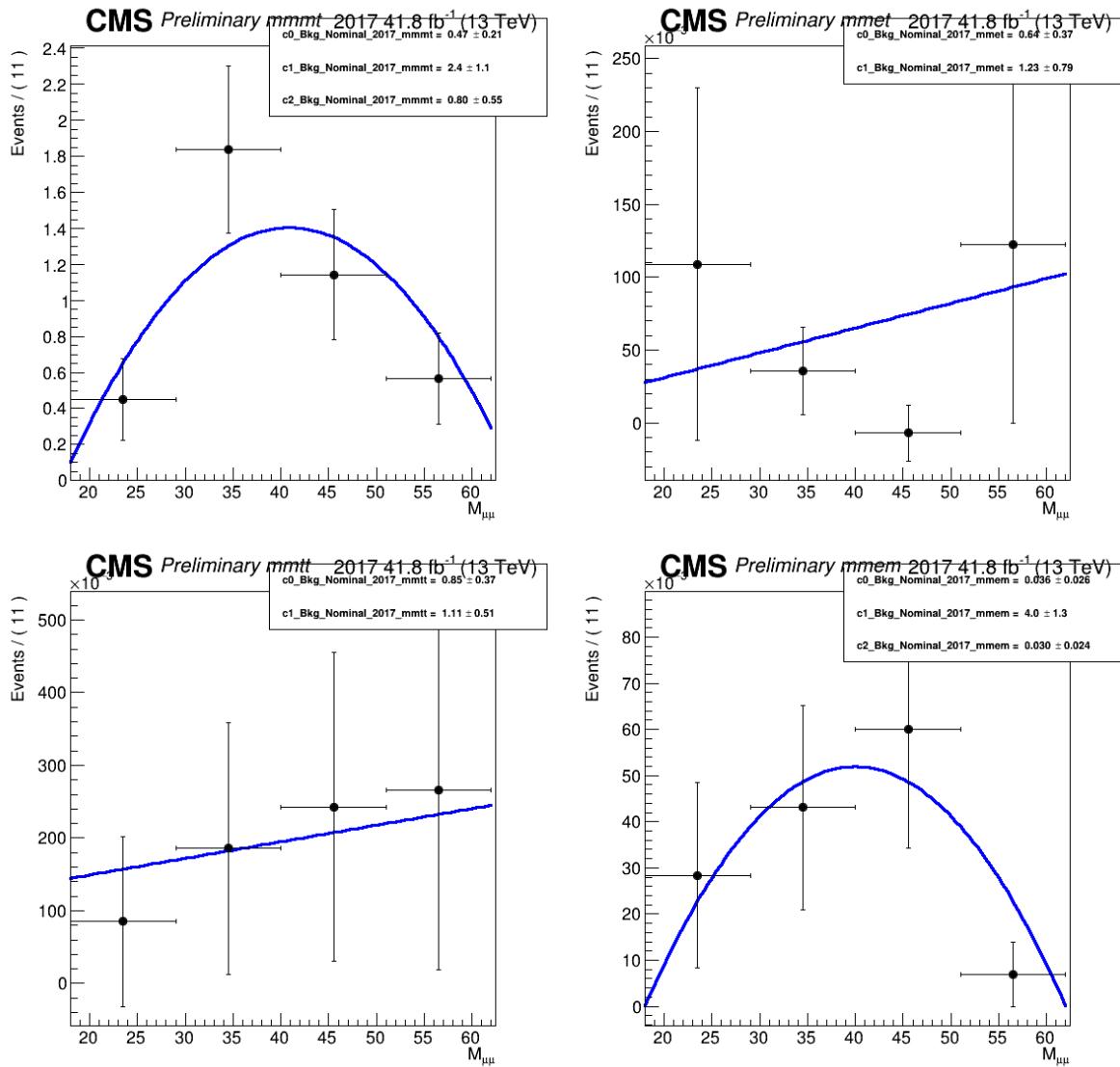


Figure 7.15: 2017 reducible background fit using Bernstein polynomials

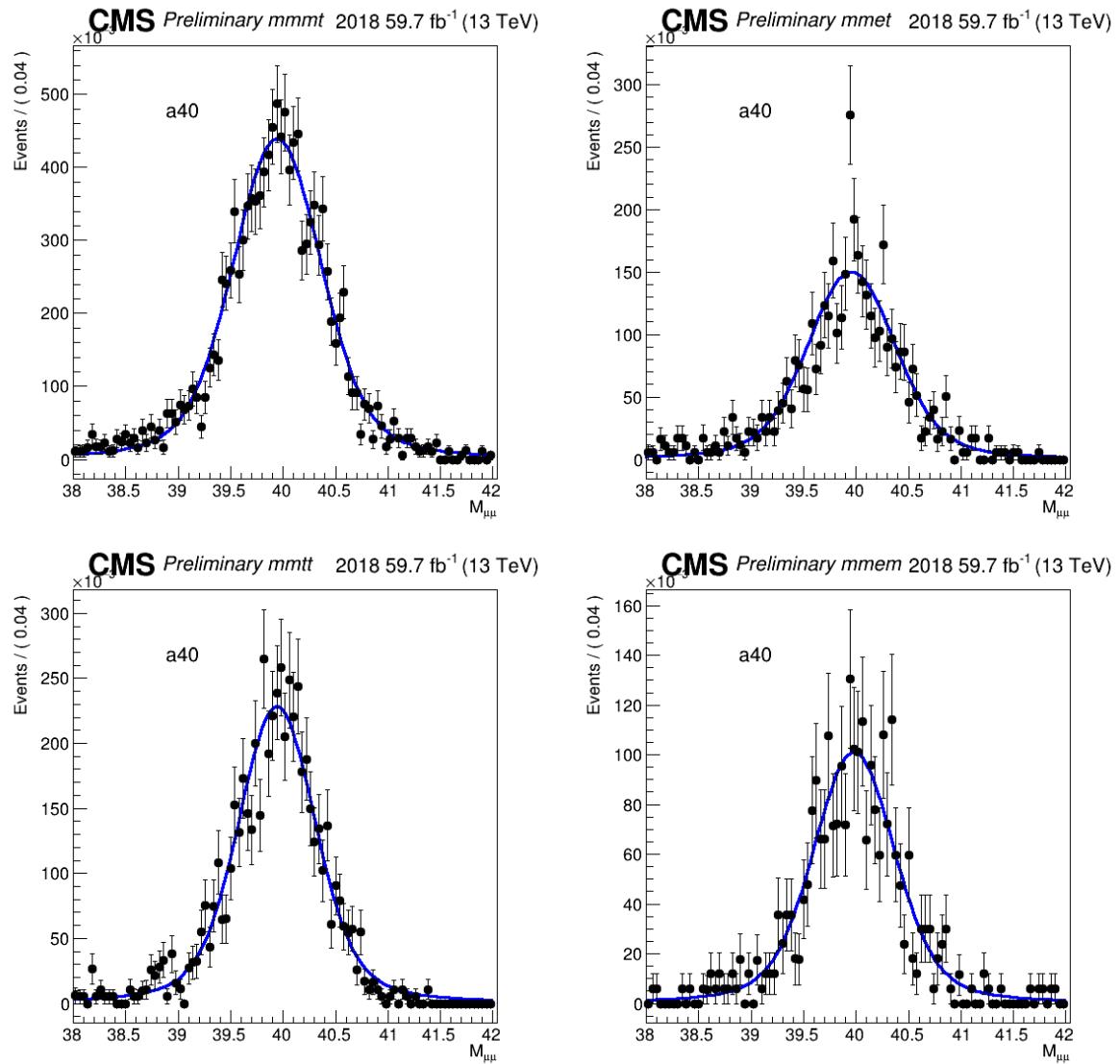


Figure 7.16: 2018 Signal fit using a Voigtian function

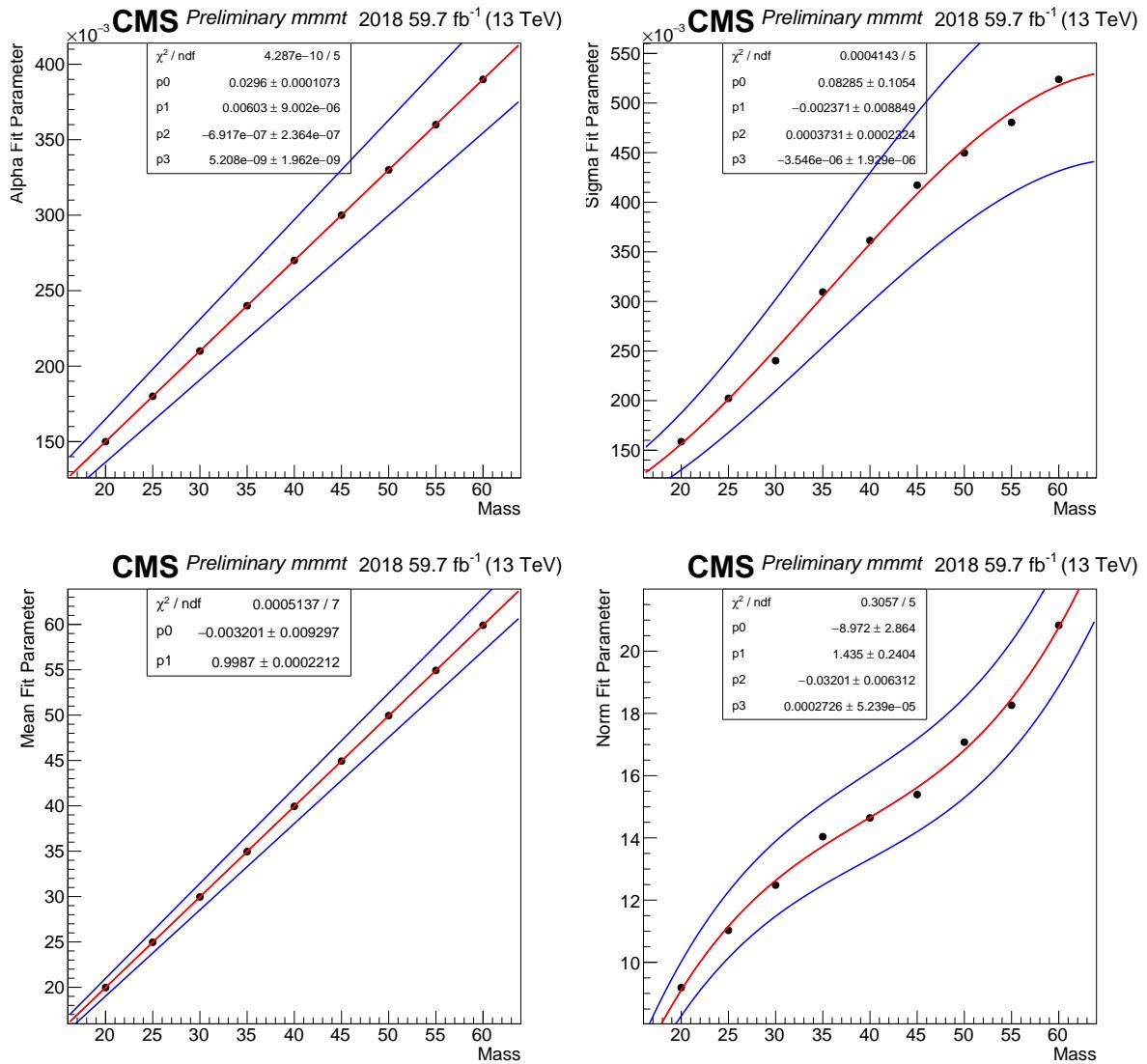


Figure 7.17: spline functions for 2018 mmmt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

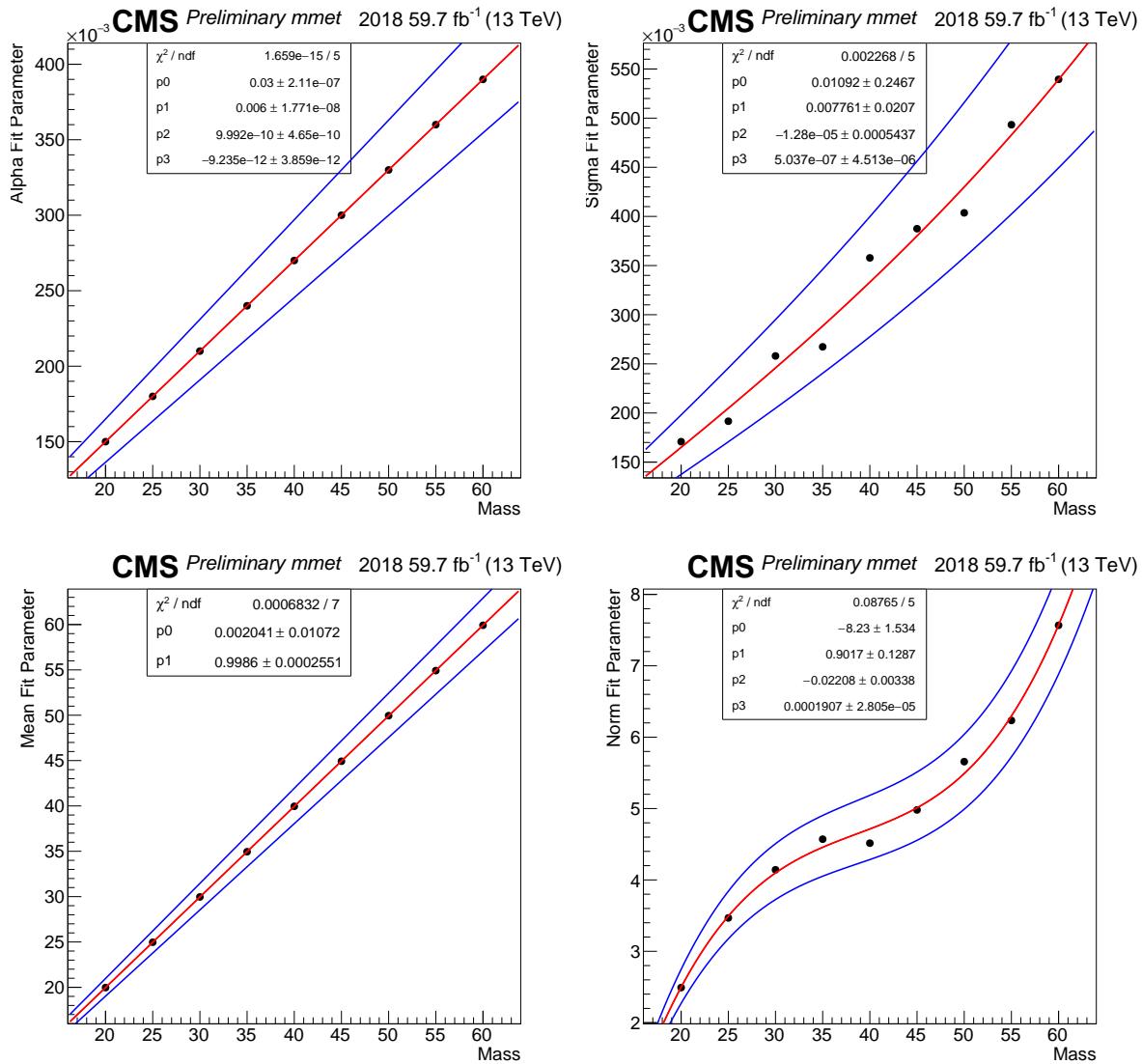


Figure 7.18: spline functions for 2018 mmet 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

