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

September, 2022

Abstract

The Higgs boson was added to the Standard Model (SM) ten years ago, and since then, physicists have been using the Higgs as an integral part of research. Using the Higgs boson to probe for new physics phenomena is part of the high energy experimental frontier. At CERN, the Compact Muon Solenoid (CMS) gathers data from proton-proton collisions to push this frontier further. In this dissertation, a search is conducted for exotic decays of the SM Higgs Boson, h, decaying to a pair of pseudoscalars, a, which then decay to pairs of muons and tau leptons. This search supports many beyond Standard Model (BSM) theories which could solve the μ coupling problem in Super Symmetry and fit into Axion-like Models (Peccei-Quinn) and Grand Unified Theories (GUTs). Due to the model independent nature of this search, the SM is also tested. Pseudoscalar masses between 20 and 60 GeV are investigated using the full Run II dataset collected at CMS corresponding to a luminosity of 137 fb^{-1}. The existence of the pseudoscalar Higgs is primarily motivated by Two Higgs Doublet Models with the extension of a Singlet (2HDM+S). Upper limits on the branching fraction are set.

Acknowledgements

This Ph.D culminates almost 10 years at CERN with CMS. Along the way I've met some truly astounding researchers, scientists, mentors, and friends.

Firstly, I would like to acknowledge Daniel Marlow, my advisor, confidant, and mentor. Dan's philosophical stance on the Ph.D has been freedom with genuine support. His generosity, guidance, and flexibility has been nothing less than profound. Even after working with him closely for six years, I still learn from his wisdom.

Prof. Christopher Tully has also been supportive, honest, and caring during my time at Princeton. I want to express my gratitude for his role in reviewing the thesis as the reader and the support he has given the Princeton CMS group with living overseas.

Special thanks to my committee Prof. Lisanti and Prof Leifer for agreeing to review and conduct the final public oration on short notice and during the busy summer time.

Additionally, I want to thank some more senior researchers that supported my work in the lab and in this analysis. Andre Frankenthal and Alexis Kalogeropoulos were helpful in guidance and providing tools to conduct the Higgs-like search. Christopher Palmer and Jingyu Luo were influential on luminosity operations and beyond. In the lab, Fengwangdong Zhang, Alan Honma, and Alessandro La Rosa made outer tracker research fun and captivating.

Going back, the start of my high energy physics journey was in Indiana. The curiosity I had at the discovery of the Higgs in 2012 guided my interest. At Purdue University, I had the pleasure of jumping into research right after the Higgs discovery by working with Daniela Bortoletto, Kirk Arndt, Gale Lockwood, and Nick Hinton. It was the best introduction to CMS research anyone could have! They welcomed me into the CMS family and pushed me all the way to Princeton.

I want to thank my friends at Princeton who welcomed a wandering Hoosier. Seth Olsen, was my first friend at Princeton and the best party planner I know. His honesty, intensity, and love of

the Rolling Stones knows no bounds. He is an inspiration in adapting, changing, and overcoming.

Luckily, I got to experience CERN with some Princeton folks too. I made fond memories around St. Genis with Gage DeZoort, Nick Haubrich, Bennett Greenberg, Stephanie Kwan, and Gillian Kopp. I have faith that this group will push the frontier of the field forward!

After the summer, I found myself as the only Princeton graduate student staying and living overseas during the covid pandemic. Dylan Teague, Grace Cummings, Andrés Delannoy, Janina Nicolini, Kevin Sedlaczek, Anders Mikkelsen, Arild Velure, Bingxuan Liu, Tal van Daalen, Syed Haider Abidi, Matthias Weber, Vivan Nguyen, and Jeff Shahinan comprise my CERN family. They were my world when the rest of the world shut down. I'm truly grateful for their company.

Dylan Teague deserves special recognition. Not only was he my main editor for this dissertation, but he also taught me more software, analysis, and physics techniques than anyone else. He is a devote friend and confidant. His appetite to learn is only matched by his vigor to climb. He has my highest regard.

... To Mom, Dad, and my dear Sophie.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Standard Model and beyond

Comprising many particles and numerous interactions between their fields, the Standard Model (SM) provides a description of the fundamental interactions of nature. The SM, which is consistent with all empirical evidence, is the leading theory of the universe.

Rather than providing a detailed description of the SM, the purpose of this introduction is to frame the analysis from a theoretical perspective. A table listing components of the SM appears below in figure 1.1.

Particles are represented by fields and interact within the theory. The simplest SM interactions of these fields are found by looking at the Dirac Lagrangian and then establishing $U(1)$ interactions that have a conserved quantity (charge). Noether originally proved that local transformation symmetries imply conserved currents which has extensive implications for particle physics and theories that predict quantum numbers and conserved quantities, such as the charge in the $U(1)$ group [1].

After analyzing the $U(1)$ group, one typically extends the theory to include more fields and structures through higher dimensional groups. The gauge principle will be examined for $U(1)$ and then expanded to $SU(2)$ before taking the product of these groups to form the Weinberg-Salam Model.

Standard Model of Elementary Particles
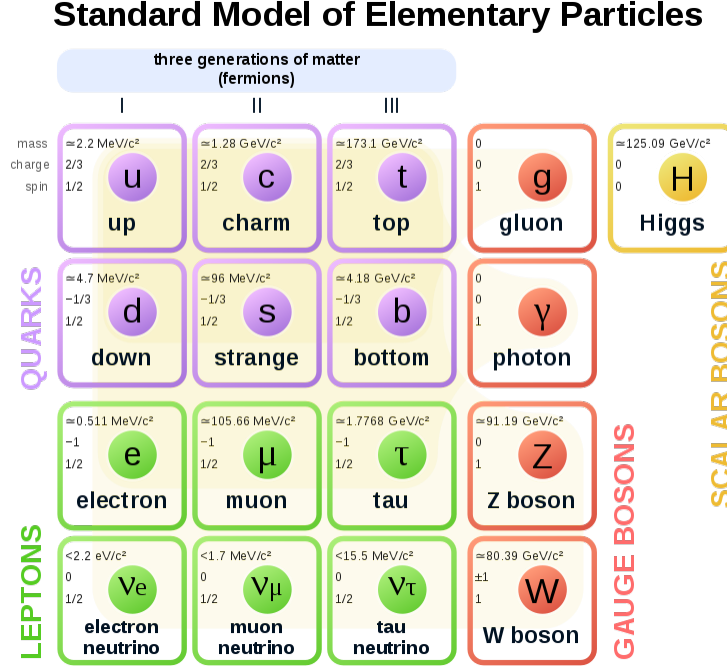
Figure 1.1: SM particles

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 with a transformation. Consider the covariant derivative on a vector field $V(x)$.

$$\mathcal{D}_\mu V(x) \equiv \lim_{\Delta x^\mu \to 0} \frac{V_{||}(x + \Delta x) - V(x)}{\Delta x^\mu} \tag{1.1}$$

How the field changes under the transformation will determine the nature of the field. The connection terms that link the field 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. The parallel component is then:

$$V_{||}(x + \Delta x) = U(x + \Delta x)V(x + \Delta x). \tag{1.2}$$

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

conservation. One defines the transformation and field, then works out the form of the covariant derivative and examines how the gauge field transforms [2].

To put it generally

$$\Psi' = U(\vec{x})\Psi \ \text{ particle under local transformation} \tag{1.3}$$

$$\mathcal{D}^\mu = \partial^\mu + igB^\mu \ \text{ covariant derivative} \tag{1.4}$$

$$B'^\mu = UB^\mu U^{-1} + \frac{i}{g}(\partial^\mu U)U^{-1} \ \text{ gauge field transform.} \tag{1.5}$$

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

$$\Psi'(x) = (1 + \frac{ie}{2}Y \cdot \beta(x))\Psi \ \text{ particle under local transformation} \tag{1.6}$$

$$\mathcal{D}_\mu = \partial_\mu - \frac{ieY}{2}A_\mu \ \text{ covariant derivative} \tag{1.7}$$

$$A^\mu \to A^\mu + \frac{1}{e}(\partial^\mu\beta) \ \text{ gauge field transform.} \tag{1.8}$$

Where Y is the charge-like parameter, e is the magnitude of charge (often combined with Y for Hypercharge), and β is the phase.

Notably, the commutator between the covariant derivatives yields the practical field that it carries. For the $U(1)$ case only, it is the electromagnetic field tensor.

$$[\mathcal{D}^\mu, \mathcal{D}^\nu]\Psi = ieF^{\mu\nu}\Psi \tag{1.9}$$

These relations are important when higher dimensional groups and more complex particle fields are considered. Yang-Mills theories take these components and analyze them 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 $U(1)$.

For $SU(2)$ with a unitary transformation of the form $U(x) = \exp\frac{ig}{2}\boldsymbol{\tau} \cdot \boldsymbol{\alpha}(\boldsymbol{x})$

$$\Psi'(x) = (1 + \frac{ig}{2}\boldsymbol{\tau} \cdot \boldsymbol{\alpha})\Psi \ \text{ particle under local transformation} \tag{1.10}$$

$$\mathcal{D}_\mu = \partial_\mu - \frac{i}{2}g\boldsymbol{\tau} \cdot \boldsymbol{W}_\mu(x) \ \text{ covariant derivative} \tag{1.11}$$

$$\boldsymbol{\tau} \cdot \boldsymbol{W}_\mu \to \boldsymbol{\tau} \cdot \boldsymbol{W}_\mu + \frac{1}{g}\boldsymbol{\tau} \cdot (\partial_\mu\boldsymbol{\alpha}) - \boldsymbol{\tau} \cdot (\boldsymbol{\alpha} \times \boldsymbol{W}_\mu) \ \text{ gauge field transform.} \tag{1.12}$$

Where g is the $SU(2)$ equivalent of e, $\boldsymbol{\tau}$ the Pauli matricies, and $\boldsymbol{\alpha}$ is the higher dimensional phase factor.

Now if we consider states and conditions from $SU(2) \times U(1)$ and add a scalar field, the structure is there for the Weinberg-Salam Model with the Higgs boson.

1.3   The Higgs mechanism

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

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_3 + i\phi_4 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.13}$$

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

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow \begin{pmatrix} e^{ig\frac{\beta}{2}} & 0 \\ 0 & e^{ig\frac{\beta}{2}} \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad U(1) \tag{1.14}$$

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow e^{\frac{ig}{2}\tau \cdot \alpha} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad SU(2) \tag{1.15}$$

and the covariant derivative is then

$$\mathcal{D}_\mu \phi = \partial_\mu \phi - \frac{i}{2} g \tau \cdot \mathbf{W}_\mu \phi - \frac{i}{2} g' B_\mu \phi, \tag{1.16}$$

yielding the Higgs field Lagrangian component

$$\mathcal{L} = (\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi) + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2. \tag{1.17}$$

The Higgs potential contains the typical "Mexican hat" shape. 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 (with L and R being lefthanded and righthanded fermion fields) [3]

$$\mathcal{L} = \bar{L} i \gamma^\mu \mathcal{D}_\mu L + \bar{R} i \gamma^\mu \mathcal{D}_\mu R + (\mathcal{D}^\mu \phi)^\dagger (\mathcal{D}_\mu \phi) \tag{1.18}$$

$$+ \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2 - G_e(\bar{L} \phi R + \bar{R} \phi^\dagger L)$$

$$- \frac{1}{4} G_{\mu\nu}^{(W)} \cdot G^{(W)\mu\nu} - \frac{1}{4} F_{\mu\nu}^{(B)} F^{(B)\mu\nu}. \tag{1.19}$$

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 these fermions mass. This equation,

along with the shape of the Higgs potential, imply spontaneous symmetry breaking and interactions that produce massive vector gauge bosons and massless photons (then the fields are typically denoted $B \to A$, $\boldsymbol{W} \to W^+, W^-, W^0$).