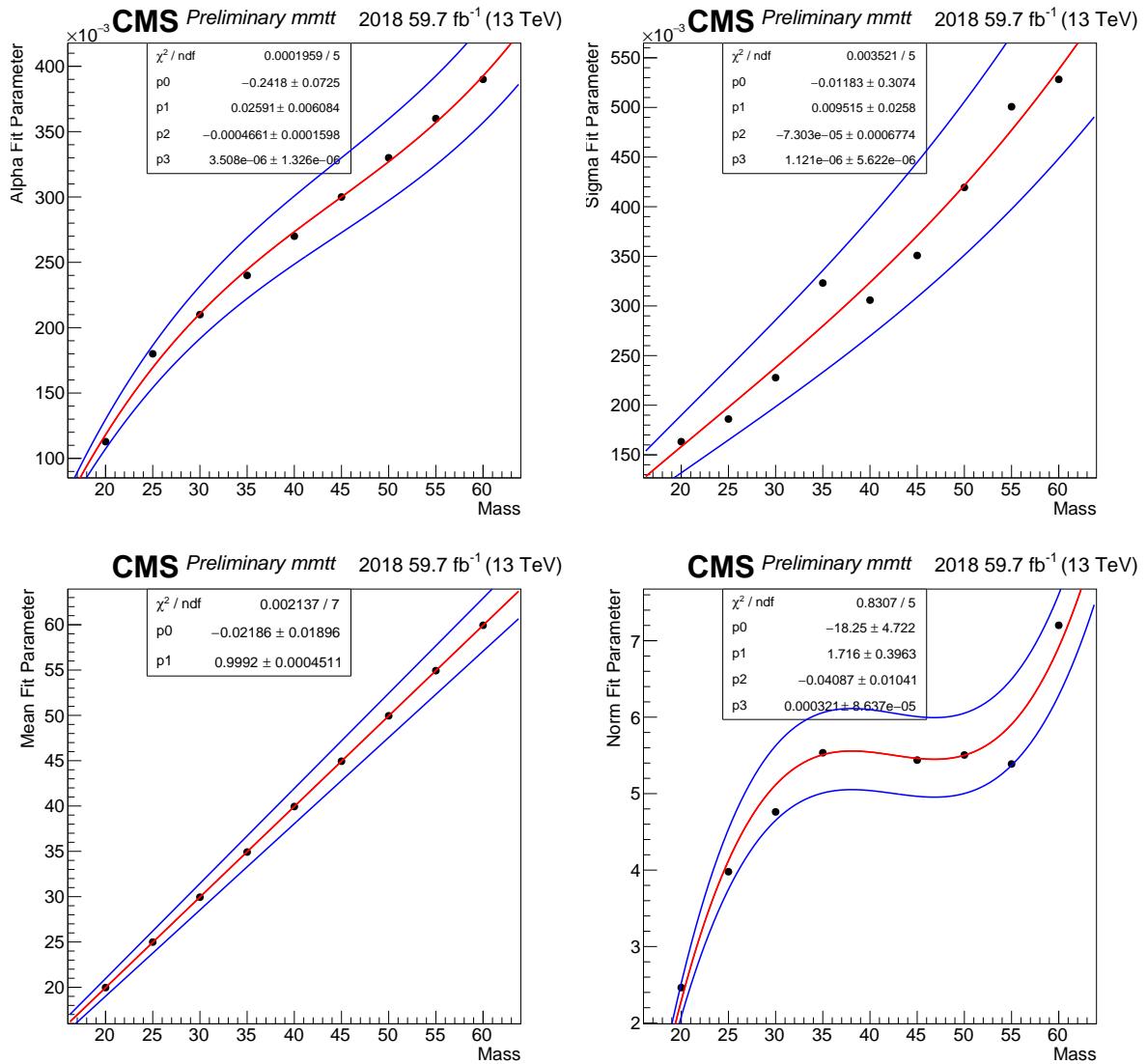


Figure 7.19: spline functions for 2018 mmtt 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

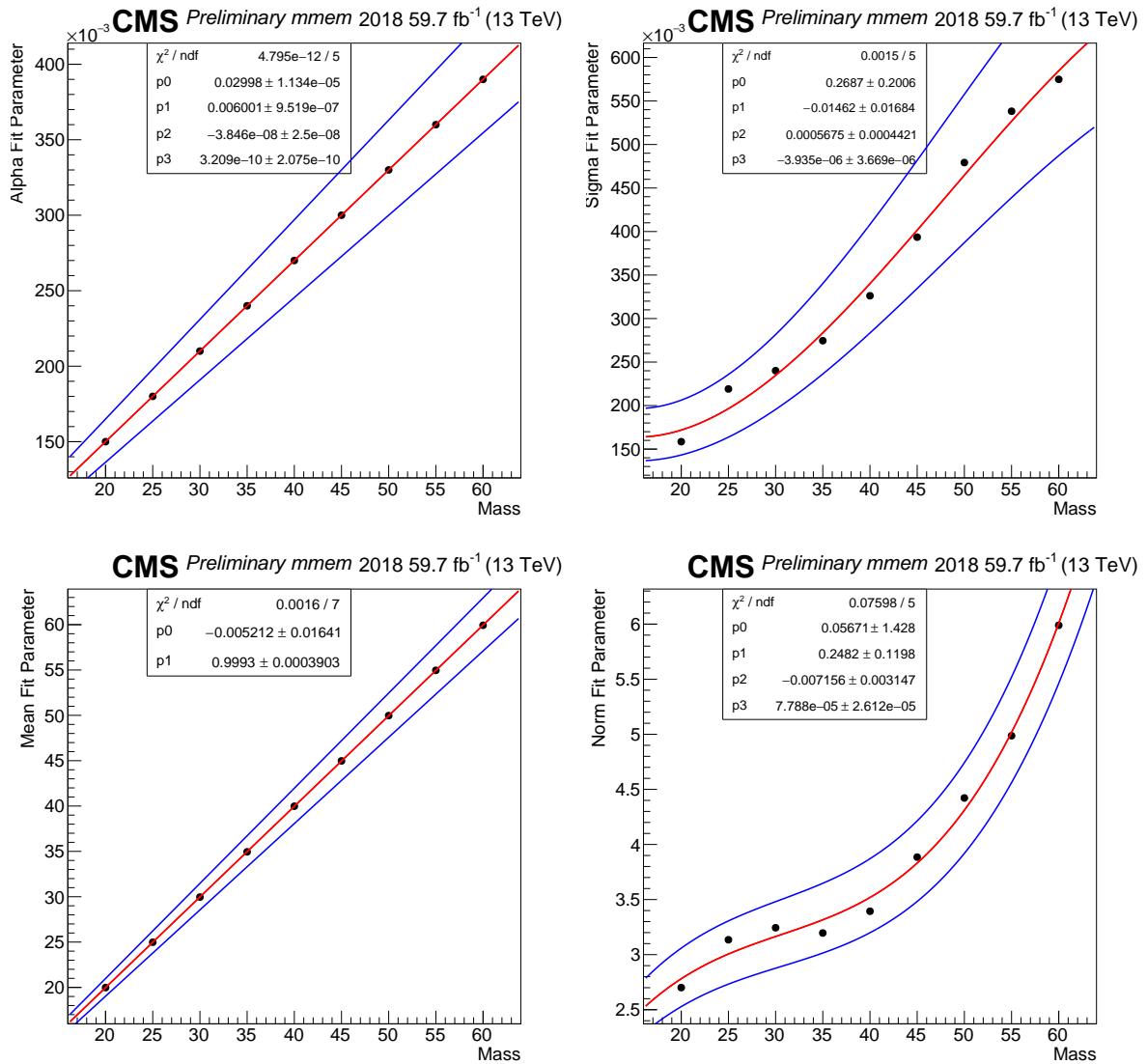


Figure 7.20: spline functions for 2018 mmem 3rd order used for Alpha, Sigma, and Normalization, linear fit for the Mean

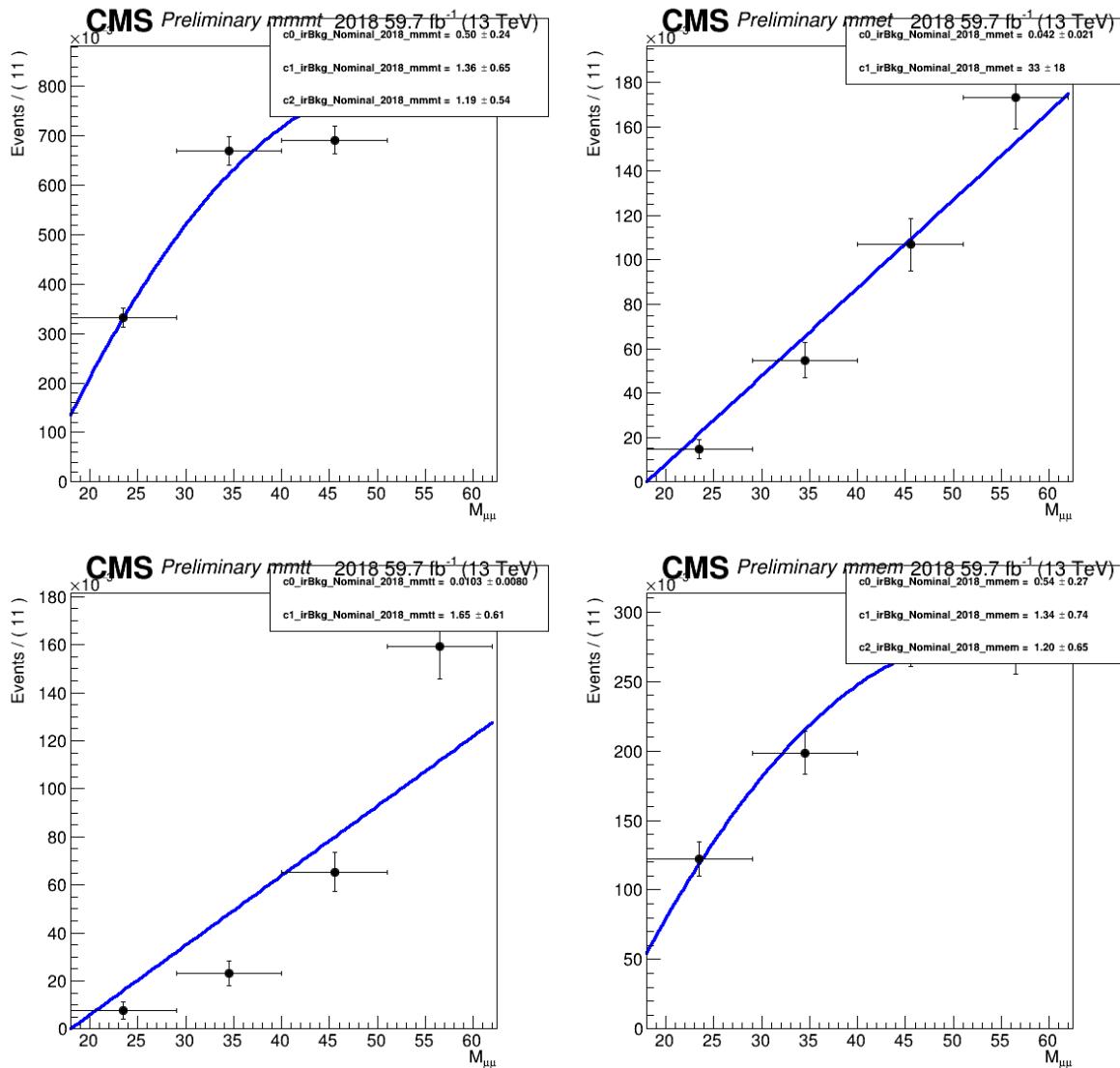


Figure 7.21: 2018 irreducible background fit using Bernstein polynomials

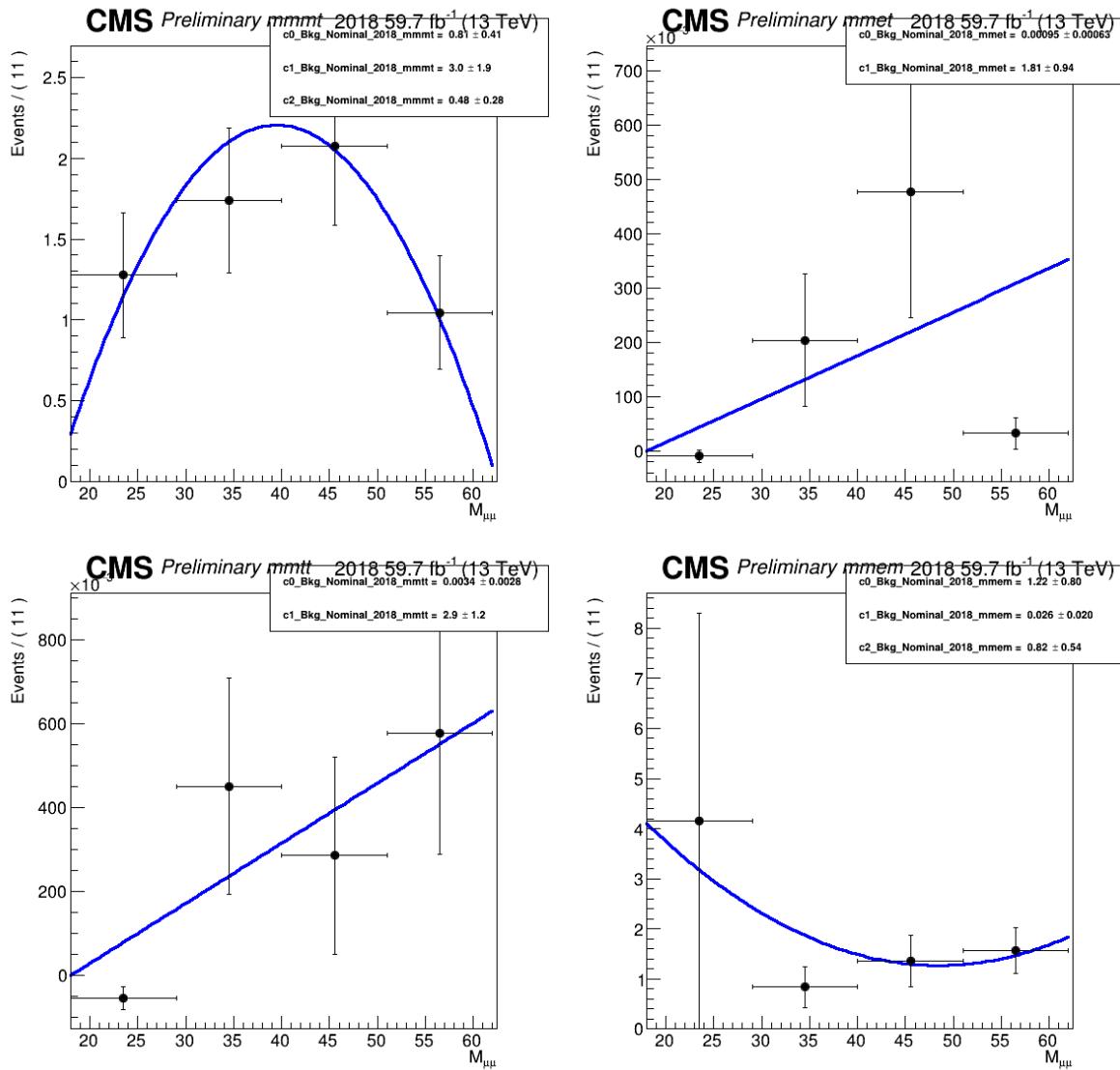


Figure 7.22: 2018 reducible background fit using Bernstein polynomials

Table 7.1:

Type of Uncertainty	Brief Description
scale	log-normal uncertainties that effect overall scale of the distribution these include $e$ , $\mu$ , and $\tau$ (split by decay mode) energy scales
c0 <sub>-</sub> , c1 <sub>-</sub> , ... cN <sub>-</sub>	Coefficients of the Bkg (datadriven) or irBkg (ZZ) parametric shape
lumi	luminosity uncertainty (1.016 representing 1.6% uncertainty)
intAlpha	alpha interpolated spline function shape uncertainty (10%)
intSigma	sigma interpolated spline function shape uncertainty (20%)
intMean	mean interpolated spline function shape uncertainty (0.1%)

Systematic pull distributions for the magnitude of the expected uncertainty from each source is listed in figure 7.23. The rest of the channels and years are located in the appendix C.

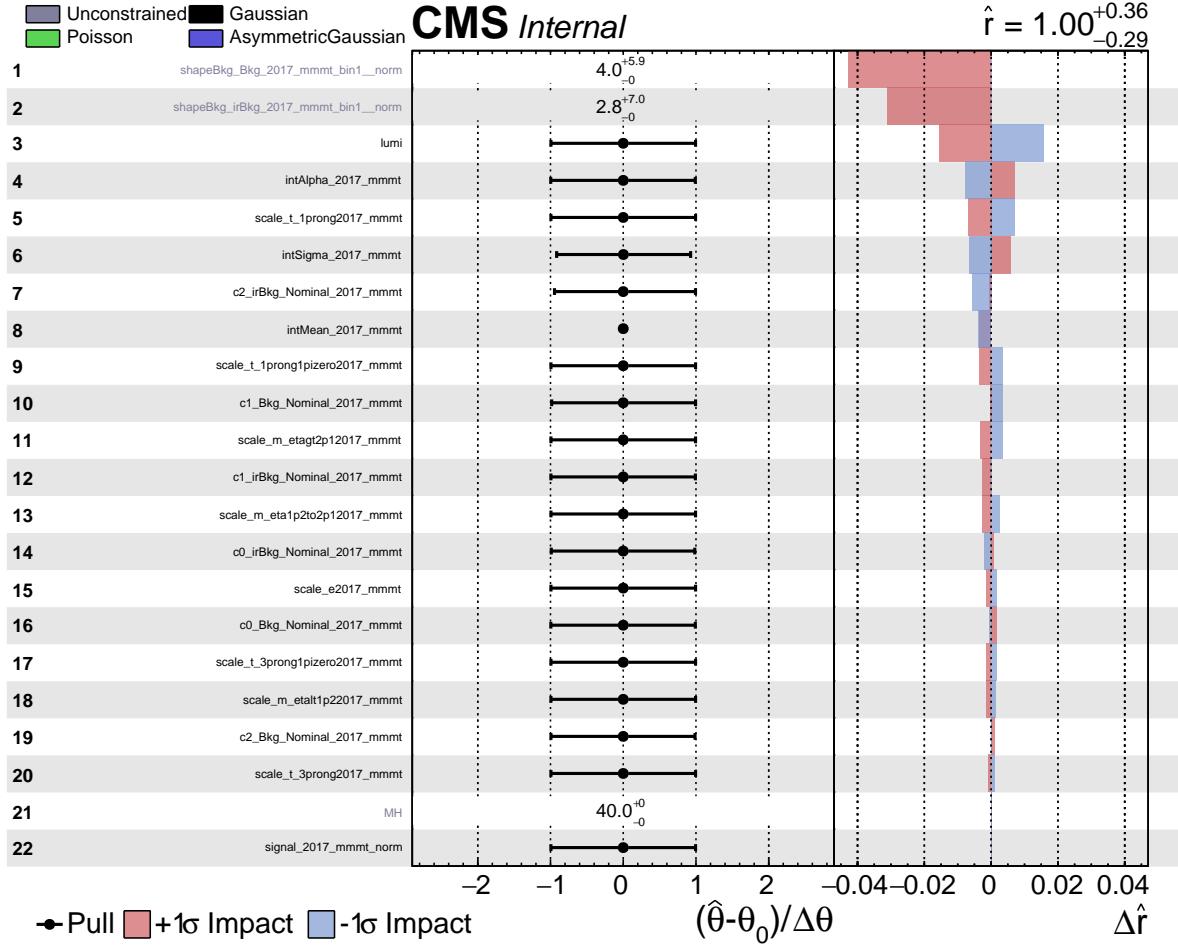


Figure 7.23: Expected systematic impacts for the fit model  $\mu\mu\tau$

# Chapter 8

## Results

After all the event selection including the signal extraction cuts on the four lepton invariant mass and the dimuon mass greater than the di-tau mass, the following yeilds represent the total events expected in each category accross all channels and years

Table 8.1: Expected Event Yields of Signal an Background Across All Years

Signal a Mass 40	Background (Data driven FF)	Irreducible (ZZ)
6.54	23.61	6.33

Each unbinned parametric likelihood fit is included for each mass point into a Confidence Level scan to set an upper limit at 95% on the branching ratio. Initial values of the signal distributions are selected to make sure that the signal strength modifier in the limit is order unity. The range of masses reflected are show between 20 GeV and 60 GeV to ensure compatibility with Higgs aa combination limits for more exotic Higgs models - like those at lower  $a$  mass. In order to estimate the upper limit at 95% CL on the branching fraction a simple poisson model can be used. For a low statistically dominated search we can choose to estimate the background yeild as no events.

The upper limit on branching fraction calculated earlier:

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}} = 0.00027$$

An upper limit of 95% is set on the Branching Fraction by adjusting the signal strength (event yield) until a p-value of 5% is reached on the joint likelihood function for the fit model. The event yield is normalized with a set branching fraction which was assumed to be the Higgs  $\sigma_{SM} \times 0.01\%$ . Multiplying the CL by 0.01% thus yields the limit on  $\frac{\sigma_h}{\sigma_{SM}} B(h \rightarrow aa \rightarrow 2\mu 2\tau)$ . Preliminary Limits

are set using the Asymptotic Limit method [15] for each mass point.

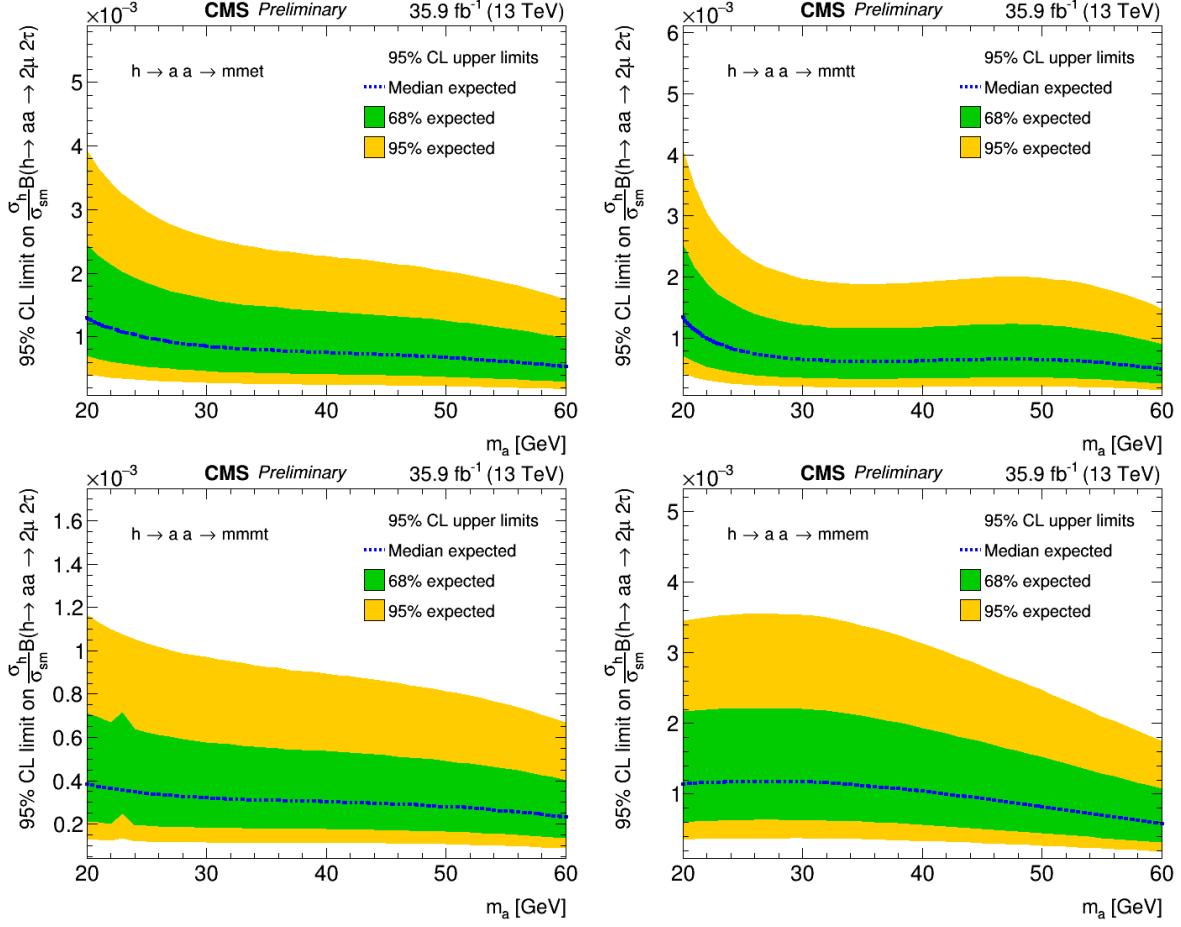


Figure 8.1: Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sections

All of the years and channels are then combined to form the combined result and the model 2HDM+S interpretations for different scenarios. Scenario type III where coupling to  $\tau$  leptons are favored is expected to be the most stringent scenario for this final state.

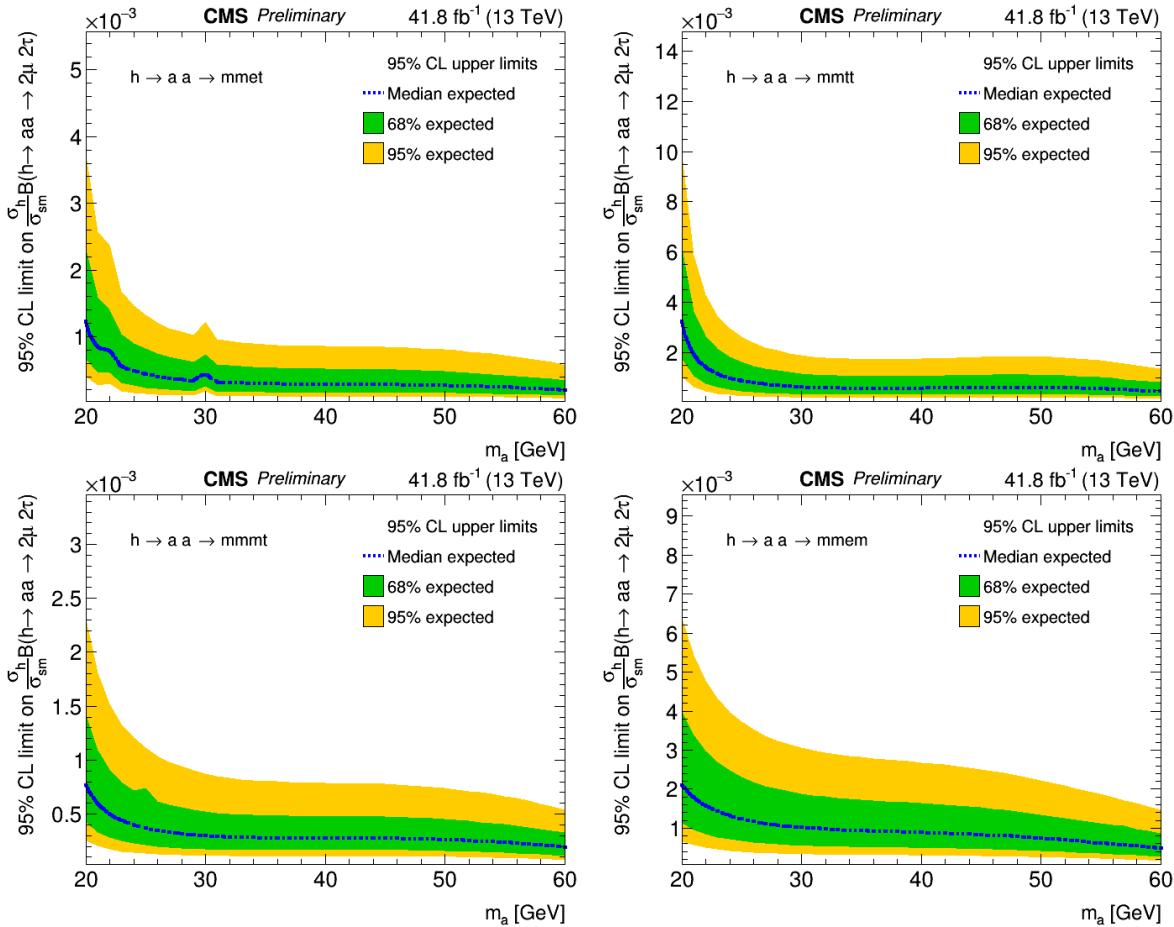


Figure 8.2: Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sections for 2017

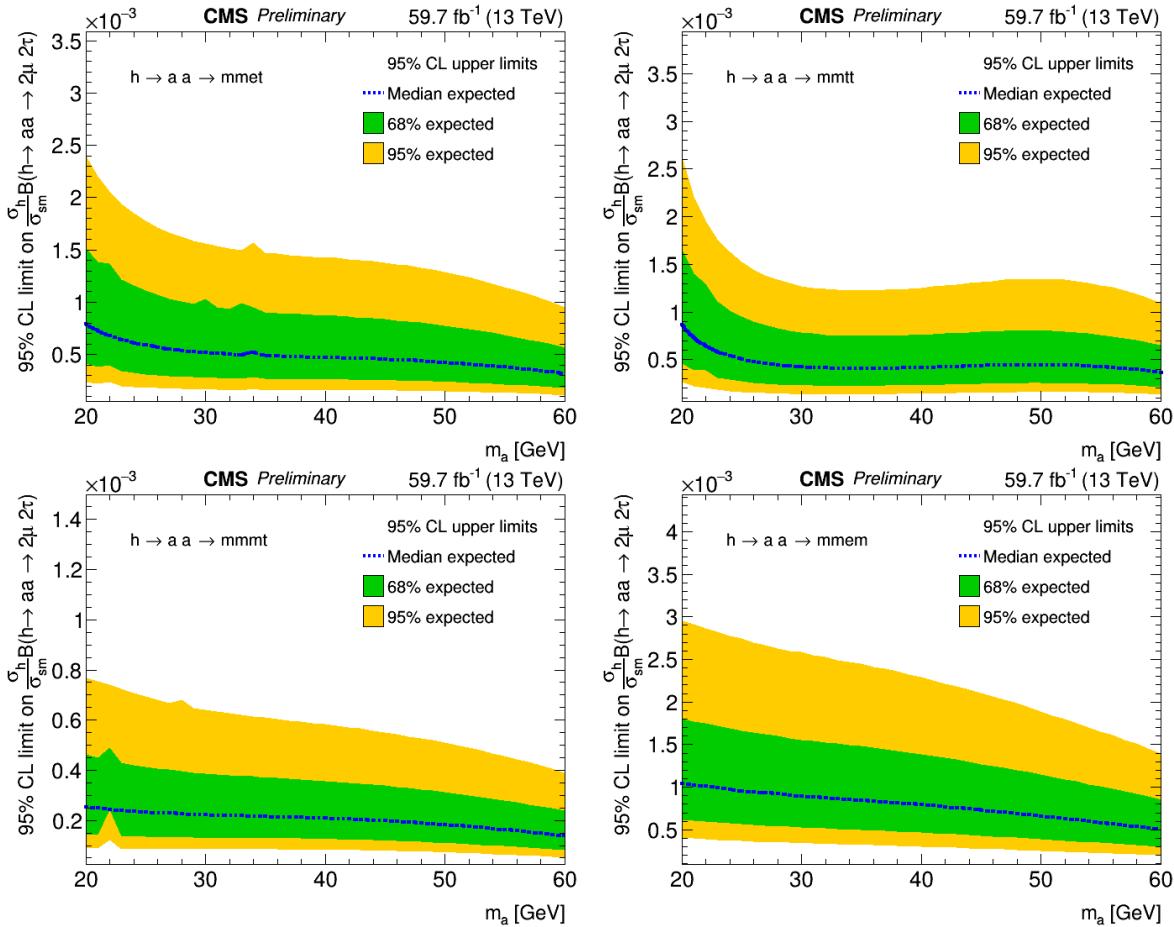


Figure 8.3: Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sections for 2018