1.4 Higgs doublet models

The SM can naturally be extended by giving more complexity to the fields. Suppose that there were multiple scalar fields instead of just the single Higgs field ϕ. Then one can consider adding another component, thus making it a doublet. Adding the doublet, and working out the relations for the field, creates the most general two-Higgs doublet model (2HDM) with the Higgs potential shown in equation 1.20 [4].

$$V = m_1^2 |H_1|^2 + m_2^2 |H_2|^2 + \frac{\lambda_1}{2}|H_1|^2 + \frac{\lambda_2}{2}|H_2|^2 \tag{1.20}$$
$$+ \lambda_3 |H_1|^2 |H_2|^2 + \lambda_4 |H_1^\dagger H_2|^2$$
$$+ \frac{\lambda_5}{2}\left((H_1 H_2)^2 + c.c.\right) + m_{12}^2 \left(H_1 H_2 + c.c.\right)$$
$$+ \left(\lambda_6 |H_1|^2 (H_1 H_2) + c.c.\right)$$
$$+ \left(\lambda_7 |H_2|^2 (H_1 H_2) + c.c.\right)$$

Expanding around the minimum of the potential yields two doublets with vacuum expectation values v_1 and v_2. They are usually mixed under a rotation parameter $\tan \beta = v_1/v_2$. 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. A denotes the pseudoscalar, h the lighter neutral SM like Higgs, and H^0 the remaining scalar. As with the notation for fields that 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 α

$$\begin{pmatrix} h \\ H^0 \end{pmatrix} = \begin{pmatrix} -\sin\alpha & \cos\alpha \\ \cos\alpha & \sin\alpha \end{pmatrix} \begin{pmatrix} H_{1,R}^0 \\ H_{2,R}^0 \end{pmatrix} \tag{1.21}$$

.

In the literature this parameter is important because $\tan\beta$ and α set the possible couplings to SM particles. Next, if a complex scalar singlet is added, couplings to SM fermions and bosons are

supported [5].

$$S = \frac{1}{\sqrt{2}}(S_R + iS_I) \tag{1.22}$$

The scalar singlet doesn't have Yukawa couplings, but rather couples to $H_{1,2}$. Through its mixing with $H_{1,2}$ the singlet can couple to SM fermions.

A possible coupling that preserves the SM Higgs by keeping θ_a small could be defined as

$$a \equiv \cos\theta_a S_i + \sin\theta_a A, \ \ \theta_a << 1 \tag{1.23}$$

This 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 small frames the Higgs pseudoscalar analysis in a region with little SM resonance—making it also a general search for any beyond SM (BSM) phenomena. There are terms in the effective Lagrangian that support the $h \rightarrow aa$ decays:

$$\mathcal{L} \subset g_{hAA}hAA + \lambda_S|S^2|^2 \tag{1.24}$$
$$\subset g_{hAA}\sin^2\theta_a haa + 4\lambda_S v_s \sin\zeta_1 \cos^2\theta_a haa$$

here ζ 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. The scenarios, supported by the effective Lagrangian 1.24, yield branching fractions as a function of a-mass and $\tan\beta$. These are enumerated below:

- Type I: Fermions couple only to the H_2 field and are independent of $\tan\beta$, the pseudoscalar coupling is 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.

For completeness, the tree level decay width to fermions f for the a is given by equation 1.25.

$$\Gamma(a \to f\bar{f}) = \frac{N_C G_F}{4\sqrt{2}\pi} g_{aff}^2 m_a m_f^2 \sqrt{1 - \frac{4m_f^2}{m_a^2}} \tag{1.25}$$

The coupling parameter g_{aff}^2 depends on the scenario and the fermions. Compelete information can be found in reference [4] and its supporting documentation.
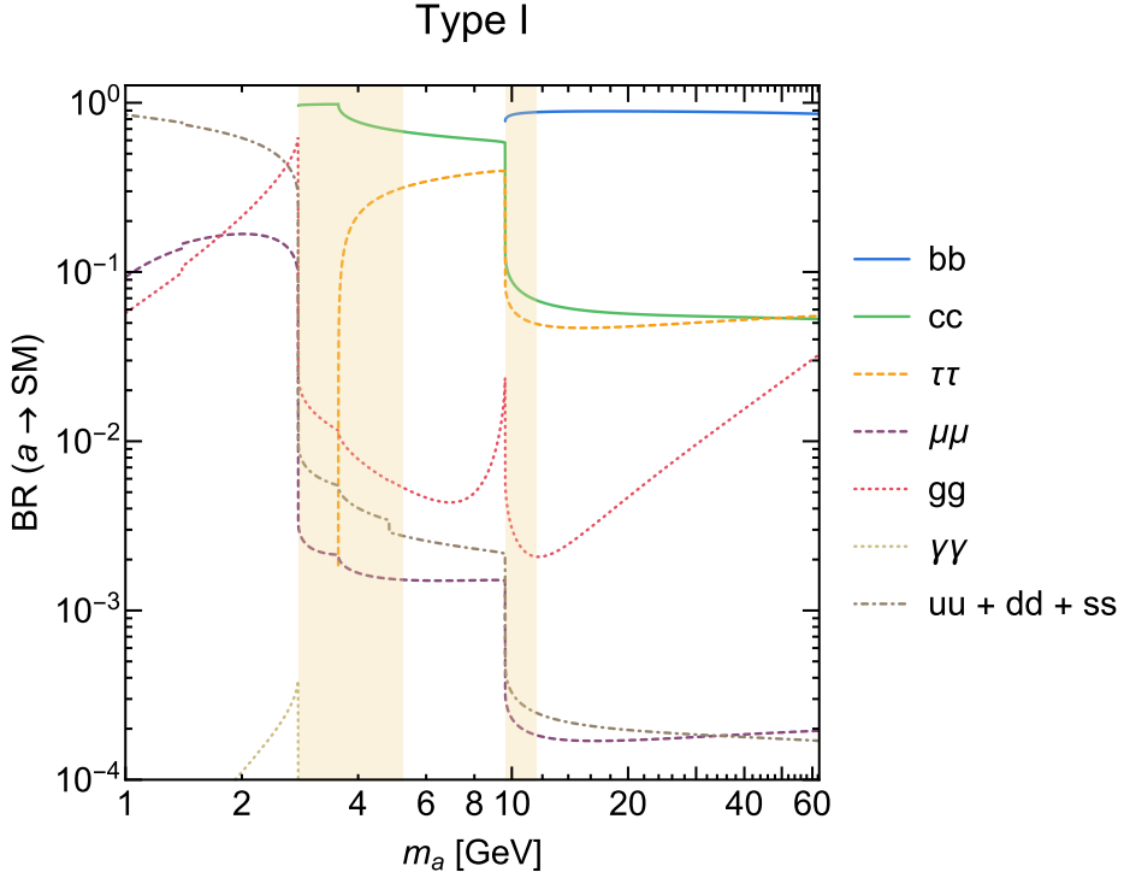
Figure 1.2: Type I 2HDM+S scenario branching fractions [4]

1.5 Previous and present searches in 2HDM+S models

As shown in the previous section, the Higgs couples to all massive SM particles and also to new particles provided that the new particles have mass. BSM theories contain ample room for the Higgs to couple to new particles, making the Higgs an excellent window to investigate any physics beyond the SM.The two Higgs doublet model (2HDM) with its extension of a scalar singlet (2HDM+S) is investigated in this analysis. These types of BSM theories can solve the μ coupling problem in

Super Symmetry (SUSY), while maintaining general support of SUSY (Holomorphy), Axion-like Models (Peccei-Quinn), electroweak baryogensis and several Grand Unified Theories (GUTs) [4].

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 fb^{-1} of data, and new competitive limits were set in reference [6].

Given the potential for improvements in the limits for different 2HDM+S types, extending this search using the full Run II dataset is motivated.

The branching ratios vary based on the value of $\tan\beta$ depending on the type of model under investigation. In particular, Types I-IV are tested. 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. In addition to the search for this model, any deviation from the SM prediction in the effective mass range would also be found.

Type II, tan β = 0.5

Type II, tan β = 5
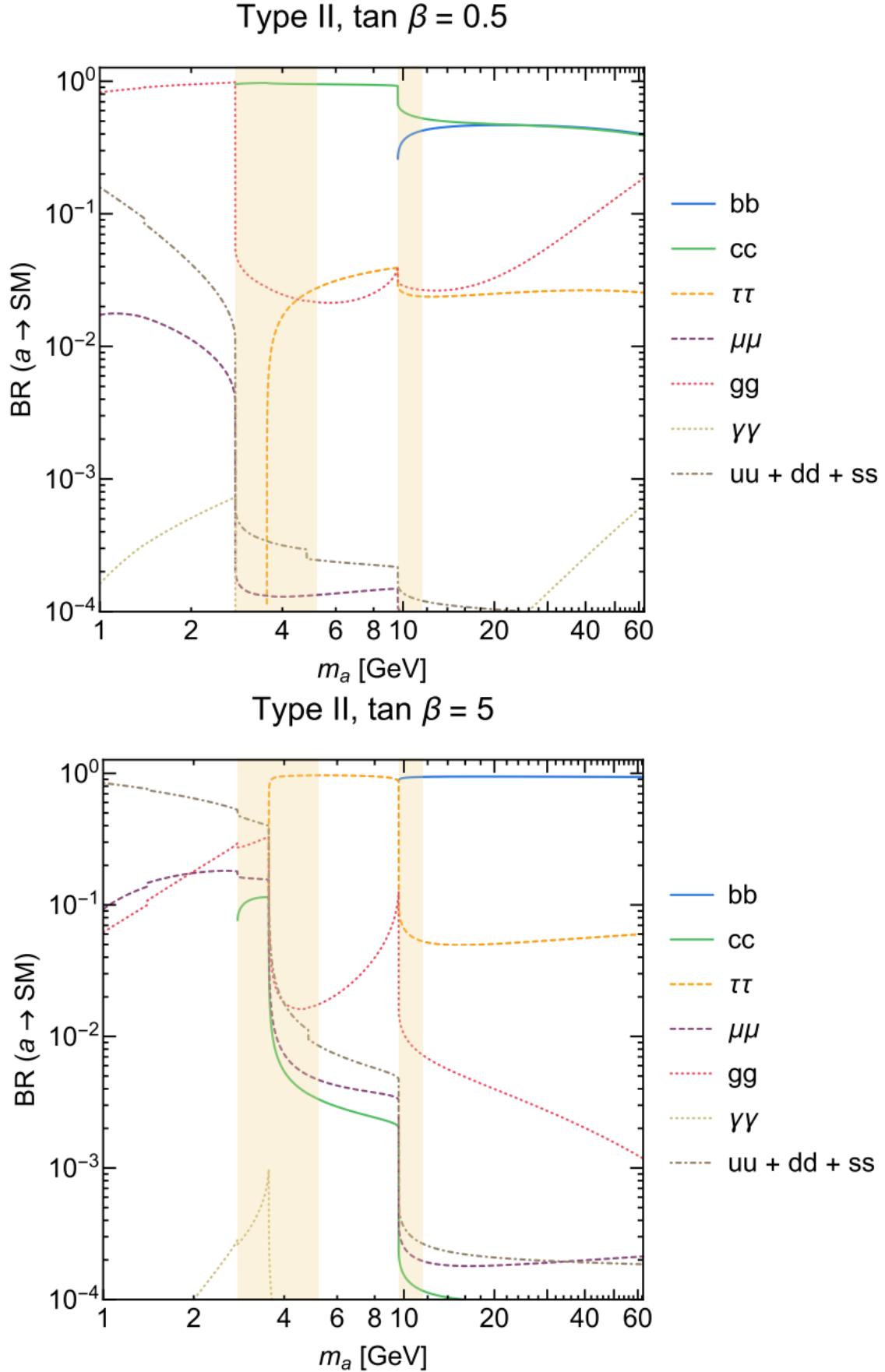
Figure 1.3: Type II 2HDM+S scenario branching fractions [4]

Type III, tan β = 0.5

Type III, tan β = 5
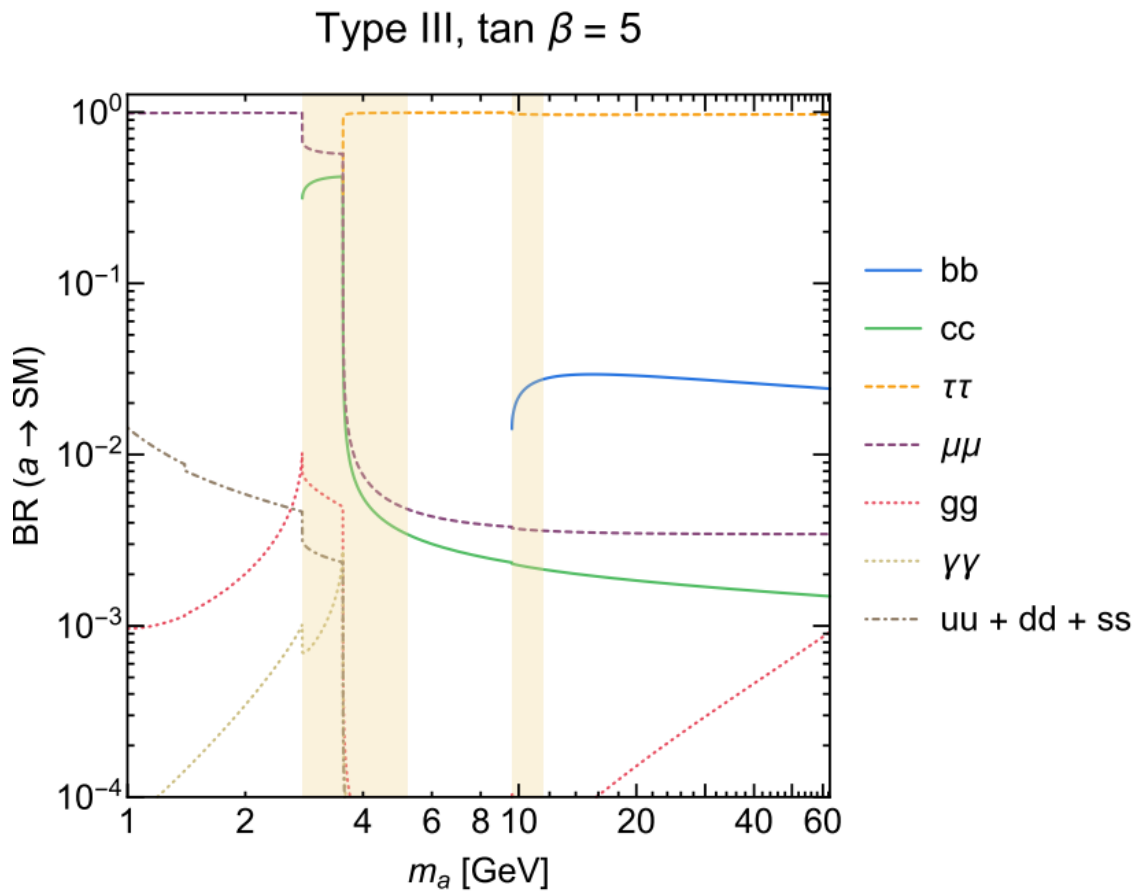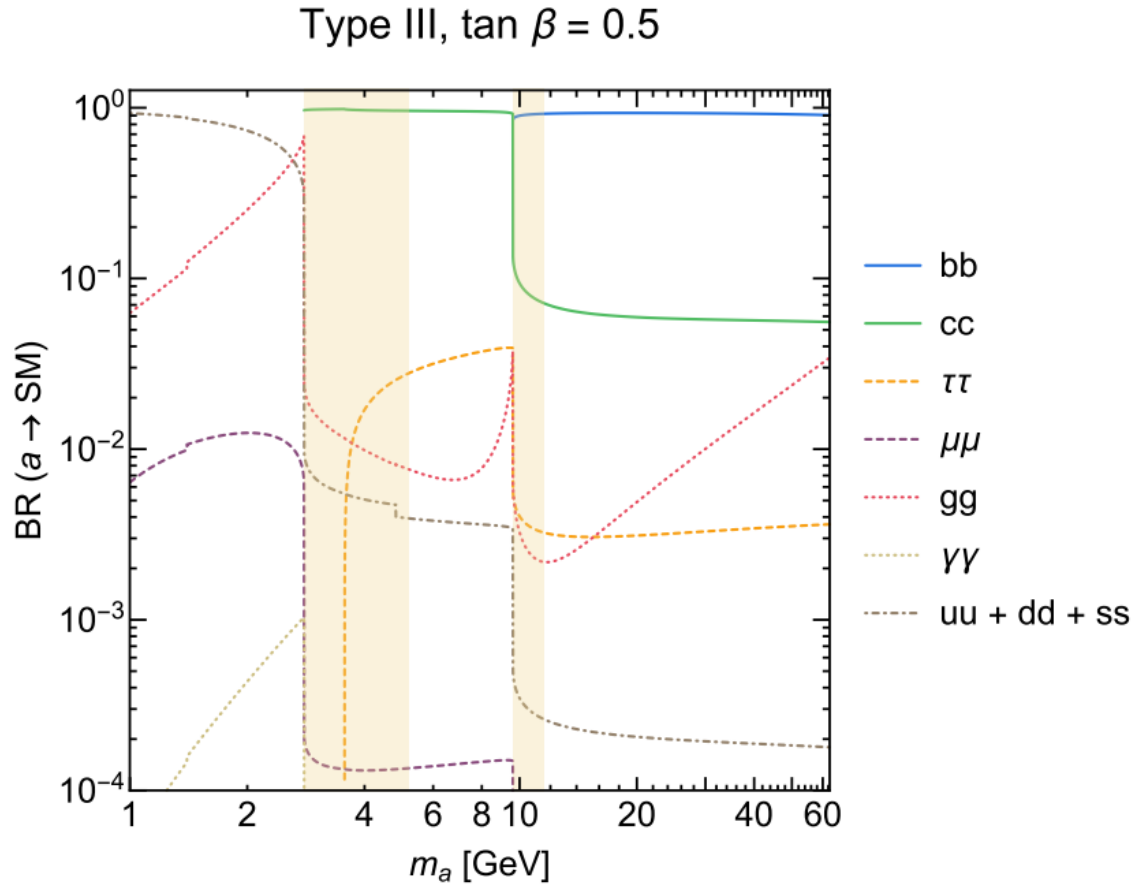
Figure 1.4: Type III 2HDM+S scenario branching fractions [4]
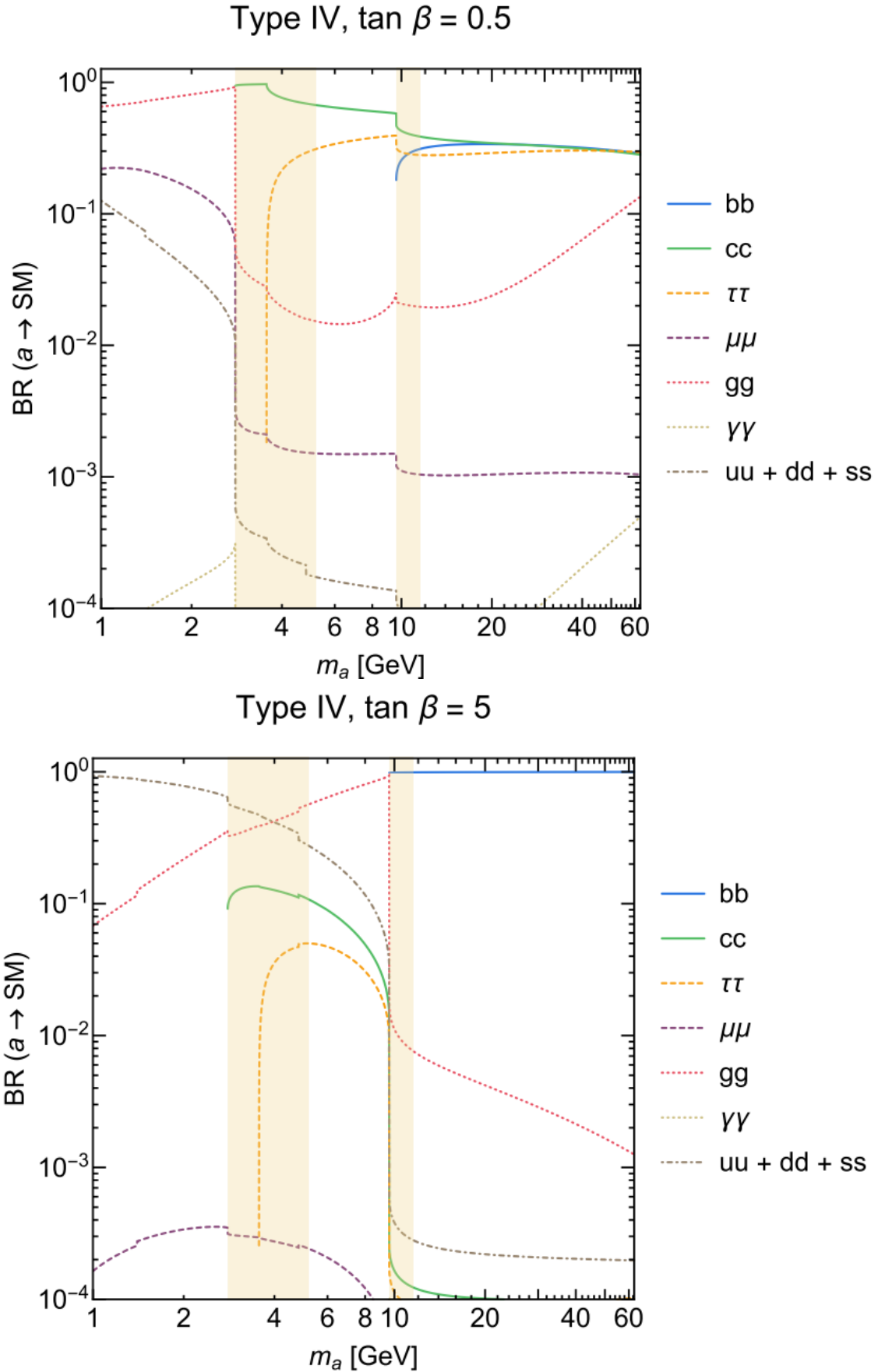
Figure 1.5: Type IV 2HDM+S scenario branching fractions [4]
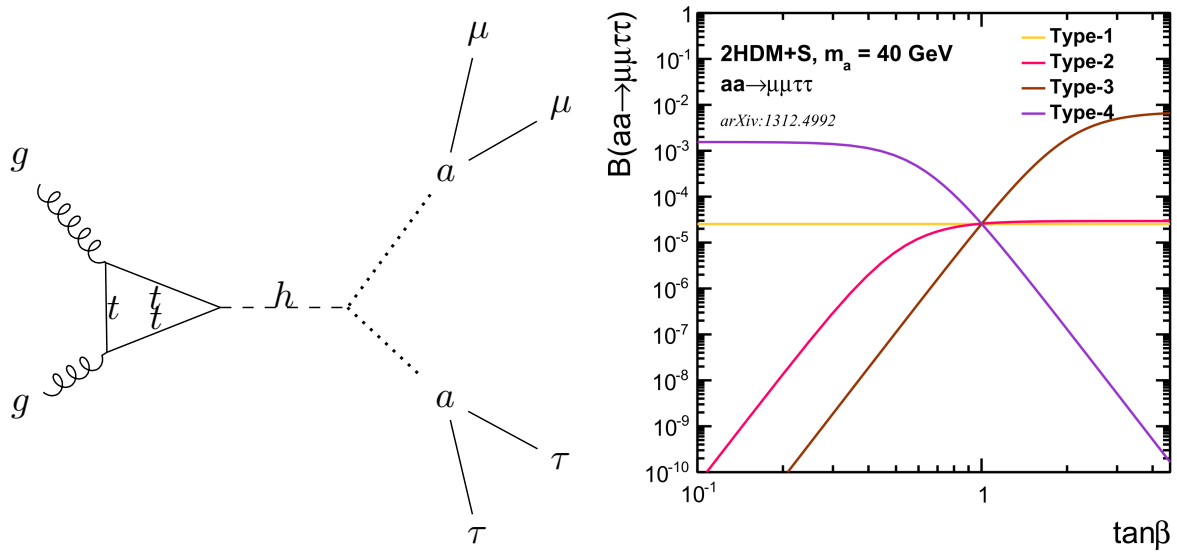
Figure 1.6: Diagram of SM Higgs decay (denoted h here) to pseudoscalar a particles (Left) and branching ratios for pseudoscalar production in different $\tan\beta$ scenarios and different 2HDM+S Types (right)