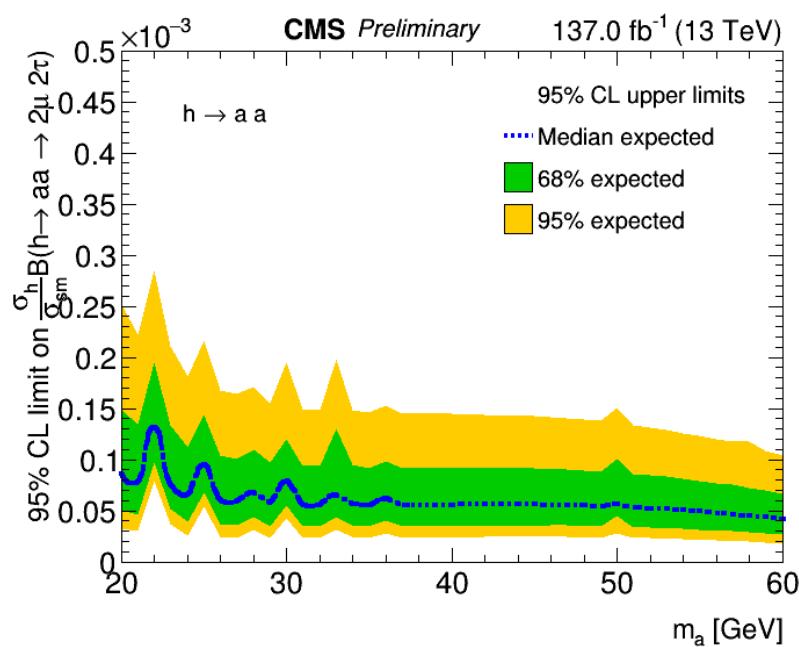


Figure 8.4: Asymptotic C.L. Limits on the Branching Fraction times ratio of the SM cross sections for the full Run II dataset  $137^{-1}\text{fb}$

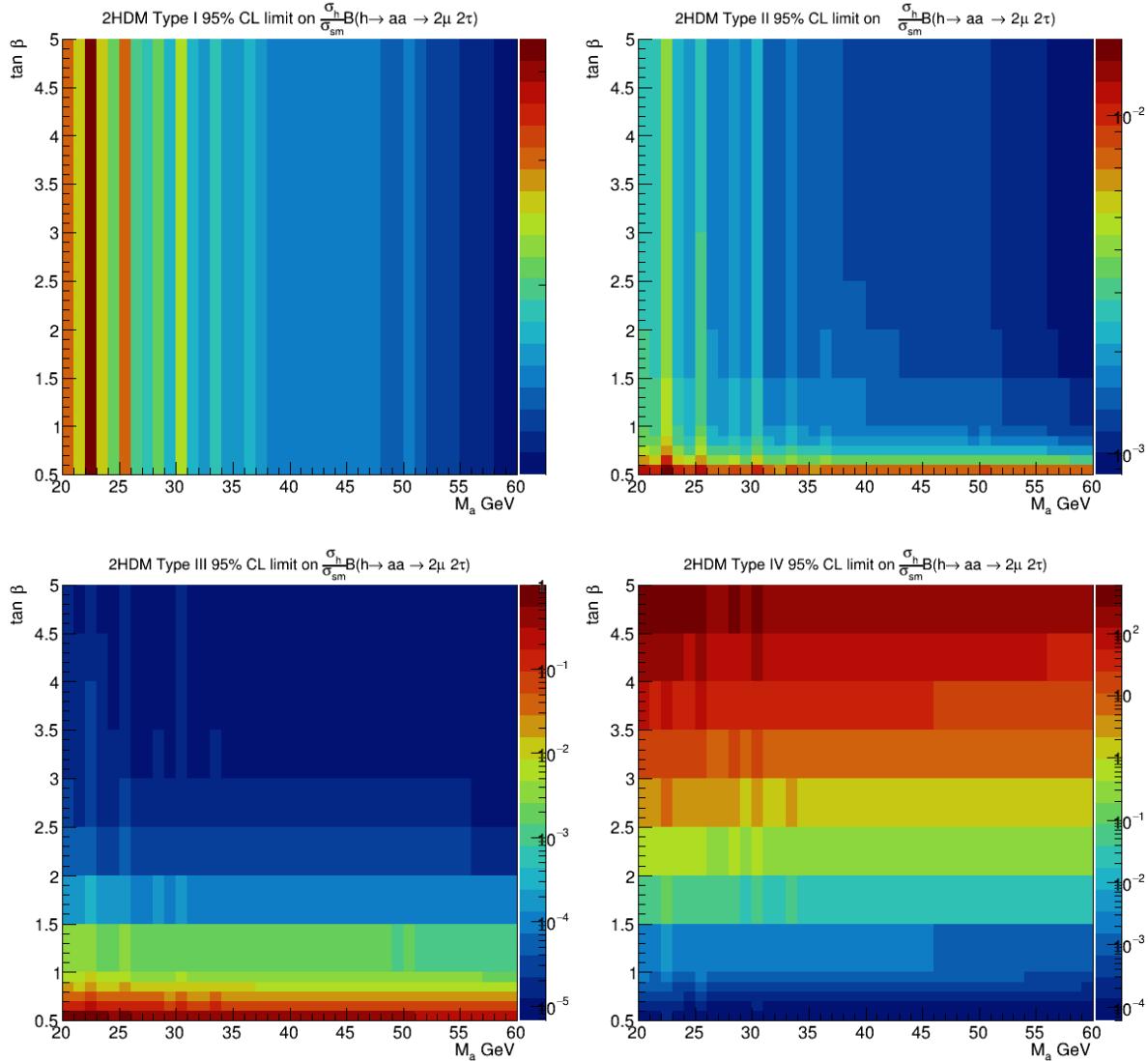


Figure 8.5: 2HDM model specific scenarios upper 95% C.L. limits on the branch fraction of Higgs to  $a a$  times the ratio of the SM cross sections for the full Run II dataset  $137^{-1} \text{fb}$

## **Chapter 9**

## **Conclusions**

## Appendix A

# Data and Simulation Samples

### A.1 Data and Simulation Used for Analysis

The full Run II dataset was used corresponding to  $137\text{fb}^{-1}$ . Data for 2016

Table A.1: List of data sets included in the analysis for the 2016 data taking period.

Data set
/SingleMuon/Run2016B_ver2-Nano250ct2019_ver2-v1/NANOAOD
/SingleMuon/Run2016B_ver1-Nano250ct2019_ver1-v1/NANOAOD
/SingleMuon/Run2016G-Nano250ct2019-v1/NANOAOD
/SingleMuon/Run2016F-Nano250ct2019-v1/NANOAOD
/SingleMuon/Run2016E-Nano250ct2019-v1/NANOAOD
/SingleMuon/Run2016D-Nano250ct2019-v1/NANOAOD
/SingleMuon/Run2016C-Nano250ct2019-v1/NANOAOD
/SingleMuon/Run2016H-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016H-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016G-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016F-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016B_ver2-Nano250ct2019_ver2-v1/NANOAOD
/DoubleMuon/Run2016B_ver1-Nano250ct2019_ver1-v1/NANOAOD
/DoubleMuon/Run2016E-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016D-Nano250ct2019-v1/NANOAOD
/DoubleMuon/Run2016C-Nano250ct2019-v1/NANOAOD

Data for 2017

Data for 2018

Table A.2: List of data sets included in the analysis for the 2017 data taking period.

Data set
/DoubleMuon/Run2017B-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2017C-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2017D-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2017E-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2017F-02Apr2020-v1/NANO AOD
/MuonEG/Run2017B-02Apr2020-v1/NANO AOD
/MuonEG/Run2017C-02Apr2020-v1/NANO AOD
/MuonEG/Run2017D-02Apr2020-v1/NANO AOD
/MuonEG/Run2017E-02Apr2020-v1/NANO AOD
/MuonEG/Run2017F-02Apr2020-v1/NANO AOD
/SingleMuon/Run2017B-02Apr2020-v1/NANO AOD
/SingleMuon/Run2017C-02Apr2020-v1/NANO AOD
/SingleMuon/Run2017D-02Apr2020-v1/NANO AOD
/SingleMuon/Run2017E-02Apr2020-v1/NANO AOD
/SingleMuon/Run2017F-02Apr2020-v1/NANO AOD
/DoubleEG/Run2017B-02Apr2020-v1/NANO AOD
/DoubleEG/Run2017C-02Apr2020-v1/NANO AOD
/DoubleEG/Run2017D-02Apr2020-v1/NANO AOD
/DoubleEG/Run2017E-02Apr2020-v1/NANO AOD
/DoubleEG/Run2017F-02Apr2020-v1/NANO AOD
/SingleElectron/Run2017B-02Apr2020-v1/NANO AOD
/SingleElectron/Run2017C-02Apr2020-v1/NANO AOD
/SingleElectron/Run2017D-02Apr2020-v1/NANO AOD
/SingleElectron/Run2017E-02Apr2020-v1/NANO AOD
/SingleElectron/Run2017F-02Apr2020-v1/NANO AOD

Table A.3: List of data sets included in the analysis for the 2018 data taking period.

Data set
/SingleMuon/Run2018A-02Apr2020-v1/NANO AOD
/SingleMuon/Run2018B-02Apr2020-v1/NANO AOD
/SingleMuon/Run2018C-02Apr2020-v1/NANO AOD
/SingleMuon/Run2018D-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2018A-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2018B-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2018C-02Apr2020-v1/NANO AOD
/DoubleMuon/Run2018D-02Apr2020-v1/NANO AOD
/DoubleMuonLowMass/Run2018A-02Apr2020-v1/NANO AOD
/DoubleMuonLowMass/Run2018B-02Apr2020-v1/NANO AOD
/DoubleMuonLowMass/Run2018C-02Apr2020-v1/NANO AOD
/DoubleMuonLowMass/Run2018D-02Apr2020-v1/NANO AOD
/EGamma/Run2018A-02Apr2020-v1/NANO AOD
/EGamma/Run2018B-02Apr2020-v1/NANO AOD
/EGamma/Run2018C-02Apr2020-v1/NANO AOD
/EGamma/Run2018D-02Apr2020-v1/NANO AOD

Table A.4: List of data sets included in the analysis for the 2016 data taking period.

Table A.5: List of data sets included in the analysis for the 2017 data taking period.

Table A.6: List of data sets included in the analysis for the 2018 data taking period.

Monte Carlo Datasets for 2018
/DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/W1jetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/W2jetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/W3jetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/W4JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/WZTo3LNu_TuneCP5_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/WW_4F_TuneCP5_13TeV_amcatnlo_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/WWZ_TuneCP5_13TeV_amcatnlo_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/WZZ_TuneCP5_13TeV_amcatnlo_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/ZZZ_TuneCP5_13TeV_amcatnlo_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/GluGluToContinToZTo2e2mu_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZTo2e2tau_13TeV_TuneCP5_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZTo2mu2tau_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZTo4e_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/GluGluToContinToZTo4e_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_EXT_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZTo4mu_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/GluGluToContinToZTo4mu_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_EXT_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZTo4tau_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_EXT_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/HZJ_HToWW_M125_13TeV_powheg_jhugen74_pythia8_TuneCP5/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext2-v1/NANOAODSIM
/ZZTo4L_13TeV_powheg_pythia8_TuneCP5/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/ggZH_HToTauTau_ZToNuNu_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/ggZH_HToTauTau_ZToQQ_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluZH_HToWW_M125_13TeV_powheg_pythia8_TuneCP5_PSweights/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/WminusToTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/WplusHToTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/ZHTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/HminusJ_HToWW_M125_13TeV_powheg_jhugen724_pythia8_TuneCP5/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/HWplusJ_HToWW_M125_13TeV_powheg_jhugen724_pythia8_TuneCP5/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M15/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M20/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M25/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M30/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M35/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M40/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M45/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M50/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M55/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2018_dtau_M60/

## Appendix B

# Fake Rate Measurements

These figures show the rest of the datadriven background estimation on the rate at which jets fake tau leptons in QCD, ttbar, and W+jet regions. The y-axis can be interpreted as the percent fake rate.

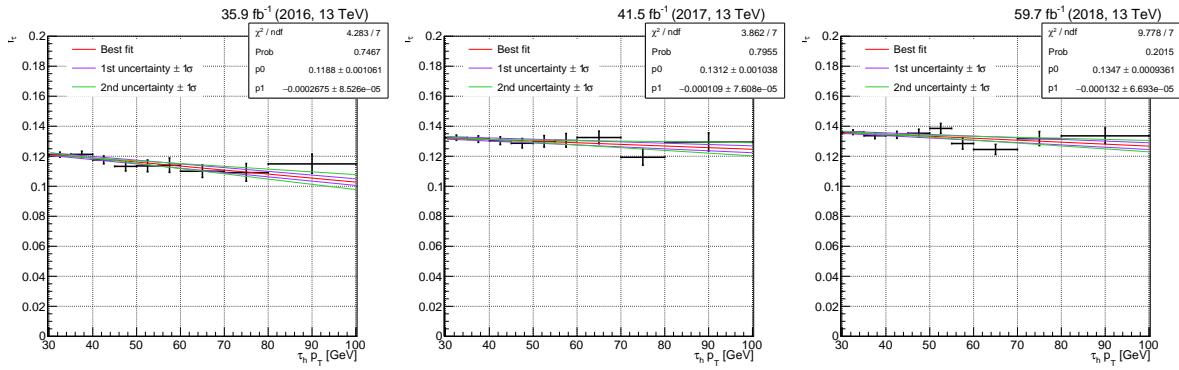


Figure B.1: Fake factors determined in the W+jets determination region with 0 jet in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

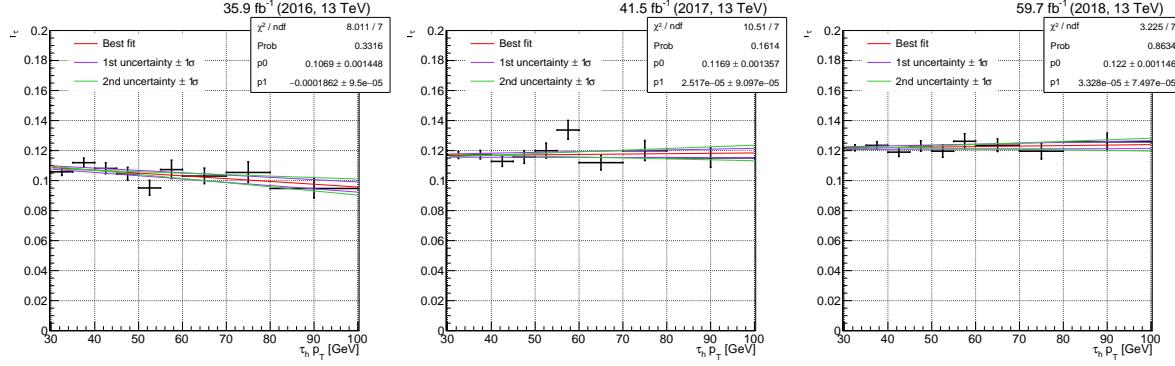


Figure B.2: Fake factors determined in the W+jets determination region with one jet in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