Chapter 2

CERN, The LHC, and The CMS Detector

2.1 CERN and the Large Hadron Collider

The *organisation européenne pour la recherche nucléaire* or CERN conducts the world's frontier particle physics experiments. Scientists at CERN represent numerous countries who work together for a greater understand of the universe. CERN hosts the Large Hadron Collider (LHC), currently the largest particle collider in the world. The LHC is 27 km in circumference and holds eight experimental caverns 150 meters underneath the earth. Accelerator physicists and engineers strive to provide high energy collisions along these eight sites. More than 1500 superconducting magnets are used to steer and focus the accelerating protons. The beams are brought into collision at four caverns where the experiments are located. The LHC is capable of colliding both protons and heavy ions. The Compact Muon Solenoid (CMS) is the general purpose detector located at collision point 5(P5) in Cessy, France [7].

2.2 The Compact Muon Solenoid detector

At around 14,000 tons, the CMS detector may not seem like it would be compact; however, it is quite dense. As shown in diagram 2.2, the detector contains many subdetectors housed within a powerful solenoid magnet. At a diameter of 6 meters, a combined weight of 12,500 tons, and a field
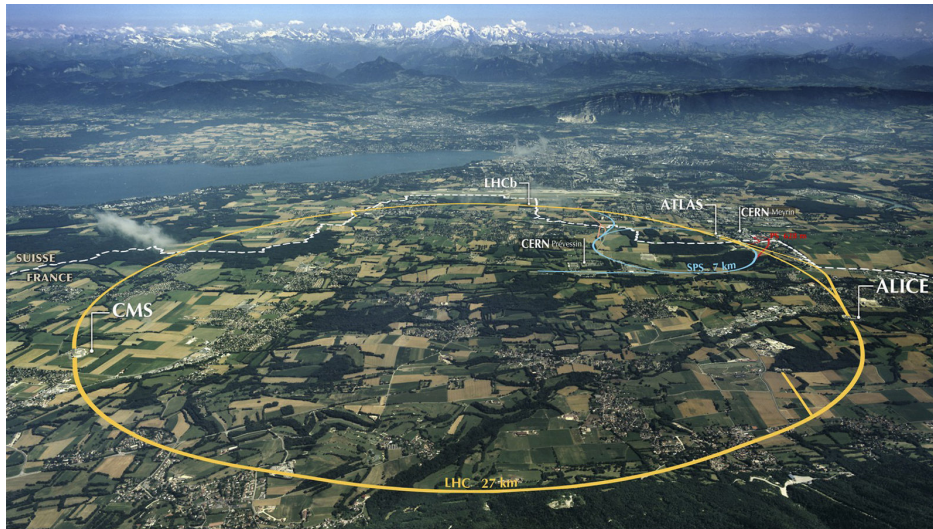
Figure 2.1: Overview of the Large Hadron Collider spanning Switzerland and France [8].

of 3.8 Tesla, the solenoid is the central feature of the Compact Muon Solenoid (CMS). Within the
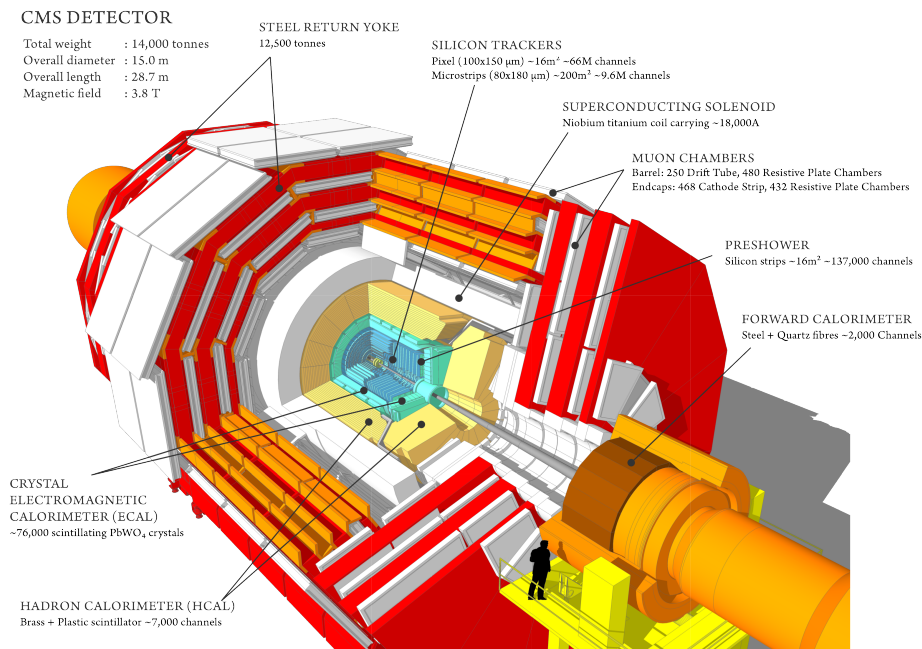
Figure 2.2: The CMS detector full 3D image with all subsystems labeled

solenoid volume, there is the silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass plus scintillator hadron calorimeter (HCAL). Each tracker and calorimeter comprises a barrel and two endcap sections. Gas-ionization chambers are embedded in the steel flux-return yoke outside the solenoid and are used to primarily detect muons.

A coordinate system centered on the nominal collision point is adopted. The y-axis points vertically outward toward the sky, the x-axis tangent to the Earth, and the z-axis along the direction of the beam pipe. The azimuthal angle ϕ and the radial coordinate r in the x and y plane, and the polar angle θ measured from the z axis are typically used to denote space points. Often pseudorapidity, defined by

$$\eta = -\ln\tan(\theta/2) \tag{2.1}$$

is used to describe the angular distance from the beam pipe. A more detailed description of the CMS detector can be found in Reference [9].

2.3 Subdetector systems

Several subdetector systems play an important role in the identification of the muons, electrons, and tau leptons that are used in the analysis. All subdetector systems are important to event reconstruction in CMS with several detectors identifying the particles that are used in this analysis. These are the tracker system, the ECAL, the HCAL, and the muon system.

2.3.1 Tracker

The tracker comprises several groups of silicon detectors. Going outward from the beam pipe, there is the pixel detector and then the silicon tracker. A silicon detector works by sensing the ionization trail left by an energetic charged particle. Typically, multiple band gaps are created through the process of lithography which adds artificial impurities of either p-type (holes) or n-type (electrons). This process is known as "doping". When a minimum ionizing particle (MIP) disturbs the latent charge—set by the bias voltage on the sensor—there is a current generated in the n and p type components which is given by the Shockley equation:

$$\mathcal{J}_{n,p} = \frac{q_0 D_{n,p} d_{n,p}}{L_{n,p}} \left(e^{\frac{q_0 V}{k_B T}} - 1 \right). \tag{2.2}$$

For reference: $L_{n,p}$ is the diffusion length, $D_{n,p}$ the diffusion coefficients, $d_{n,p}$ the charge/hole density, V the bias voltage, q_0 the standard charge unit, k_B the Boltzmann constant, and T the temperature. This current is sensed by the electrodes etched onto the silicon substrate. A graphical display of surface current using simulation as a function of time is shown in figure 2.3.1. A wealth of silicon information can be found in reference [10].
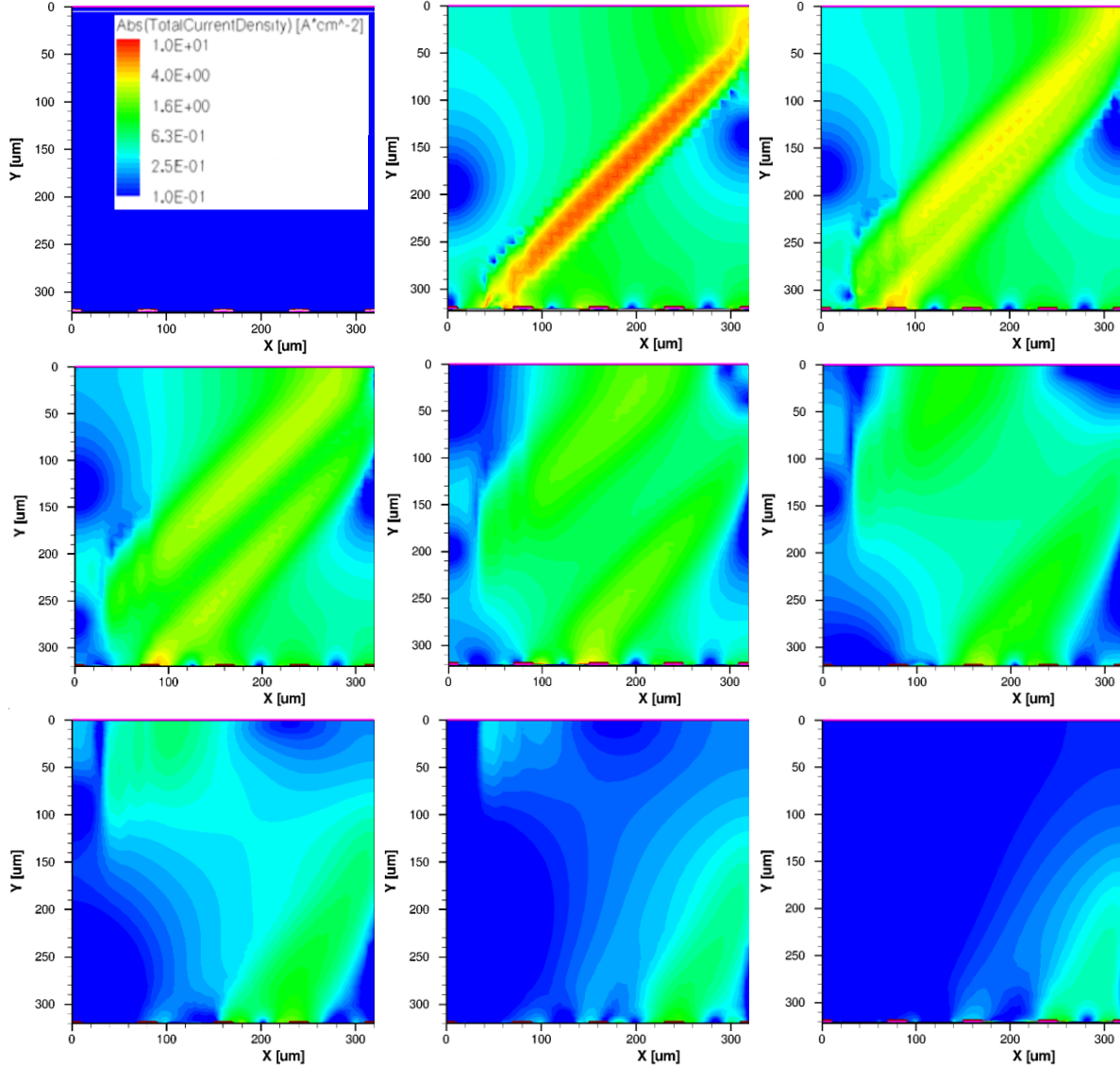
Figure 2.3: Simulation of a MIP traveling across a silicon strip sensor at $45°$ over time from 0.0, 1.1, 1.5, 2, 3, 4, 5, 6, 7 nanoseconds. The induced surface current dissipates and would be collected by the channels of the silicon module [10].