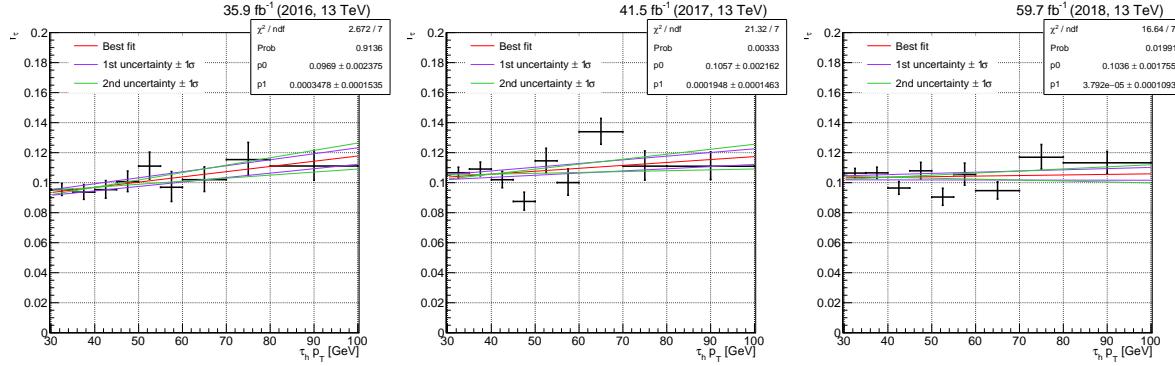


Figure B.3: Fake factors determined in the W+jets determination region with at least two jets in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

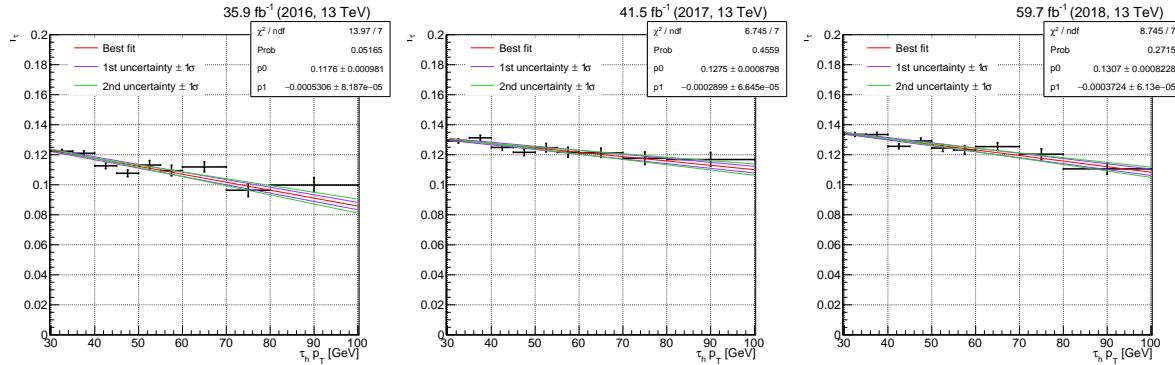


Figure B.4: Fake factors determined in the W+jets determination region with 0 jet in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

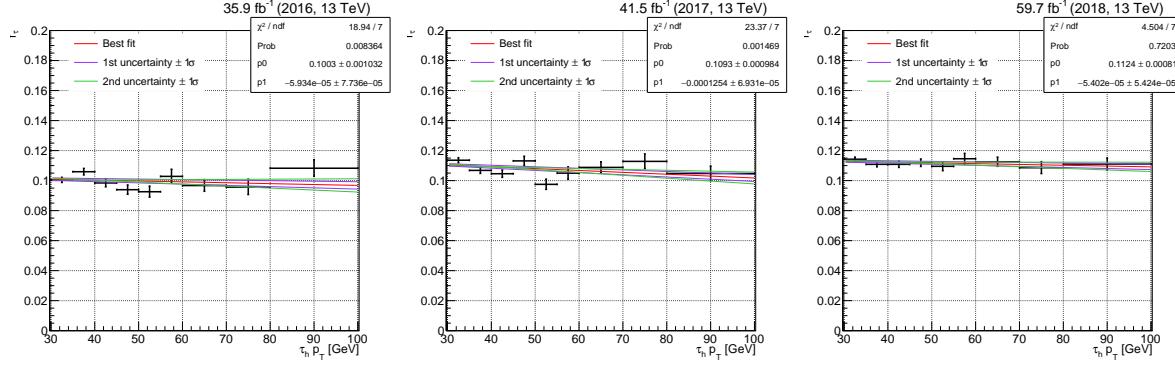


Figure B.5: Fake factors determined in the W+jets determination region with one jet in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

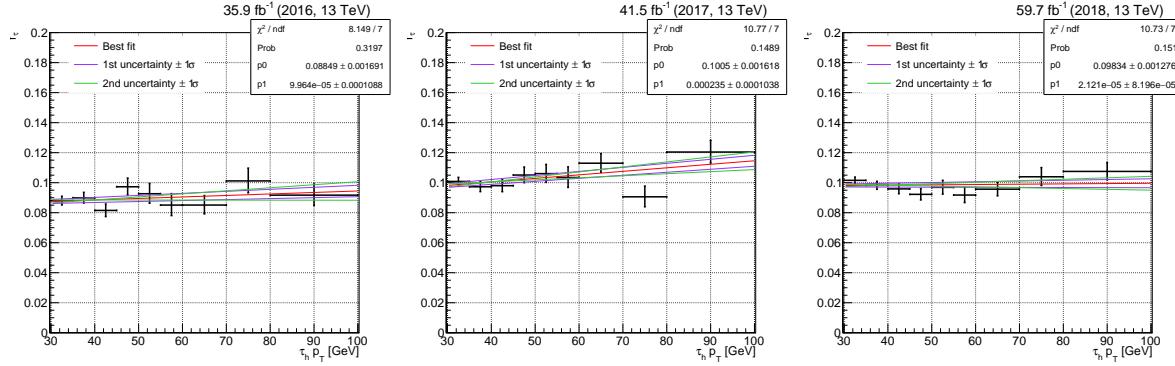


Figure B.6: Fake factors determined in the W+jets determination region with at least two jets in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

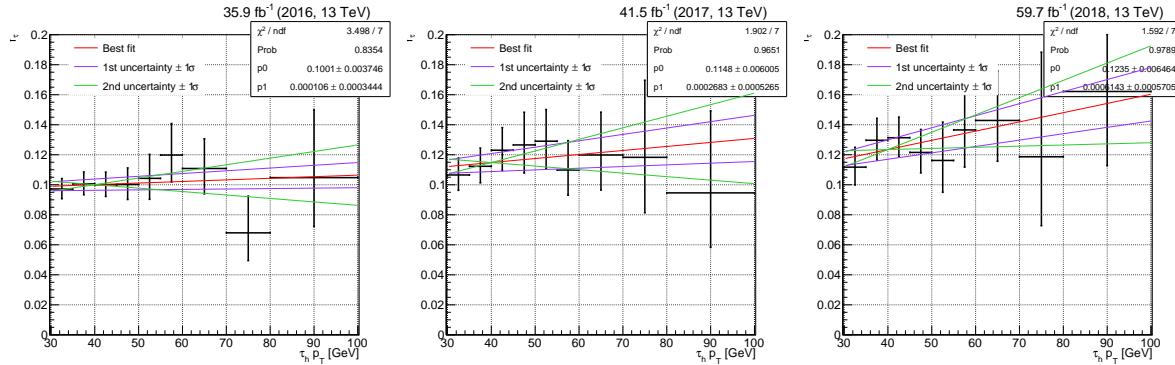


Figure B.7: Fake factors determined in the QCD multijet determination region with 0 jets in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

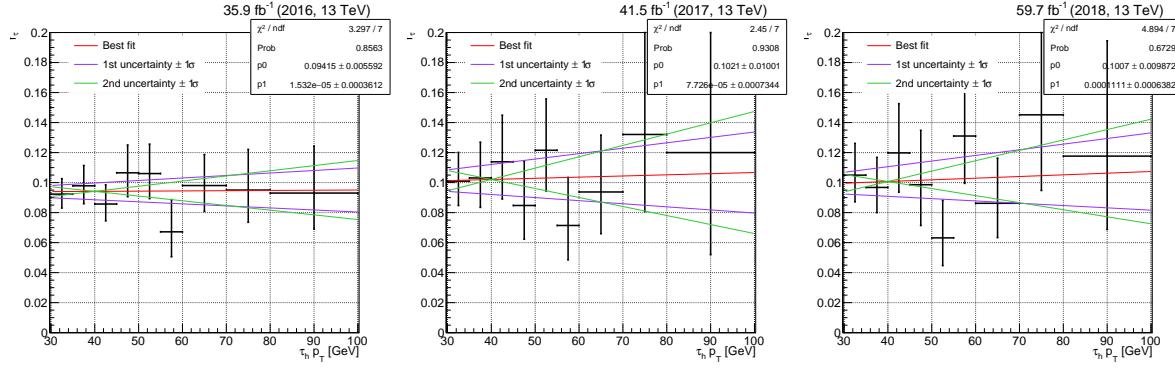


Figure B.8: Fake factors determined in the QCD multijet determination region with one jet in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

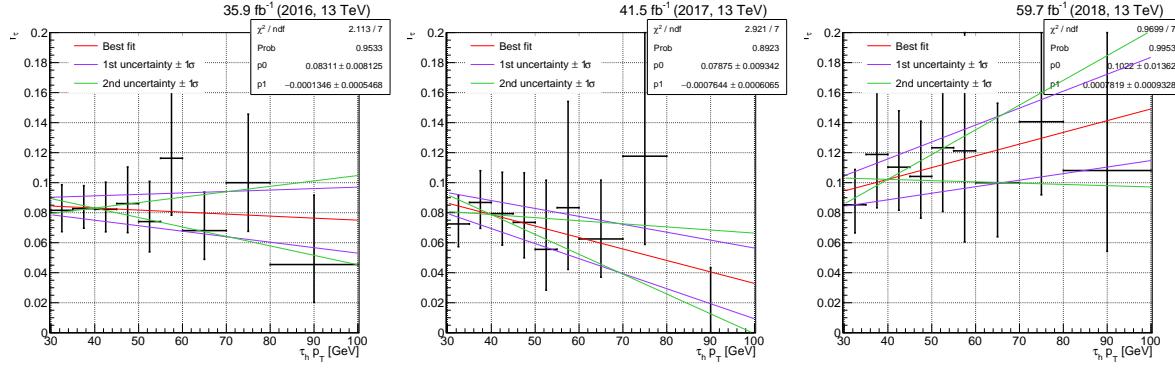


Figure B.9: Fake factors determined in the QCD multijet determination region with at least two jets in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

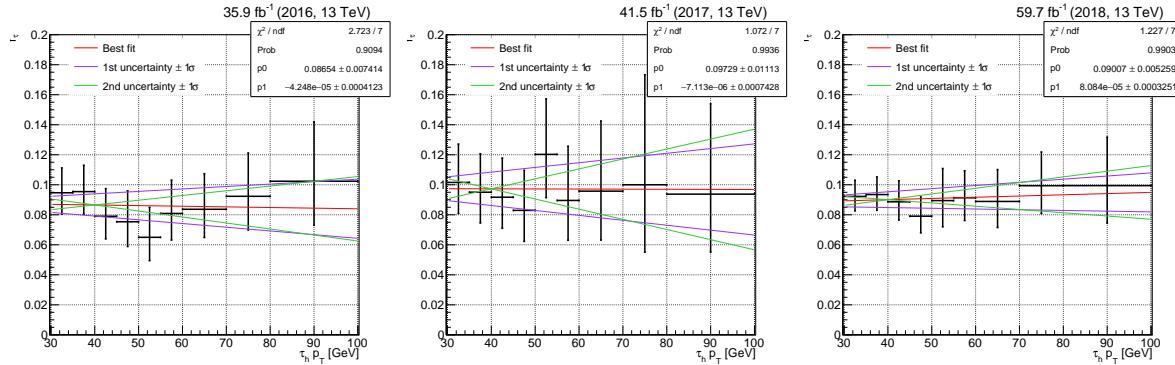


Figure B.10: Fake factors determined in the QCD multijet determination region with one jet in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

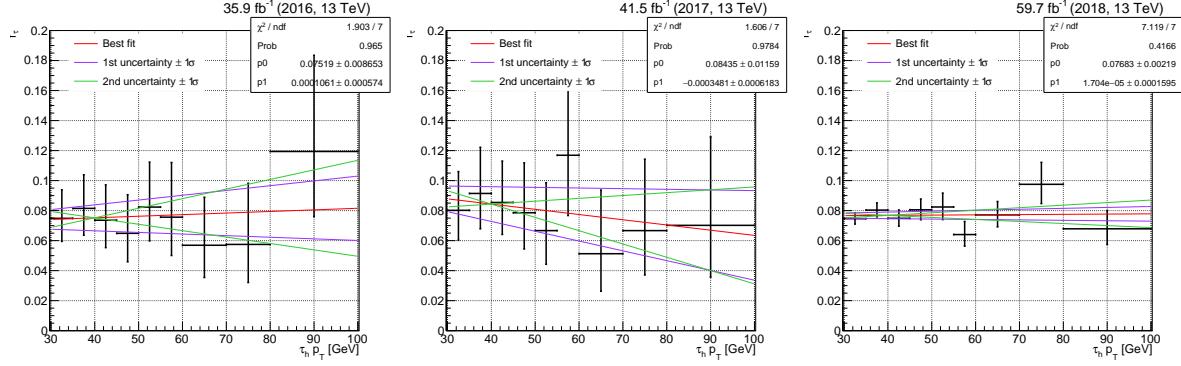


Figure B.11: Fake factors determined in the QCD multijet determination region with at least two jets in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h$   $p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

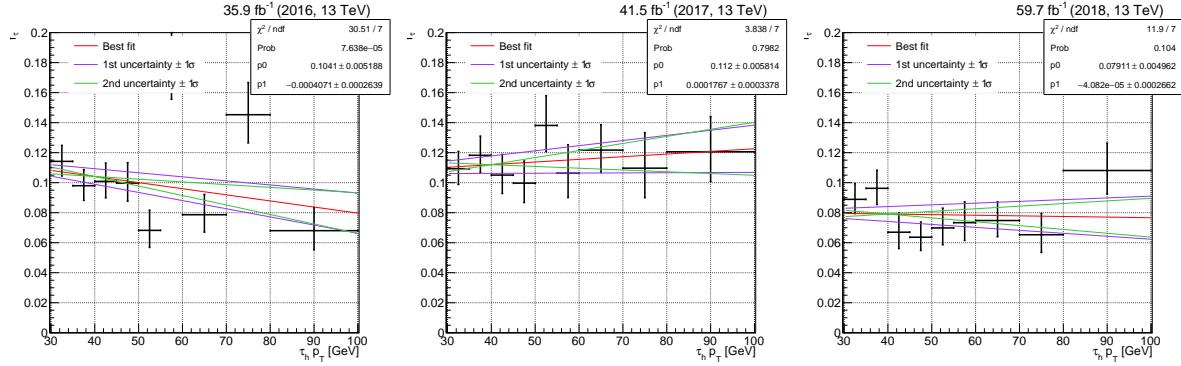


Figure B.12: Fake factors determined in the  $t\bar{t}$  determination region in data in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h$   $p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

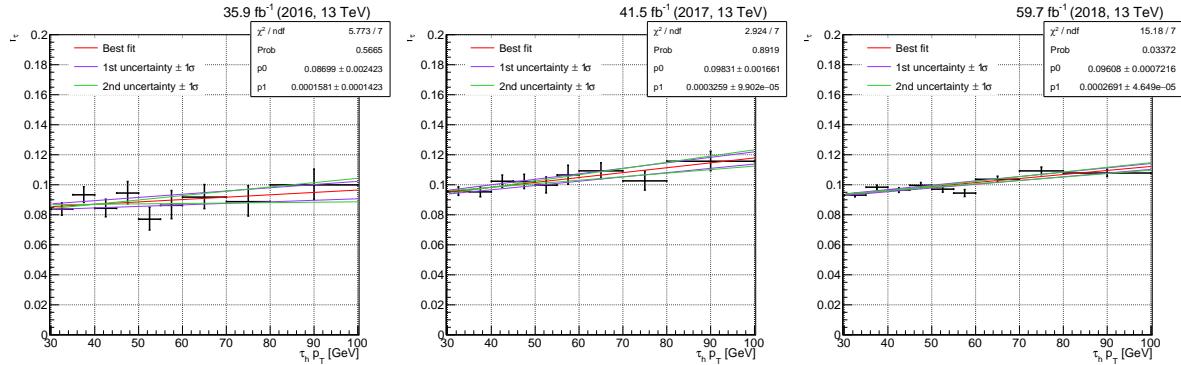


Figure B.13: Fake factors determined in the  $t\bar{t}$  simulation in the  $e\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h$   $p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

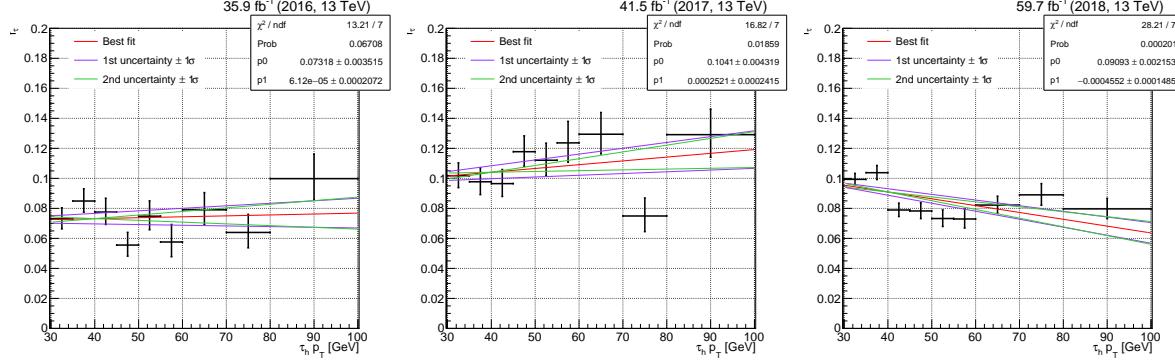


Figure B.14: Fake factors determined in the  $t\bar{t}$  determination region in data in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

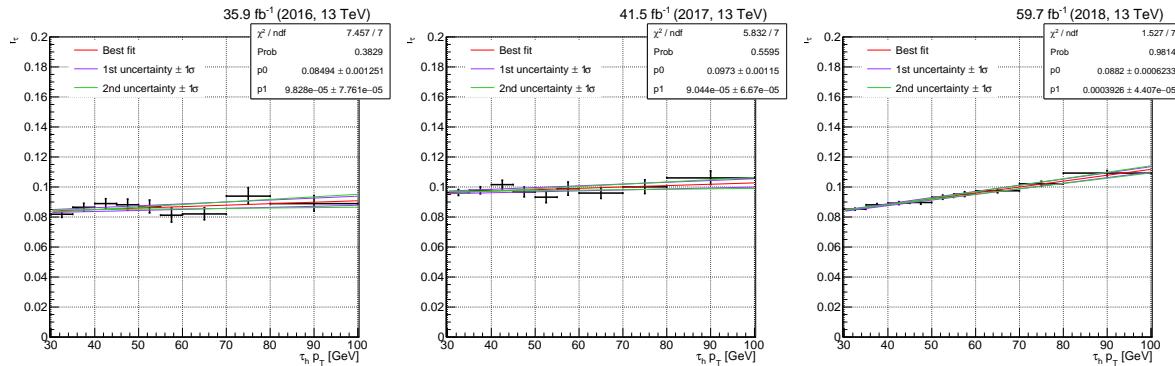


Figure B.15: Fake factors determined in the  $t\bar{t}$  simulation in the  $\tau_h$  final state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the  $\tau_h p_T$ . The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

## Appendix C

# Systematic Uncertainties

Below are the rest of the channels and years from the systematic uncertainty discussion similar to mmmmt 2017 result shown in the main part of the paper 7.23.

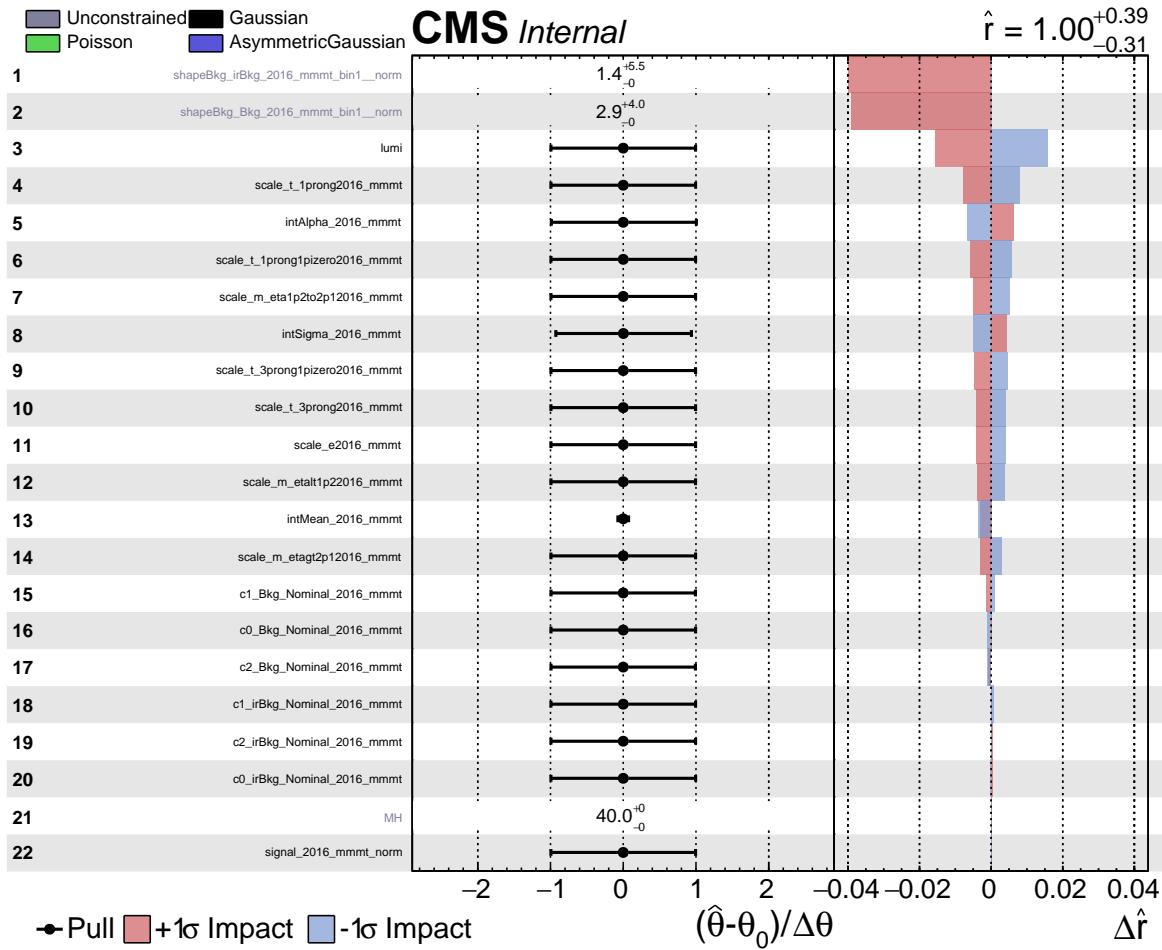
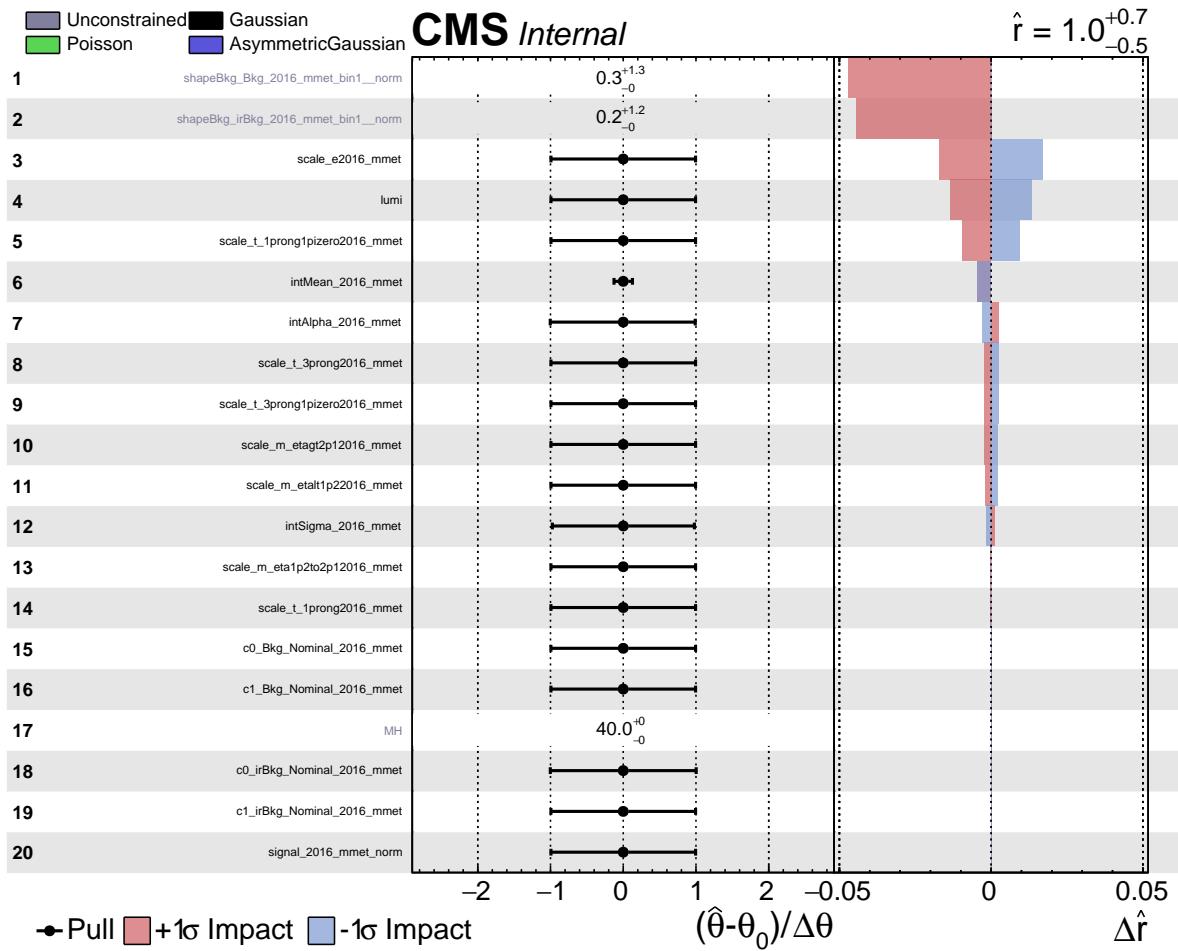