Pixel detector

The pixel detector contains the Barrel Pixels (BPIX) and the Forward Pixels (FPIX). Similar 2×8 silicon detector modules make up both the BPIX and FPIX systems.

In 2016, the phase I 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. After wirebonding, the gantry system encapsulated the wirebonds for protection from corrosion and magnetic field resonance. Purdue was one of the manufacturing sites alongside

University Nebraska-Lincoln.

Using LabVIEW, we developed a state machine to assemble and encapsulate these pixel modules. A pattern recognition and 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 assembly module—is shown in figure 2.3.1.
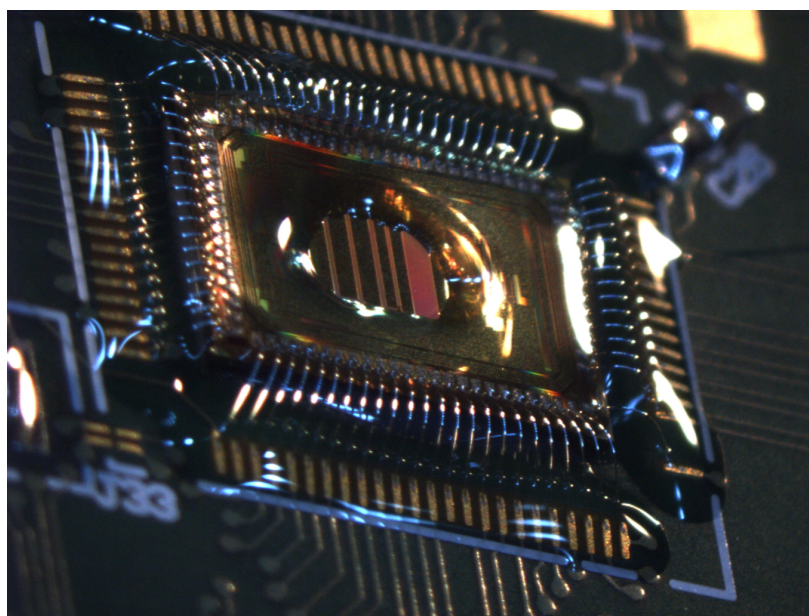
Figure 2.4: Encapsulated token bit manager of a FPIX module currently installed in CMS

In 2017, this system was installed in CMS, increasing the number of disks to three and the number of barrel layers to four. The design of CMS is such that the inner subdetector systems may be taken out of the solenoid and serviced. These forward disks, which are especially important for the reconstruction of boosted charged particles, are located in high regions of η and are overlapped for improved hermeticity.

Silicon tracker

The silicon tracker comprises larger silicon modules by area than the pixel system and is located further from the beam pipe. A representative layout of the silicon tracker and the pixel system can be found in figure 2.3.1 [9].
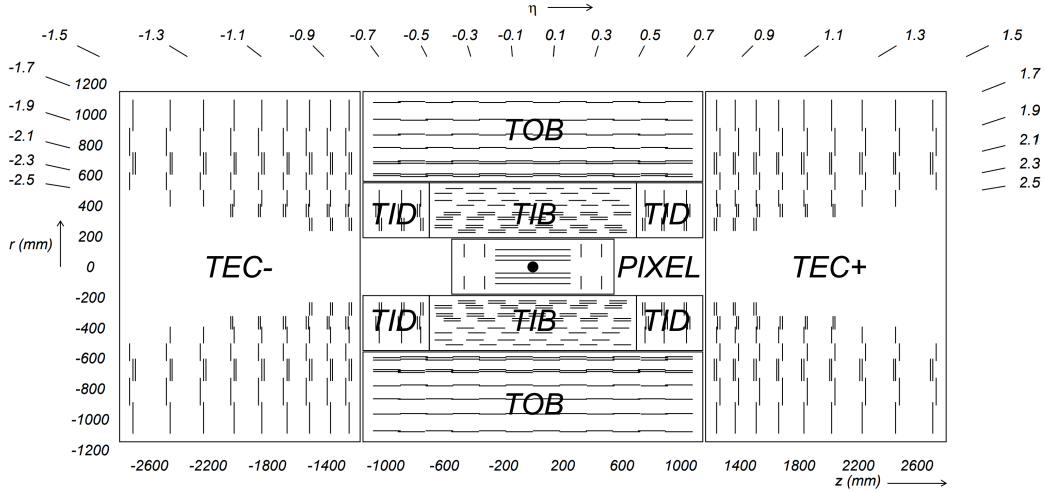
Figure 2.5:   The silicon tracker system, consisting of the inner pixel system (BPIX and FPIX),
The Tracker End Caps (TEC), Tracker Inner Detector(TID), Tracker Inner Barrel (TIB), and the
Tracker Outer Barrel (TOB) by position in r, z , and η [9]

2.3.2   ECAL

The principal components of the ECAL are the 76200 lead tungstate ($PbWO_4$) crystals, which
scintillate and absorb energy from incoming particles. These detector components are also separated
in barrel and endcap regions. A photomultiplier is attached to each the crystal to detect its light
signal. The relative energy resolution is

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.3\%)^2$$

The first, second, and third terms in equation 2.3.2 reflect the stochastic, noise, and constant terms
respectively as determined in a calibration run with a 440 nm blue laser [9, 10].

2.3.3   HCAL

The HCAL is the primary subdetector to identify "jets", which are collimated collections of hadrons.
Brass plates interwoven with plastic scintillators are used to induce particle showers. Light signals
from the scintillators are propagated through wavelength shifting fibers, and read out through an
optical decoding unit before ultimately landing at a hybrid photodiode. There are barrel (HB)
and endcap (HE) sections inside the solenoid. The hadronic outer (H0) and super forward detector
(HF) sit outside the solenoid. An overview of the HCAL system is shown in figure 2.6 [9, 10].
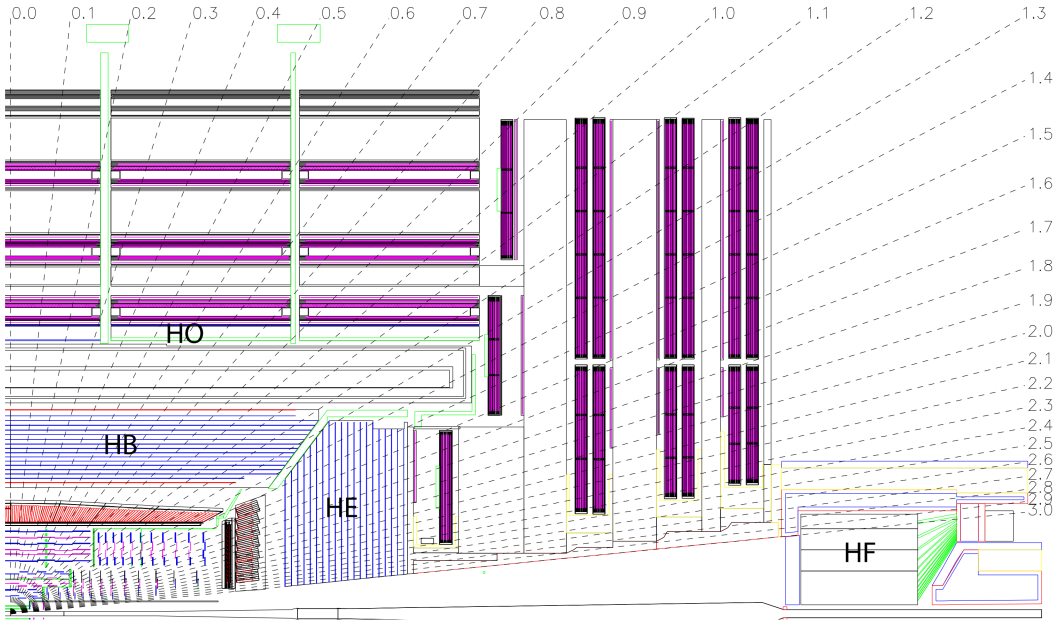
Figure 2.6:   Overview of the HCAL system from the z,η plane showing the hadron barrel (HB), encap (HE), outer (HO), and the forward (HF) subsystems [9]

2.3.4   Muon system

The muon system comprises drift tubes (DT) which cover a pseudorapidity region ($|\eta| < 1.2$) split into four stations interleaved in the flux return plates. In the higher $|\eta|$ endcap regions, cathode strip chambers (CSC), which provide fast response time, fine segmentation, and radiation resistance, are used. In both high and low regions resistive plate chambers (RPCs) are used [11]. For a visual representation of a cross section of the muon system, please refer to figure 2.3.4. Notably, the track of the muon is bent inside the solenoid by the Lorentz force, and then reverses after it exits.

2.4   Level 1 Trigger system and High Level Trigger

Particle collisions happen at a rate of 40 MHz, resulting in about 20 minimum bias events at each bunch crossing. The bandwidth that would be needed to record all these collisions is prohibitively high [7, 13]. The Level 1 trigger (L1) and High Level Trigger (HLT) work to reduce these rates by selecting events of interest.

The L1 trigger system, which comprises many subsystems, can process data at the beam collision rate. Algorithms are in place that take input from the calorimeters, the muon systems, and other
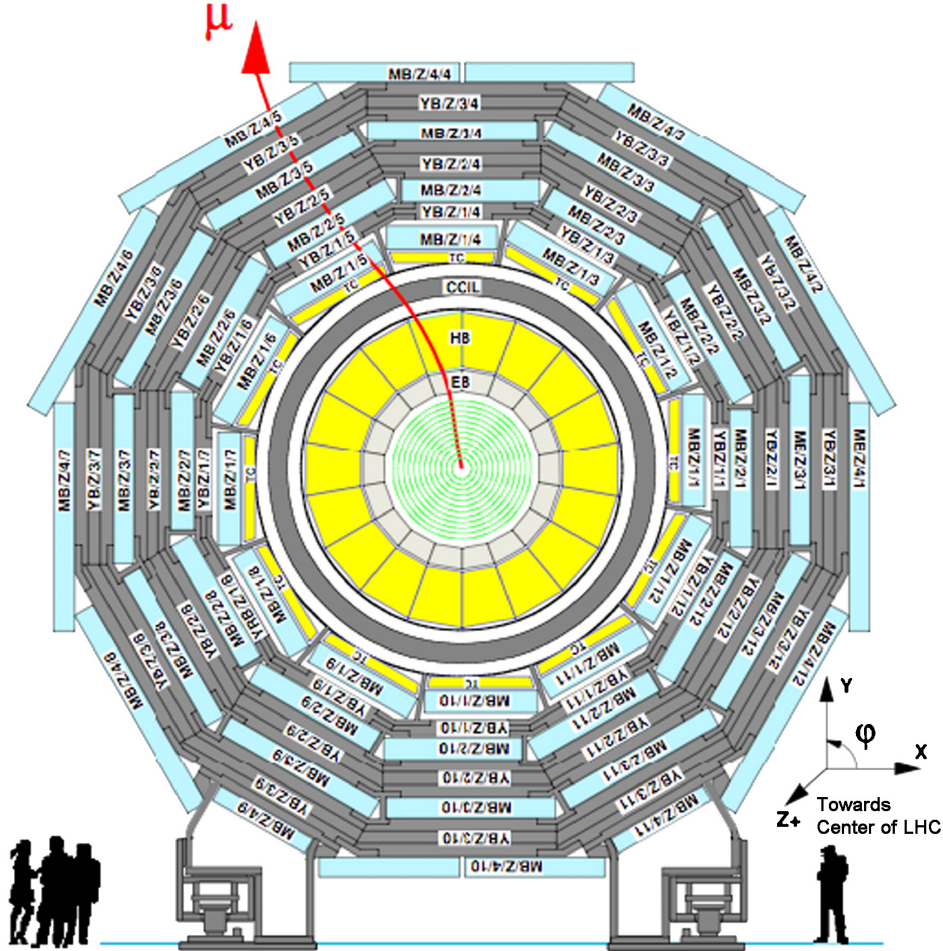
Figure 2.7: Muon system involving multiple subdetector systems: tracker, solenoid, and the muon gas chambers around the iron yoke (grey) [12].

detectors in the form of "trigger primitives", and use pattern recognition, along with fast summing techniques, to trigger on events. Many of these algorithms are run on Field Programmable Gate Arrays (FPGAs). After 144 beam crossings, the Global Trigger (GT) initiates readout for events of interest at the front end electronics. The L1 system also outputs physics objects to seed the reconstruction algorithms used by the HLT [13]. The HLT maximum input rate is 100 kHz and the output rate is on the order of kHz. It is constrained by the processing power available, the data recording and transfer rate of Tier 0 (T0), and the prompt reconstruction algorithms. Late in Run II, "Scouting" and "Parking" data were used to make more efficient use of the available bandwidth. Scouting reduces the event size by saving only objects reconstructed by the HLT. Parking reduces the immediate load on the T0 system by postponing prompt reconstruction to when CMS is not running [14].