Figure C.1: Expected systematic impacts for the fit model  $\mu\mu\mu\tau$

Figure C.2: Expected systematic impacts for the fit model  $\mu\mu e\tau$

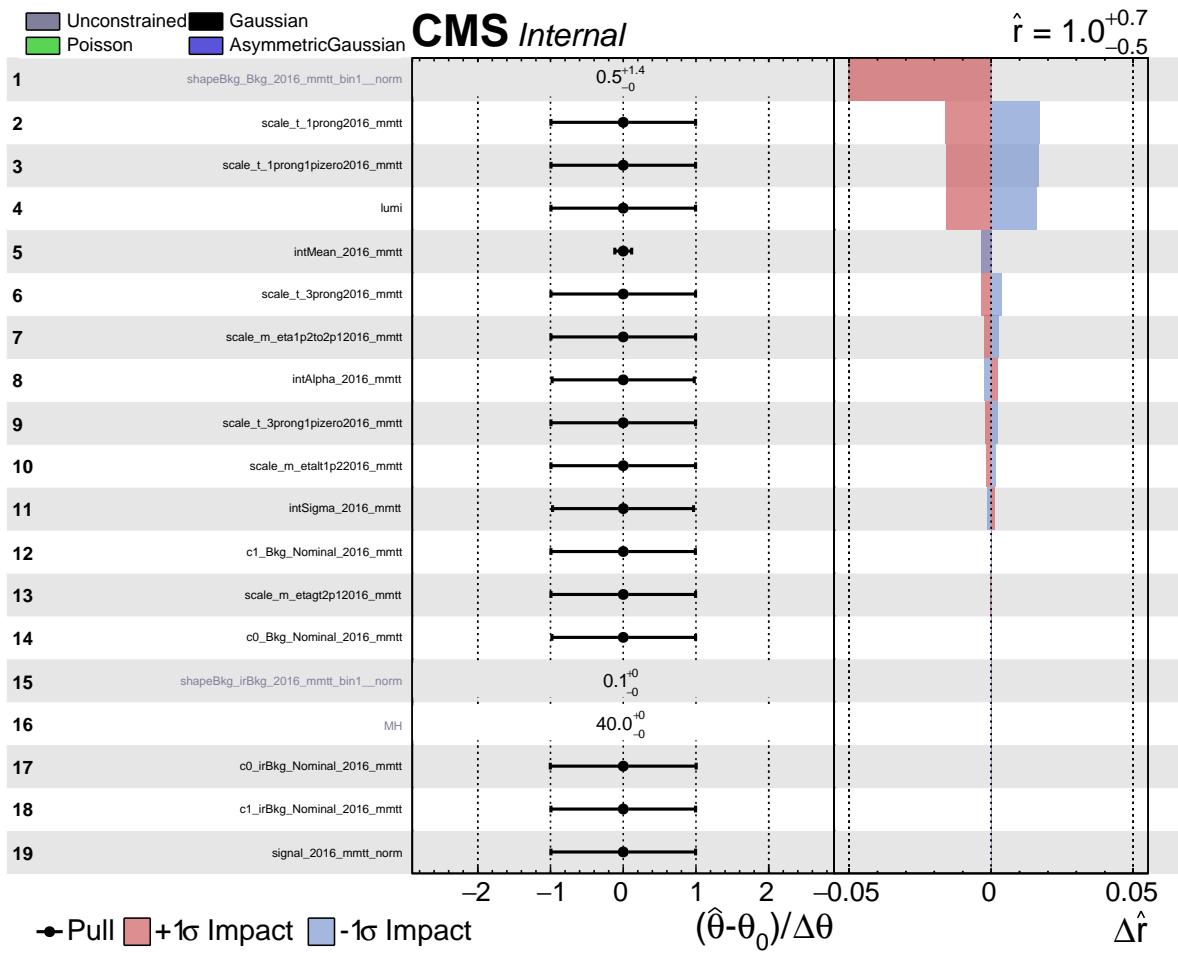


Figure C.3: Expected systematic impacts for the fit model for 2016  $\mu\mu\tau$

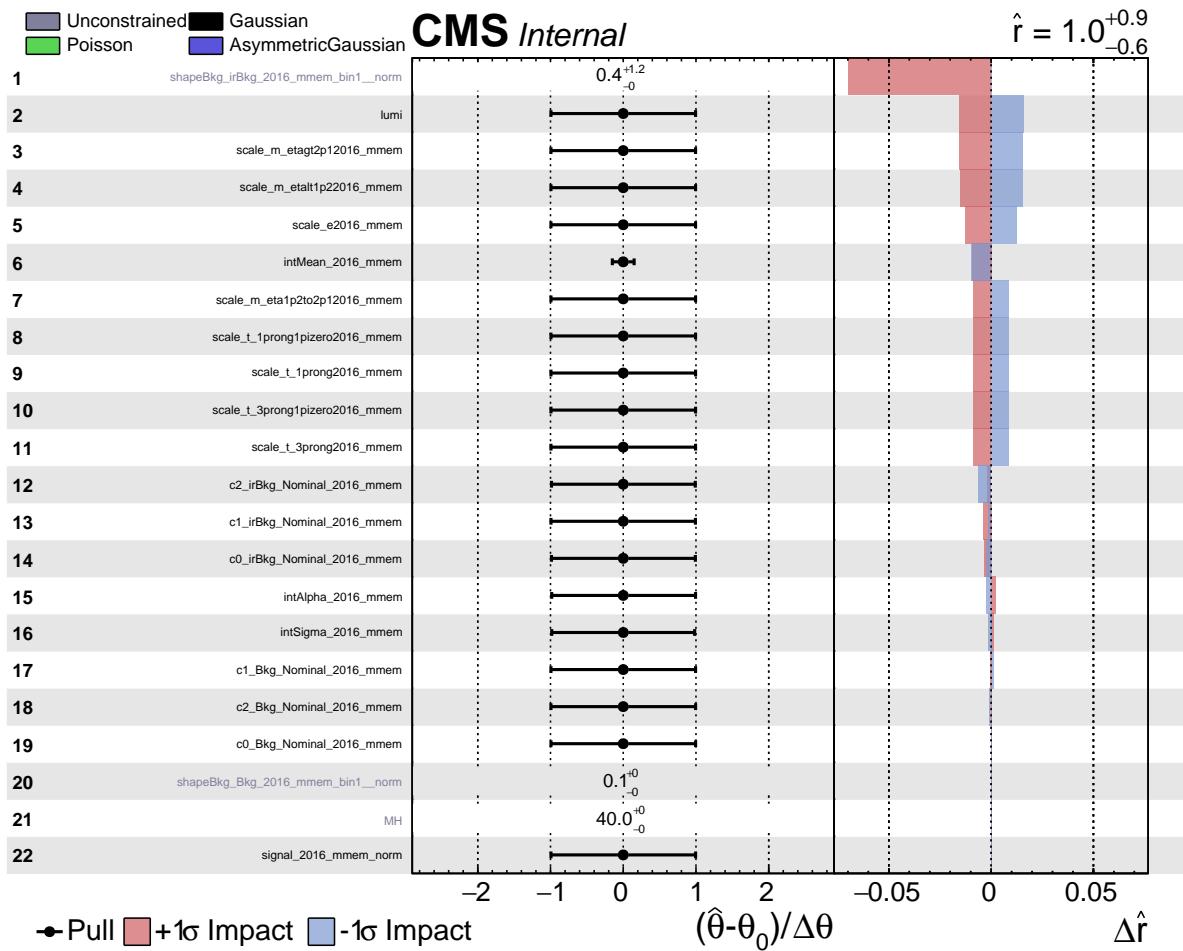


Figure C.4: Expected systematic impacts for the fit model for 2016  $\mu\mu e\mu$

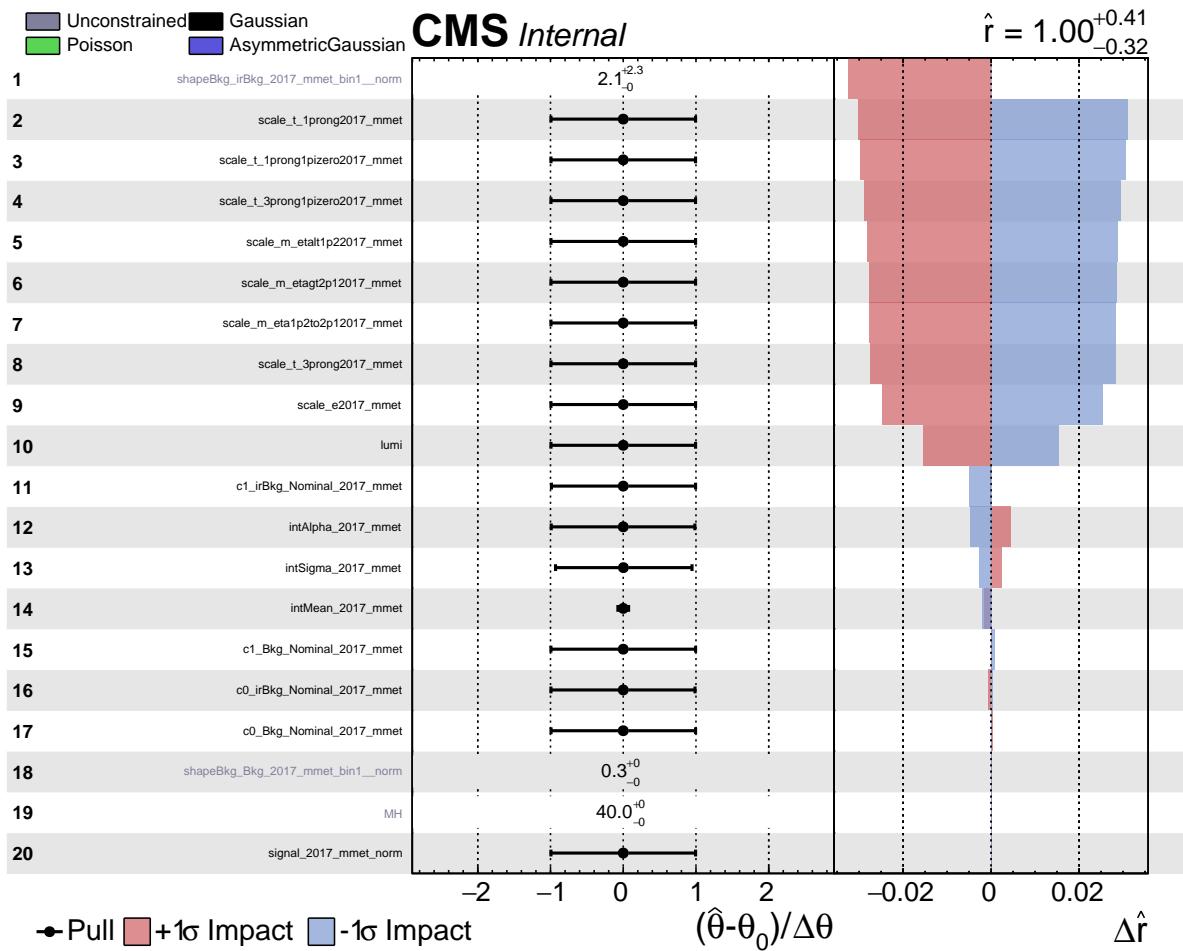


Figure C.5: Expected systematic impacts for the fit model  $\mu\mu\tau$

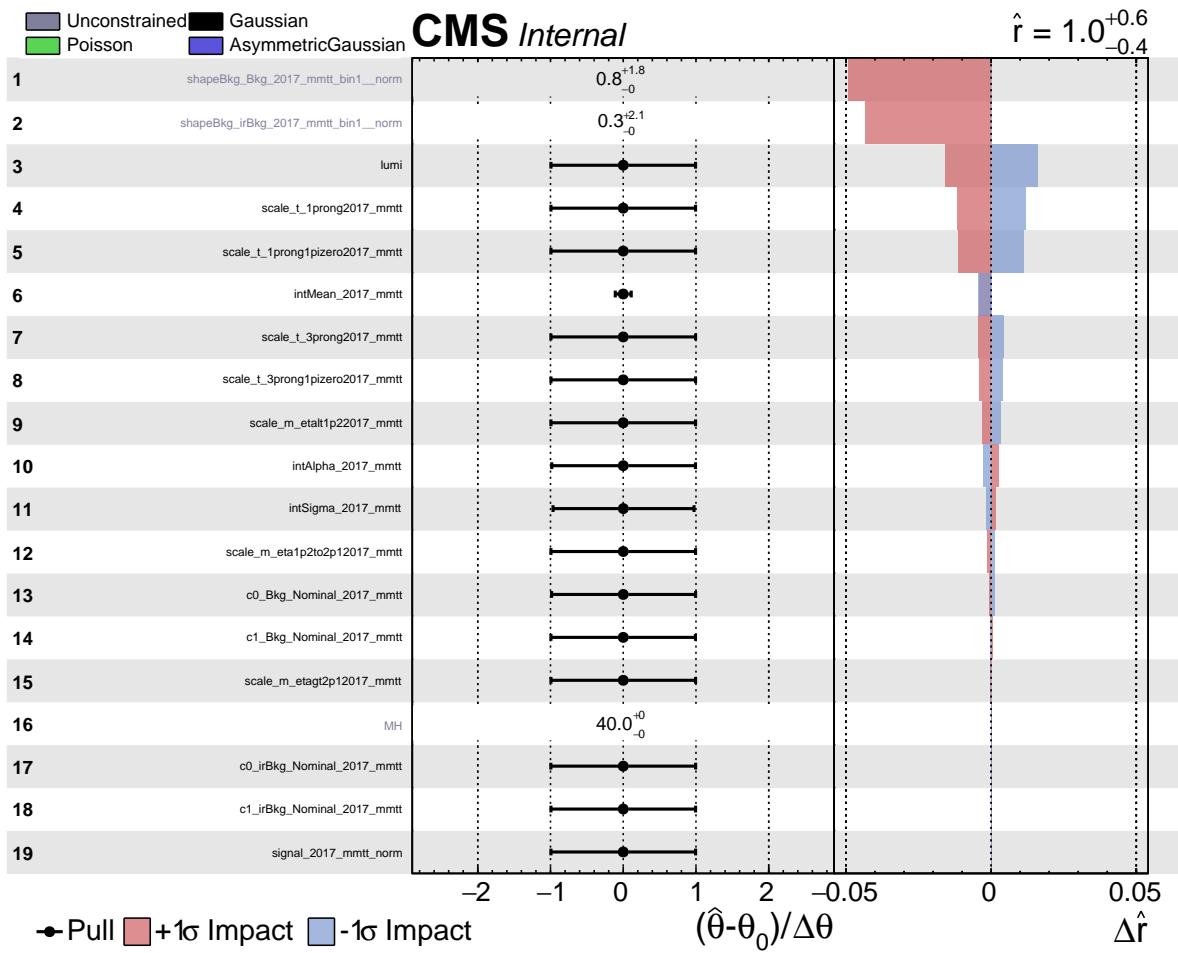


Figure C.6: Expected systematic impacts for the fit model for 2017  $\mu\mu\tau\tau$

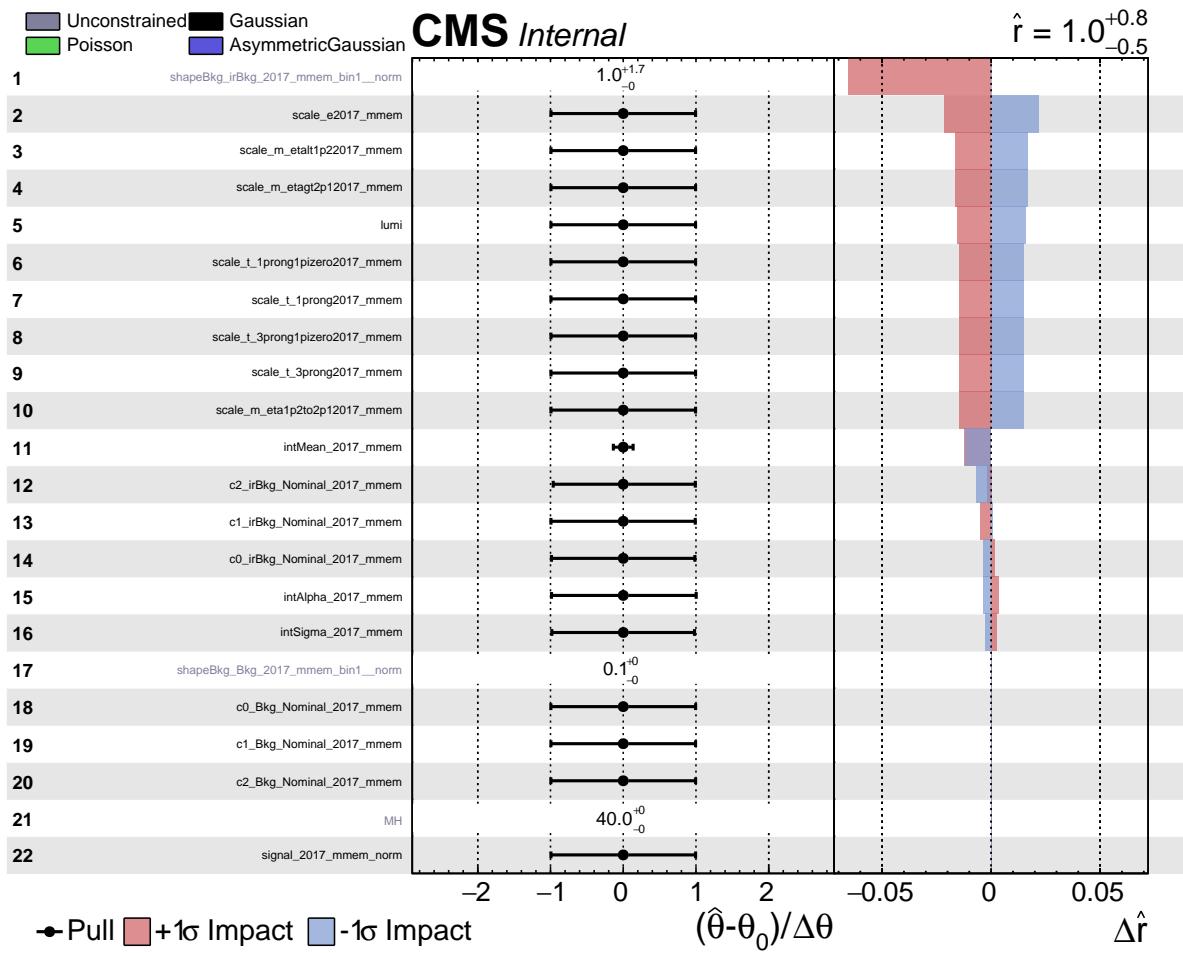


Figure C.7: Expected systematic impacts for the fit model for 2017  $\mu\mu e\mu$