2.5   Particle flow algorithm

The event data model requires the association of higher level physics objects—like leptons—with energy deposits and tracks in the detector. The particle flow algorithm has the goal of associating these primary detector signatures with these particles so a direct comparison to Monte Carlo (MC) simulation can be done. The list of particle objects includes jets, missing transverse energy, taus, charged-leptons, photons, and bottom quark jets among others. To outline the algorithm: charged particle tracks reconstructed in the tracker, energy clusters from the ECAL, HCAL, preshower detector (ES), and forward calorimeter (HF) are topologically linked into blocks. The linking is done through many associations of energy deposits and tracks in ϕ, η space. These blocks are then interpreted as particles. Further details can be found in reference [15].

2.6   Computational Infrastructure

Over 200 peta-bytes of information have been gathered in Run II. A schematic overview of the computing infrastructure can be found in figure 2.6 [16].
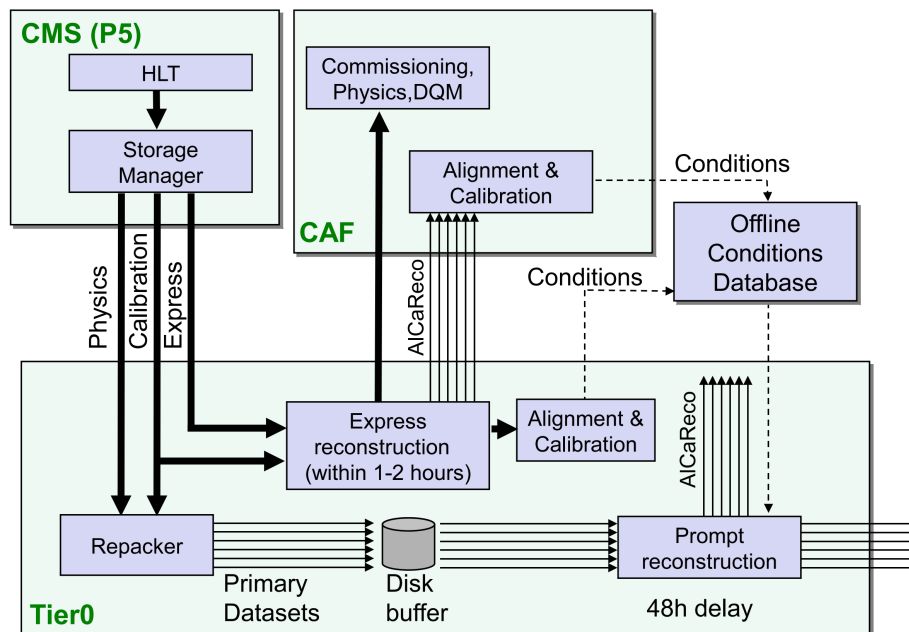
Figure 2.8: Schematic overview of T0 computing: data can be available from different sections allowing for data quality monitoring and also storage to several databases [16].

The T0 facility, which comprises 32,000 24-core processors, is where higher level reconstruction

of physics objects is done. The data is stored and processed at many different sites, organized by tiers. There is one Tier-0, seven Tier-1, about one-hundred fifty Tier-2, and numerous Tier-3 centers. The sum of these tiers is the "grid" or the Worldwide LHC Computing Grid (WLCG), which in total combines 900 000 CPUs from over 170 sites in 42 countries. Tools like `XrootD` and `Rucio` allow physicists from around the world to access centrally supported data. An abundance of up to date information can be found `https://home.cern/science/computing/grid`.

Chapter 3

Luminosity

3.1 Luminosity at the LHC

Luminosity sets the scale for the number of events recorded at the LHC. It is how bright the beam is and 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. The number of expected events for any given process is the luminosity $\mathcal{L}$ times the cross section σ

$$N_{\text{event}} = \mathcal{L}\sigma_{\text{event}}, \tag{3.1}$$

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{LHC}} \gamma_r}{4\pi\epsilon_n \beta*} \left(1/\sqrt{1 + (\frac{\theta_c \sigma_z}{2\sigma*})^2} \right). \tag{3.2}$$

N_b is the number of particles in the bunch crossing, n_b the number of bunches, f_{LHC} the revolution frequency of the LHC, γ_r the relativistic factor, ϵ_n the normalized beam emmittance, $\beta*$ the beta function at the collision point (related to the crossing angle), θ_c is the full crossing angle, σ_z the RMS bunch length and σ^* the transverse RMS beam size at the interaction point.

3.2 Luminometers

Several subsystems are used to measure luminosity at CMS. Particularly, the pixel luminosity telescope (PLT), HF with summed transverse energy (ET) and occupancy (OC), fast beam condition monitor (BCM1F), and tracker/pixel based luminosity detectors to name several. In this section,

tracker based luminosity will be the focus. The pixel system is integral in the reconstruction of events for physics analyses and in measuring the luminosity.

There was calibration of the new pixel detector in early 2017, from the BPIX and FPIX upgrades mentioned in section 2.3.1. The Lumi-POG—luminosity physics object group—commisioned luminosity measurement using the clusters from the new pixel detector in an automated workflow. We measure the luminosity by counting pixel clusters in a low channel occupancy setting and scaling it by a visual cross section—the measured cross section for the instrument. Using the relation in equation 3.3, the instantaneous luminosity can be obtained once the number of clusters and the visible cross section σ_{cluster} are measured.

$$\langle N_{\text{cluster}} \rangle \equiv \frac{\sigma_{\text{cluster}}}{f_{\text{LHC}}} \mathcal{L}_{\text{SBIL}} \tag{3.3}$$

$\mathcal{L}_{\text{SBIL}}$ 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. σ_{cluster} is the cross section that is measured in a separate analysis involving Van-de-Meer scans (beam dynamic scans). More details can be found here [17].

3.2.1 Tracker luminosity

For the pixel luminosity, a two component correction is applied on the fly to correct for self-radiative effects on the pixel modules under particle fluence and for inefficiencies. One detail that is important in estimating the luminosity from the pixel detector is ensuring that the data has consistent performance. Several times in a year, the Lumi-POG and Beam Radiation Instrumentation Luminosity (BRIL) groups analyze the performance of each subdetector 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 was vetted by looking at relative module performance over the runs of data-taking for those years. If the modules didn't have consistent performance, they were removed from the final result.

Even though up to half of the modules were vetoed after this procedure, there was plenty of statistics due to high numbers of pixel clusters per event. This module veto decision was made by taking the total clusters in each module and scaling them so that the overall total clusters are one. Then, on a per-module basis, the performance relative to the total clusters was compared. This

helped the analyzer look at consistent performance and manage a list of passing modules used in a final luminosity measurement using the pixel detector.

For Run III data-taking, integration of the cluster counting procedure was included at the HLT. A data compression of 10^3 was made by taking the low level data from the silicon pixels and storing them in a simple data container, which saved a peta-byte of data. Luminosity measurements using these data containers and methods are being investigated as more CMS-physicists are interested in using the central tracking system for luminosity measurements.

Chapter 4

Lepton Identification and Object Selection

4.1 Lepton identification

The following three subsections briefly describe how certain subsystems of the CMS detector work together to identify muons, electrons, and tau leptons.

4.1.1 Muon identification systems

As CMS implies in its name, muons are certainly a focal point in particle detection. Looking at muons that come from the interaction vertex—prompt muons—the tracker plays an important role in identifying charged particle tracks such as muons. The tracker system works in conjunction with the gas chambers to reconstruct muons, and the solenoid bends the muon's tracks allowing the momentum to be measured for an accurate mass resolution. By design, muons should be the only particle that reaches the gas chambers, making for great muon purity. During reconstruction, muons are identified and are ultimately divided into four working points (very loose, loose, medium, tight). These points are defined based on their efficiencies and depend on the χ^2 of the track and momentum of the candidate muon [18, 19].

4.1.2 Electron identification systems

The main subdetector involved with electron identification is the ECAL. To identify electrons, a cluster of energy in the ECAL is associated with a track that is constructed in the silicon detector system. The tracks are identified in the typical fashion using the Kalman Filter tracking technique to pick good quality tracks, and then the tracks are refitted using a Gaussian Sum Filter. These tracks are then associated with an ECAL super cluster—grouped energy deposits—by requiring matching in η, ϕ space

$$|\Delta\eta| = |\eta_{\text{SC}} - \eta_{\text{in}}^{\text{extrap}}| < 0.02 \tag{4.1}$$

$$|\Delta\phi| = |\phi_{\text{SC}} - \phi_{\text{in}}^{\text{extrap}}| < 0.15 \tag{4.2}$$

This method has an overall efficiency of about 93% [20].

4.1.3 Tau identification systems

Tau leptons decay both hadronically and leptonically as shown in figure 4.1.3:
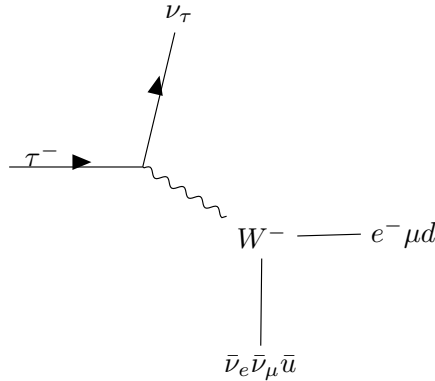
Figure 4.1: diagram depicting the possible decays of the tau lepton: 65% to a hadronic tau, 18% to an electron, and 17% to a muon (each with associated lepton neutrino).

It is the heaviest lepton, so heavy it can decay to intermediate mesons such as the ρ, a, and π mesons. Ergo, when it comes to tau identification, many algorithms are needed to properly identify them using information across the CMS detector. Decay modes are listed in table 4.1 below.

The Hadron Plus Strips (HPS) combines the use of the tracker system and the ECAL for hadronic tau identification [22]. To identify τ_h candidates, the HPS algorithm is used to identify the major modes of the hadronic tau decay [22]. Typically, events with hadronic prongs—charged

Table 4.1: Possible hadronic tau decays. h doesn't indicate a Higgs particle, but a hadronic prong instead [21]

Decay Modes	Resonance	$\mathcal{B}(\%)$
Leptonic Decay		35.2
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
Hadronic Decay		64.8
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other		3.3

hadrons—are considered in combination with a number of neutral pions and missing transverse energy from the neutrinos. Neutral pions almost always decay to photons, so a hadronic tau identification algorithm should combine the identification of charged hadrons and neutral pions. The τ_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 [23, 24]. In addition to the HPS algorithm, a Deep Neural Network (DNN) was constructed to further aid in identification by discriminating between genuine tau leptons and those that originate from quarks or gluon jets, electrons, or muons. In the DNN, the tau four-momentum and charge, the number of charged and neutral particles constituents, the isolation variables, the compatibility of the leading tau track with the primary vertex, the properties of a secondary vertex in case of a multiprong tau decay, observables related to the η and ϕ distributions of energy reconstructed in the ECAL strips, observables related to misidentified taus as electrons, and the estimated number of primary vertices density in the event are all used. In total, 47 high-level input variables are incorporated [25]. In practice, the DNN has discriminators against muons, electrons, and jets that fake genuine taus and has efficiencies that go from 40% to 90%, in a 10% granularity of the discriminating variable. The medium working point is used for each of these discriminators.