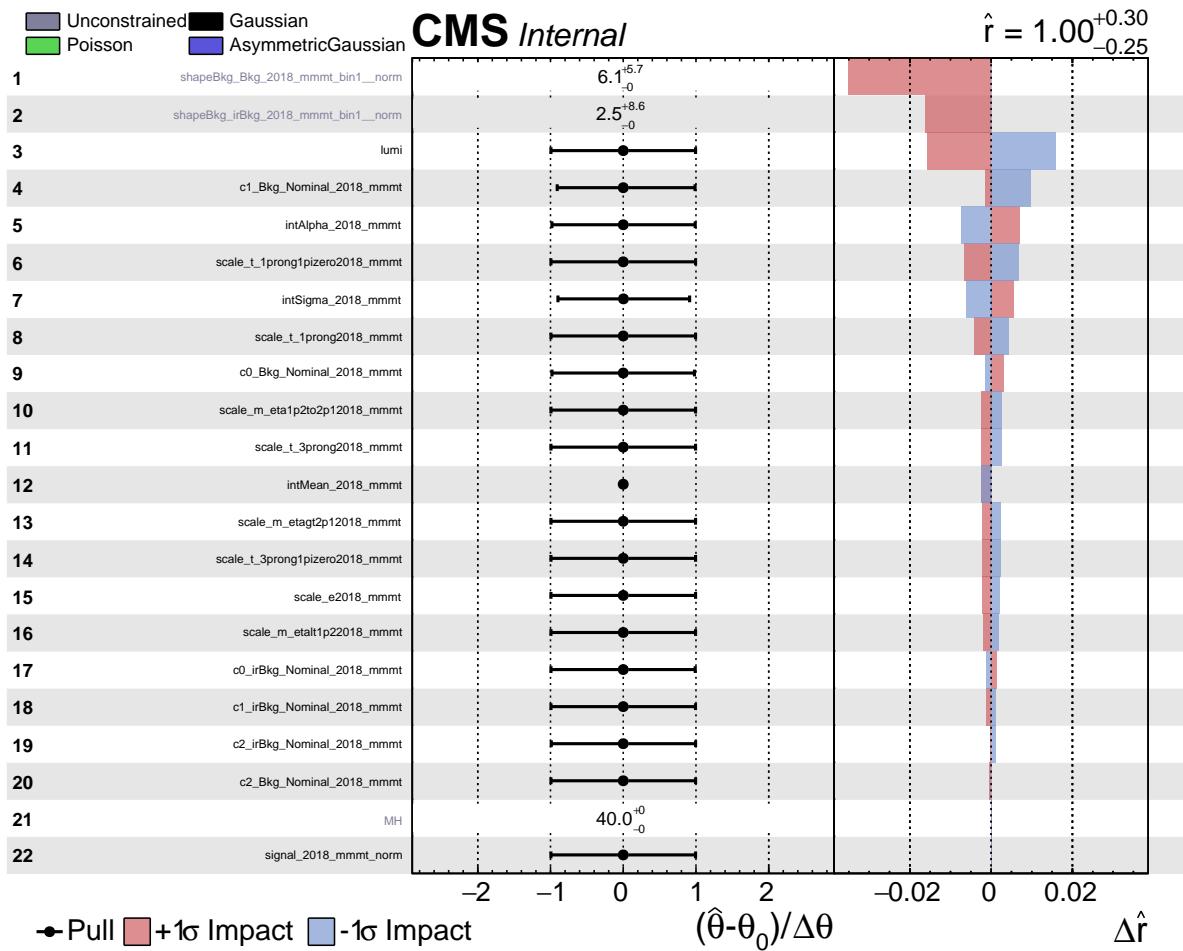


Figure C.8: Expected systematic impacts for the fit model  $\mu\mu\mu\tau$

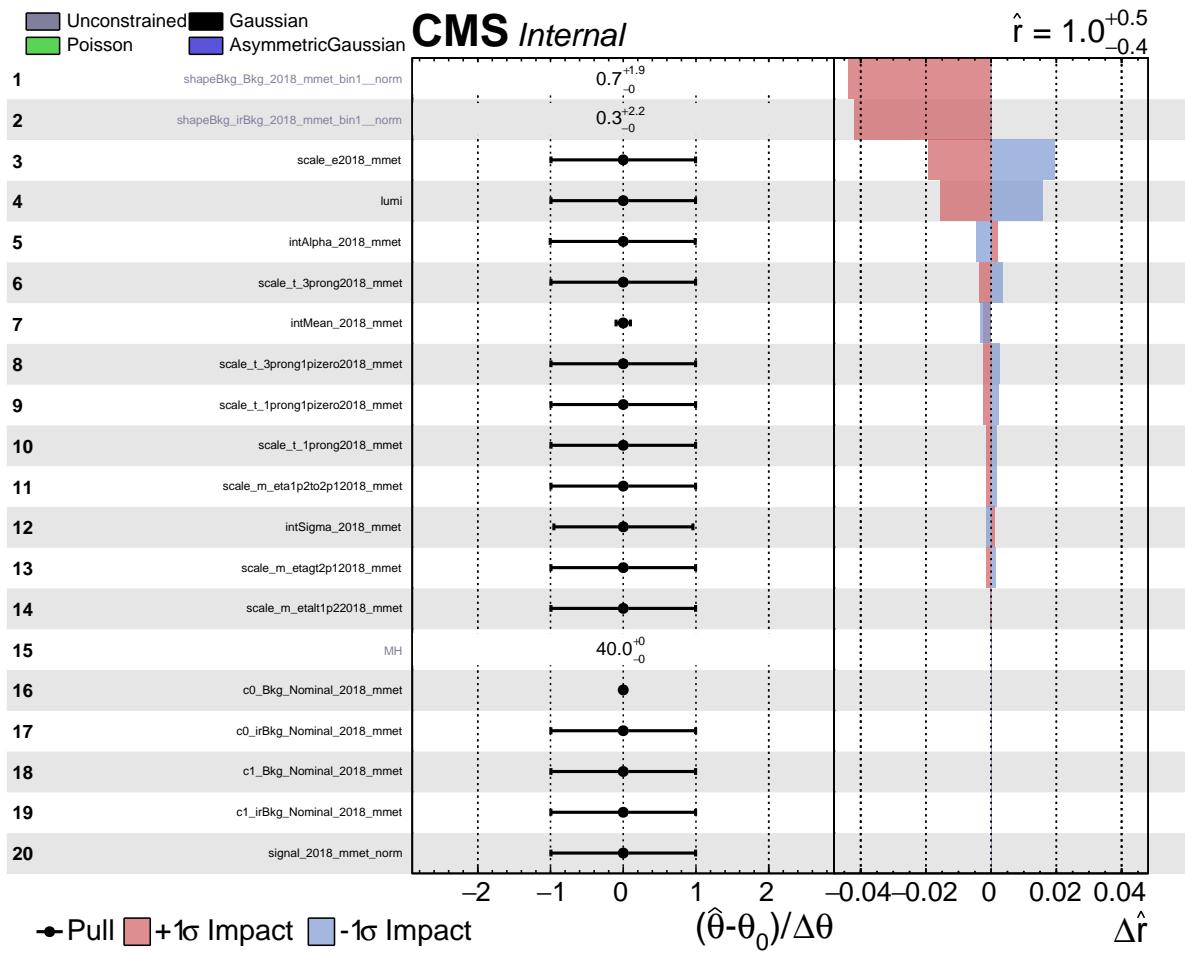


Figure C.9: Expected systematic impacts for the fit model  $\mu\mu e\tau$

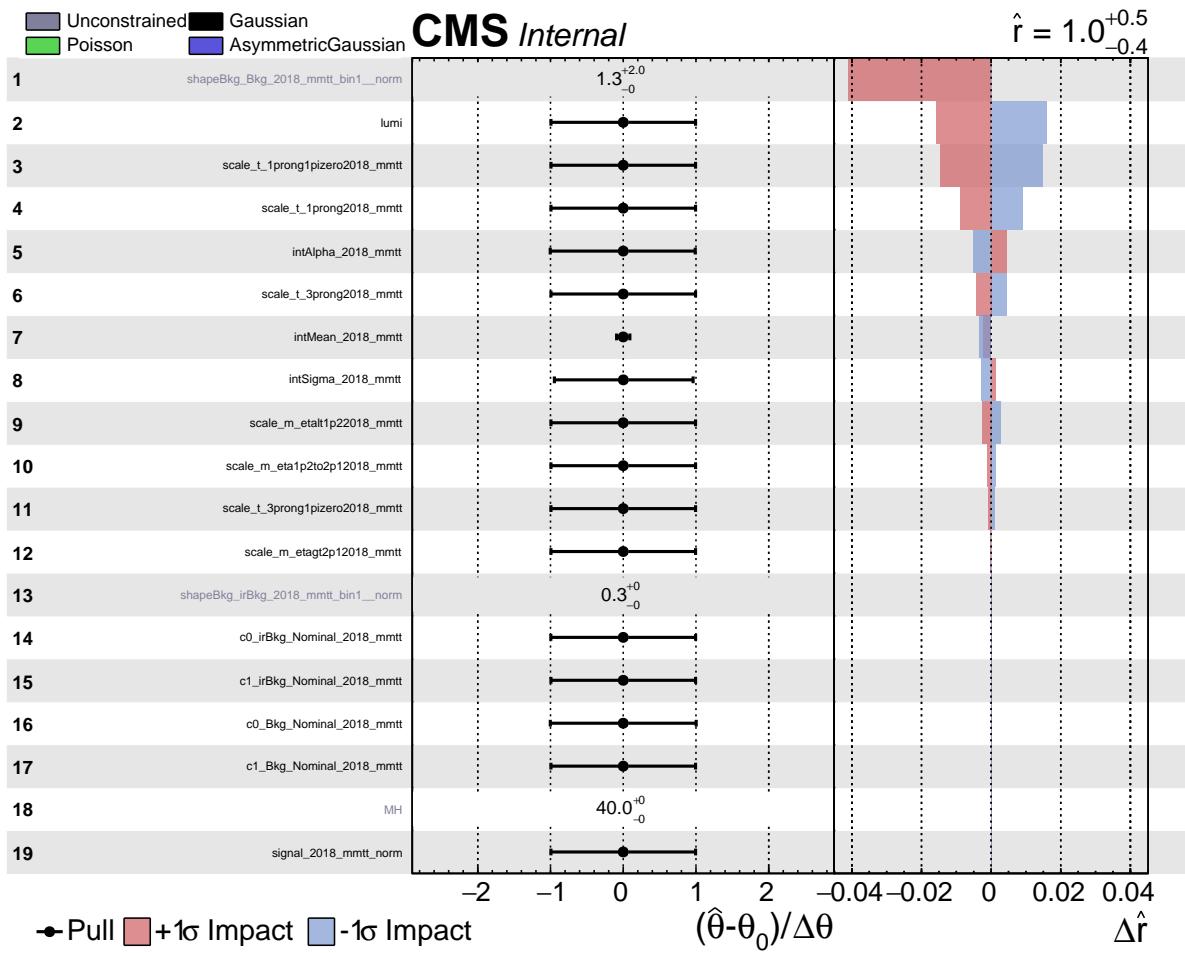


Figure C.10: Expected systematic impacts for the fit model for 2018  $\mu\mu\tau\tau$

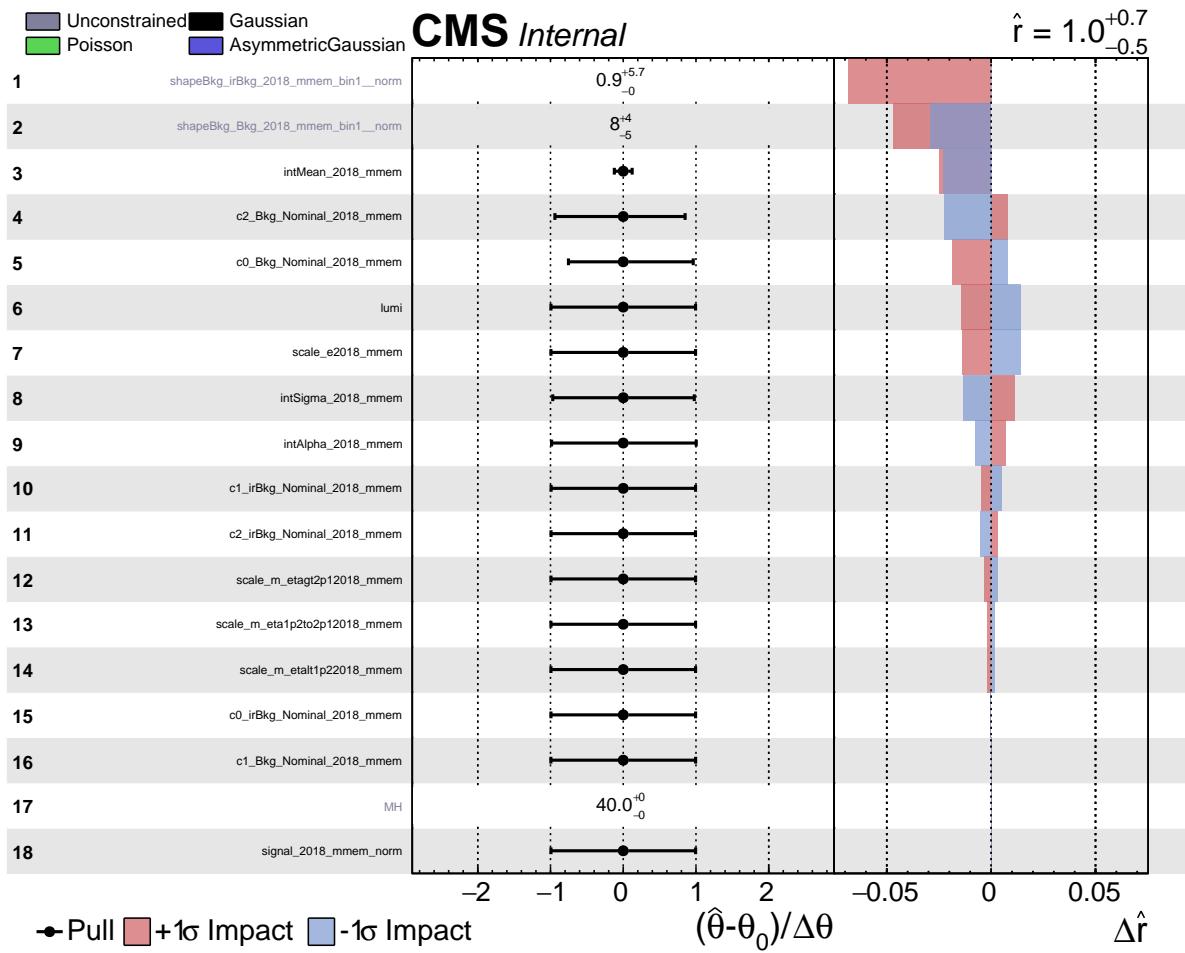


Figure C.11: Expected systematic impacts for the fit model for 2018  $\mu\mu e\mu$

# References

- [1] C. G. Tully, *Elementary particle physics in a nutshell*. Princeton, NJ: Princeton Univ. Press, 2011.
- [2] T. Lancaster and S. J. Blundell, *Quantum field theory for the gifted amateur*. Oxford: Oxford University Press, Apr 2014.
- [3] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, “Theory and phenomenology of two-higgs-doublet models,” *Physics Reports*, vol. 516, pp. 1–102, jul 2012.
- [4] D. Curtin, R. Essig, S. Gori, P. Jaiswal, A. Katz, T. Liu, Z. Liu, D. McKeen, J. Shelton, M. Strassler, Z. Surujon, B. Tweedie, and Y.-M. Zhong, “Exotic decays of the 125 GeV higgs boson,” *Physical Review D*, vol. 90, oct 2014.
- [5] A. M. Sirunyan *et al.*, “Search for an exotic decay of the higgs boson to a pair of light pseudoscalars in the final state of two muons and two  $\tau$  leptons in proton-proton collisions at  $s=13$  tev leptons with the CMS detector,” *Journal of High Energy Physics*, p. 283, 2017.  
$$\sqrt{s} = 13$$
- [6] S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *JINST*, vol. 03, p. S08004, 2008.
- [7] S. Chatrchyan, G. Hmayakyan, V. Khachatryan, A. M. Sirunyan, R. Adolphi, G. Anagnostou, S. Brauer, ... Hansel, T. Nobauer, M. Droege, U. Roser, A. Schonning, F. Stockli, and P. Trub, “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment,” *JINST*, vol. 3, p. S08004. 361 p, 2008. Also published by CERN Geneva in 2010.

- [8] CMS Collaboration, “Particle-flow event reconstruction in CMS and performance for jets, taus, and  $E_T^{\text{miss}}$ ,” CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.
- [9] I. Kratschmer, “Muon reconstruction and identification in CMS Run I and towards Run II,” tech. rep., CERN, Geneva, Oct 2014.
- [10] CMS, “Search for the standard model higgs boson decaying to a pair of tau leptons in 2016 data,” CMS Note 2016/355, 2016.
- [11] V. Khachatryan *et al.*, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV,” *JINST*, vol. 10, p. P06005, 2015.
- [12] A. M. Sirunyan *et al.*, “Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV,” *JINST*, vol. 13, p. P10005, 2018.
- [13] CMS Collaboration, “Identification of hadronic tau lepton decays using a deep neural network,” 2022.
- [14] “Measurement of Higgs boson production in the decay channel with a pair of  $\tau$  leptons,” tech. rep., CERN, Geneva, 2020.
- [15] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *The European Physical Journal C*, vol. 71, feb 2011.