4.2 Data and simulation

For this analysis, muons are paramount, so there must be a certain number of muons triggered for the event to be selected. Two final states contain electrons, so datasets containing electrons are

also used. The single muon, double muon, and electron plus photon datasets are used depending on the year. These datasets contain the triggers that are the most important for object selection. Single muon triggers with single, isolated muons at various thresholds are implemented, along with double muon triggers and triple muon triggers. More information on triggers and selection is given in the event selection section 5.2.1.

The simulation typically used to compare with data is MadGraph5@NLO with PYTHIA 8 for hadronization [26]. These CMS centrally-generated samples are then digitized using GEANT4 [27] to the same format as real data events collected and processed at HLT. This raw data is then reconstructed to physics objects—such as tracks and higher level objects like leptons. A direct comparison between data and simulation can be made after calibrating simulation in control regions.

Data taken from CMS during the entire Run II period was examined, corresponding to 137 fb^{-1} of integrated luminosity. An exhaustive list of data and simulation Monte Carlo (MC) can be found in 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 GeV in intervals of 5 GeV with the parent Higgs produced through gluon fusion. These masses are sufficient for the parametric modeling, which is used in the fit model 8.3 to obtain a 1 GeV peak resolution. Signal samples were produced centrally by CMS for 2016, but privately produced for 2017 and 2018. The scripts and conditions used are located here:

`https://github.com/samhiggie/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/ .

4.3   Physics object selection

Baseline selections for objects are recommended by the various Physics Object Groups (POGs). The process of making selections on variables is known as "making cuts". Ultimately, these selections take the form of cuts for leptons based primary on kinematic variables such as the momentum. Selections are made to identify muons, electrons, and tau leptons. These three leptons form the objects under selection and comprise the final states in the analysis. All events containing at least

two leptons that pass the trigger requirements are selected depending on the momentum of the leptons in the event. Special triggers are used for the different final states based on the number of muons in the event. Identifying particles in object selection is critical, particularly differentiating between lepton candidates that come from the interaction vertex (prompt) and those that appear from decays down the line (nonprompt). Relative isolation is typically defined in order to ensure there is no overlap between candidate leptons, to make sure that each lepton is not associated with other physics objects like jets. More details on this variable and its usage in the particle flow algorithm can be found here [28].

$$I^\ell \equiv \frac{\sum_{\text{charged}} p_{\text{T}} + \max\left(0, \sum_{\text{neutral}} p_{\text{T}} - \frac{1}{2} \sum_{\text{charged, PU}} p_{\text{T}}\right)}{p_{\text{T}}^\ell}.$$ (4.3)

$\sum_{\text{charged}} p_{\text{T}}$ 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 $\sum_{\text{neutral}} p_{\text{T}}$ is a similar quantity for neutral particles. Track association isn't possible with neutral particles and thus not with primary vertex information; therefore, to take pileup into consideration, an estimate of the transverse momentum from the pile up contribution is subtracted (PU p_{T}).

As mentioned in chapter 2, the muon identification system uses the tracker to identify charged tracks and the muon chambers to identify the particles later in their trajectories after they exit the solenoid. Typically, "good" muons are those that are both associated with a track and their subsequent identification in the drift tubes, CSCs, or RPCs. The average muon lifetime is 2.2 μs so they travel quite far from the interaction point. The physics object group's recommendations are followed, which select muons with $p_{\text{T}} > 15\,\text{GeV}$ and $|\eta| < 2.4$ in addition to selecting only "good" muons.

Electrons originating from the tau decay are reconstructed by track association along with energy deposition in the ECAL. Events are vetoed for candidate electrons that also show a substantial energy deposition in the HCAL for better sample purity. The hits and track quality from two separate algorithms, along with the geometrical and energy matching from the ECAL are used in a multivariate analysis (MVA) technique to select good electrons for analysis [20].

Tau leptons must pass the HPS algorithm and the medium DNN working point to be selected. Both techniques are detailed above in section 4.1.3.

All leptons also have a momentum threshold and isolation requirements to place them in a kinematic region where good agreement between data and MC is expected in control regions. A table listing these selections are found in table 5.2.

Table 4.2: Baseline cuts and identification for lepton selection δZ and δxy are distances from the interaction vertex

lepton	baseline cuts
Muon	$\delta Z < 0.2$, $\delta xy < 0.045$, $p_{\mathrm{T}} > 5.0 GeV$, Iso. $<= 0.2$, $\eta \geq 2.4$
Electron	$\delta Z < 0.2$, $\delta xy < 0.045$, $p_{\mathrm{T}} > 7.0 GeV$, Iso. $<= 0.15$, $\eta \geq 2.5$
Tau	$\delta Z < 0.2$, $\delta xy < 0.045$, $p_{\mathrm{T}} > 18.5 GeV$, Iso. $<= 0.2$, $\eta \geq 2.3$, Med. DNN

4.4 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 POGs, standard techniques are applied to ensure proper simulation. Corrections to energy scales for the leptons in the analysis are most critical. 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 affect the results are considered in the overall error in the statistical inference model.

4.4.1 Muon energy scale

Corrections to the muon's energy scale are computed for the muons that pass the selection for this analysis. Medium muons with track based isolation that pass any of the isolated single muon triggers at 22 GeV, 24 GeV, and 27GeV are then rescaled for pileup (explained in 4.4.6), efficiency, di-lepton p_T, and electroweak re-weighting based on accurate gauge boson measurements. After selection, the scale factors for energy corrections are measured and parametrized in η and ϕ in multiplicative and additive corrections [29].

$$\rho^{\mathrm{cor}} = \kappa(\eta, \phi)\rho + Q\lambda(\eta, \phi). \tag{4.4}$$

The correction coefficients κ, and λ are measured in a tag and probe method and Q is charge, ρ to energy.

In practice, these are just scale factors applied to the energy scale in certain η, ϕ regions.

Correction (%)	
η region	scale factor
$0 - 1.2$	0.4
$1.2 - 2.1$	0.9
> 2.1	2.7

Table 4.3: Measured μ energy scale correction for genuine μ(s) across all years.

4.4.2   Electron energy scale

Electron energy scale and resolution requires corrections to be applied to MC in order to match data [30]. 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 η of the electron shown in table 4.4.2.

Correction (%)	
η region	scale factor
$0 - 1.2$	1.0
$1.2 - 2.1$	1.0
> 2.1	2.0

Table 4.4: Measured e energy scale correction for genuine e(s) across all years.

4.4.3 τ energy scale

There is a nominal energy shift based on the type of tau decay. Due to the electroweak interactions, τ leptons decay hadronically and leptonically 4.1. When the tau decays hadronically, many different intermediate mesons are produced. Each type of meson decay has a different signature, particularly when they hadronize and deposit their energy within HCAL. The Tau-POG has measured the central value systematic deviation as a form of scalar factor that is applied for an accurate result of measuring the tau's energy. This is split by the prongs (charged hadrons) and π^0 s.

Correction (%)			
Decay mode	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

Table 4.5: Measured τ_h energy scale correction for genuine τ_h(s) across all years.

The deviation in the model is measured by taking the difference between 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 + π_0 decay mode as a function of tau energy; all other tau decay modes are also measured.
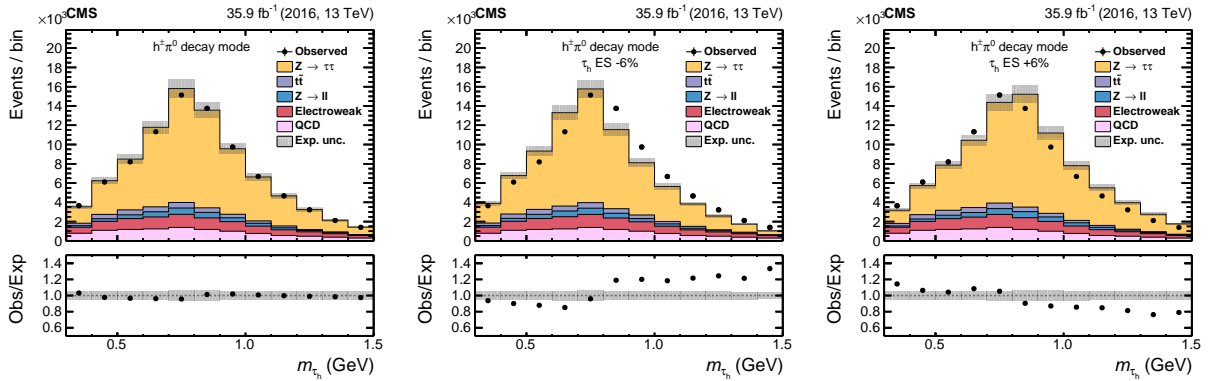
Figure 4.2: Tau mass distributions considering the nominal τ_h energy scale in simulation (left), or the τ_h energy scale shifted by -6 (middle) or +6% (right), in the $\mu\tau_h$ final state, for the 1 prong + π^0 decay mode.

4.4.4 τ_h identification efficiency

The τ_h identification efficiency can be different in Data and MC [31]. To correct for this difference, measurements are made using genuine Z+jets production (Drell-Yan) to two τ leptons, where one decays leptonically and the other hadronically. The invariant mass of the system is used 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 event selection regions shown in section 5 with an emphasis of simulation containing real taus. This measurement is done by the Tau POG, and the scale factors are provided to CMS.

While used in the primary event and the parameter of interest in the fit, the efficiency's error

is not considered in the overall systematic error as they are expected to have very little impact on fit and limits based on the 2016 result.

4.4.5 $e \rightarrow \tau_h$ and $\mu \rightarrow \tau_h$ misidentification rate

The efficiency of the discriminators against electrons or muons misidentified as τ_h candidates can also be different between simulation and data. These data/MC scale factors are binned by barrel/endcap region of the measured $\eta(\tau_h)$, and by τ_h decay mode. Scale factors are measured to correct this difference and are applied to electrons or muons faking tau leptons in MC. Full information on misidentification measurements and application in analyses can be found in reference [31].

The misidentification scale factors are derived by pass and fail regions. The regions are set up by selecting events where a reconstructed τ_h passes the DNN working point and also fails the DNN discriminant against muons or electrons. The regions considered are QCD multijet, W+jets, and Z+jets—similar to the regions used in chapter 6. QCD multijet is estimated from a same sign lepton region within data. W+Jets normalization is carefully selected from a region with high transverse mass. The visible mass distributions of the events in these regions are fit, and the overall signal yield remains constant in the pass and fail regions. The expected impact on the systematic error from these anti-lepton discriminators is expected to be very small so they are not included in the uncertainty model either.

4.4.6 Pileup re-weighting

Re-weighting is done to rescale the events for the effective number of primary interactions (pileup) during collisions. A minimum bias dataset is used, depending on the year, to calculate the expected number of primary interactions for MC re-weighting. During Run II, pileup in a crossing or underlying event was around 50 and in Run III it will exceed 200.

4.4.7 Electron and muon identification efficiency

Scale factors derived within the SM Higgs decay to tau tau group are applied for muons [32], and the EGamma POG scale factors are applied for electrons [33] to account for the identification efficiency. The expected impact on the systematic error from these identification efficiencies is expected to be very small so they are not included in the uncertainty model either.

4.4.8   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 yield of the samples. The event weights for simulation are scaled to the expected yields for each sample.

To correct for differences between leading order and next to leading order cross sections, scale factors called K-factors are used. They are applied to W+Jets and Drell-Yan samples. For Drell-Yan a factor of 1.1637 is used and for W+Jets a factor of 1.221 is used.

4.4.9   Low momentum 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 encaps using the tag and probe technique in the 2016 analysis. These factors are used to correct for low p_T muon selection, in addition to the Muon POG's recommendation, to support correct simulation of data for all years. The scale factors are listed in the table 4.6

Table 4.6: Scale factors to correct for low momentum muon selection being less than the trigger threshold.

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

4.4.10   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 change and provide a cross check, distributions for the τ, μ, and e energy scale shifts in the parameter of interest (the mass of the di-muon system from the leading p_T pseudoscalar a particle) are plotted. An example of this type of systematic for $\mu\mu\tau\tau$ channel is shown in figure 4.4.10.

In the end, the uncertainty considered in the fit model would then be the percent yield up and down as a flat error (log-normal) that affects the normalization of the template. Because the model
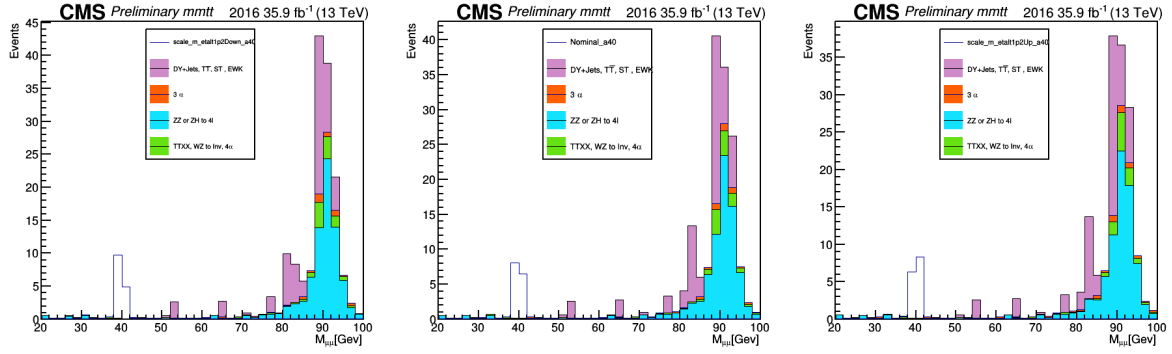
Figure 4.3: Systematic shifts in the dimuon mass spectrum for 2016 $\mu\mu\tau\tau$ for the muon energy scale shift down (left), nominal (mid), and up (right), no data is shown on this plot as it directly reflects the signal region without the extraction cuts.

is so statistically limited, the log-normal is sufficient in capturing the changes in the distributions over the fit range 8.3.

Chapter 5

Event Selection

5.1   Framing an analysis

In order to conduct a concrete hypothesis test, two perspectives are taken. The null hypothesis is the SM, which will comprise all possible events that are categorized as *background*. The alternative hypothesis is the SM with the addition of the *signal*, ie., $H \to aa \to \mu\mu\tau\tau$.

5.2   Defining signal and control regions

To optimize the analysis, regions of the data and simulation are cut away in order to increase the number of signal events relative to background events. In order to not bias the result, regions are setup to investigate the agreement of simulation with 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   Triggers for event selection

The trigger requirements are inclusive, selecting events that pass single, double, and triple muon triggers. Events that are triggered by the single muon trigger criteria contain muons that are isolated with either 22, 24, and 27 GeV energies. Double muon triggers have a 17 GeV threshold for the leading muon and 8 GeV for the subleading muon. Triple muon triggers are used for the channels that have three muons in the final state and have a descending threshold of 12, 10, and

5 GeV respectively. In addition, to properly select objects that coincide with the trigger, triggers are matched to their corresponding objects. The lepton is matched to the seed and filter bit that is generated at the L1 system. Trigger filter bit matching ensures that the objects and events are correctly triggered.

5.2.2 Optimizing lepton pair selection

A simple selection algorithm was used to identify good lepton pairs that come from the pseudoscalar a. Standard working point cuts are made, and two oppositely charged, isolated muons with the largest scalar summed p_{T} are chosen to form the first decay products of the a. Two opposite charged τ leptons with the largest scalar summed p_{T} are chosen for the second a. This approach increased the signal acceptance compared to choosing mass window cuts to form the a pairs. The pair matching efficiency study done with the preliminary dataset from 2016 is listed in table 5.1. The dip in efficiency may be explained by the boosted or resolved a particles depending on their

Table 5.1: Lepton pair matching efficiency

a - Mass	15	20	25	30	35	40	45	50	55	60
Efficiency	0.87	0.82	0.79	0.79	0.79	0.80	0.80	0.83	0.85	0.87

mass and decay products. If the a mass is low then it is more relativistic, resulting in collimated leptons. If the a has a higher mass, then it is produced closer to rest and the leptons are identified at a large angle of separation. It is likely that the particle flow algorithm has a more difficult time in identifying decay products of the a particles in between the mass extremes.

5.2.3 Optimizing final state event selection

After picking the leading prompt muons from the a decay, the next step is to identify the other a decay by using various leptons in the final state. The final state comprises four leptons: two muons coming from the leading a and two tau leptons coming from the subleading a. These τ leptons can decay leptonically or hadronically, and this analysis counts events from both types of decay. Therefore, event selection is driven to find two prompt muons and all decay products of the tau leptons. Four final states are used: $\mu\mu e\mu$, $\mu\mu e\tau$, $\mu\mu\mu\tau$, and $\mu\mu\tau\tau$. The states $\mu\mu\mu\mu$ or $\mu\mu ee$ are not included, as the expected number of events would be extremely low based on the double leptonic

tau decay. For notation, when the final state is listed with a tau, such as $\mu\mu\mu\tau$, the τ is presumed to decay hadronically. The third muon in this context would be coming from a leptonically decaying τ. Additionally, due to convention in plotting, the μ(s) are often marked as m(s) and τ(s) as t(s) like in plot 6.3. In addition to the kinematic requirements listed in 4.3, several cuts are made to select final state events. The following list contains cuts common to all channels:

- leading muons must have opposite charge coming from the a

- tau decay products must have opposite charge coming from the other a

- no b-quark identified jets

- signal extraction cuts (not shown in data MC control plots, but used in statistical test)

    - invariant mass of the 4 lepton system cuts based on signal to background ratio and overall event yeild

        * $\mu\mu\mu\tau$ all years at $M_{4l} < 120 GeV$

        * $\mu\mu e\tau$ all years at $M_{4l} < 120 GeV$

        * $\mu\mu e\mu$ all years at $M_{4l} < 110 GeV$

        * $\mu\mu\tau\tau$ 2017 and 2018 at $M_{4l} < 130 GeV$ and 2016 at $M_{4l} < 135 GeV$

    - $M_{\mu\mu} > M_{\tau\tau}$ (to account for energy loss from neutrinos).

Table 5.2:   additional final state selection cuts

finalstate	cuts
$\mu\mu e\mu$	Iso. μ from $\tau <= 0.2$, Iso. e from $\tau <= 0.15$
$\mu\mu e\tau$	τ_h DNN against μ and e
$\mu\mu\mu\tau$	τ_h DNN against μ and e,Iso. μ from $\tau <= 0.15$
$\mu\mu\tau\tau$	τ_h DNN against μ and e

Chapter 6

Background Estimation

Due to the stringent cuts in the tight signal region and absense of resonance, MC simulation suffers from low statistics for some background processes. Therefore, in addition to the MC simulation, a data driven method is used to estimate a significant portion of background that is not reliably estimated using MC.

The hadronic τ decays produce jets; therefore, jets coming from other processeses effectively fake the hadronic τ signature. This is a non-trival fake rate that needs to be measured and accounted for in this analysis.

In order to conduct the data driven method, a proportion is made to extract the jet faking tau background. Generally, this proportion is constructed using a region orthogonal to the statistical hypothesis test. For example, one requirement could be a charge sign inversion on a lepton pair used in the final state. In this orthogonal region and in the tight signal region, tight and loose identification criteria are made to extrapolate the scale factor that estimates the tau's fake rate. The tight identification should be excellent at selecting genuine tau leptons and the loose identification more inclusive to all tau leptons, even those that are fake. Therefore, using the orthogonal reigons—same sign and opposite sign—along with the loose signal region, one can extrapolate the number of events in the tight signal region. Due to the four regions, the method is also referred to as the "ABCD" method.

6.1 Brief outline of the fake rate method

The fake rate function in the same sign (SS) region is *known*. Events passing loose identification—that includes tight indentification—in the opposite sign (OS) region is *known*. Events passing in the tight signal region is *unknown*. Assuming that the loose and tight identification is not dependent on the sign of the leptons. Then one can make the equivalence statement:

$$\frac{\text{Events}_{\text{SS Tight}}}{\text{Events}_{\text{SS Loose}}} \doteq \frac{\text{Events}_{\text{OS Tight}}}{\text{Events}_{\text{OS Loose}}}. \tag{6.1}$$

To make the expression more precise, the fake rate function is typically parametrized in lepton candidate transverse momentum. Also, prompt MC is subtracted from data, which is motivated by estimating the true jet faking tau background (non-prompt taus). If tau leptons in MC are identified as prompt then it is unlikely that the tau is a jet, so they are removed:

$$f(p_T) = \frac{\text{Data Events}_{\text{S.S. Tight}} - \text{Prompt MC Background}_{\text{S.S. Tight}}}{\text{Data Events}_{\text{S.S. Loose}} - \text{Prompt MC Background}_{\text{S.S. Loose}}}. \tag{6.2}$$

After the measurement is made for each tau candidate, the fake rate is applied as an event weight—$w(f(p_T))$ a transfer function explained 6.4—to the opposite sign loose region in order to extrapolate to the tight signal region. So isolating the events in the tight signal region and flipping the relation in equation 6.1, one obtains the result:

$$\text{Events}_{\text{OS Tight}} = w(f(p_T)) \cdot \text{Events}_{\text{OS Loose}}. \tag{6.3}$$

6.2 Measurement of the fake rate

To measure the fake rate, multiple categories are considered and motivated through the processes which produce jets. As outlined in the SM Higgs decays to tau leptons analysis and its supporting document on fake rate measurements, several regions are used to determine the fake rate [32]. The separate jet "enriched" background processes are used for each final state

- QCD multijet targeting the majority of jet$\rightarrow \tau_h$ fake events in the $\mu\mu\tau\tau$ final state,

- W+jets targeting jets mostly in the $\mu\mu e\tau$ and $\mu\mu\mu\tau$ final states,

- $t\bar{t}$ events targeting fully-hadronic or semi-leptonic decays.

The fake rates are then measured as a function of p_T of the object and for final states involving hadronic tau leptons, and are further split into subcategories depending on the decay mode. W+jets with no jets and one jet, QCD multi-jet with no jets, one jet, and more jets, and $t\bar{t}$ inclusive jets make up the total number of background categories that are measured. At high hadronic tau p_T (greater than $100\,\text{GeV}$), negative fake rates are possible because of low statistics and the linear fit model extrapolation, so if the candidate tau has a p_T of greater than $100\,\text{GeV}$ the rate at $100\,\text{GeV}$ is applied. In order to combine the fake rates from these "enriched" background processes and use it in an ABCD approach, the fraction of events for each of the background process are combined in an overall fake rate that is still parametrized by p_T and category.

Therefore the following steps are done for each final state to measure the fake rate:

1. Determine fake rate scale factor parametrized in candidate lepton p_T in the QCD, W+jets, and $t\bar{t}$ regions

2. Make corrections based on the other lepton in the channel for closure

3. Make corrections based on the differences between the first step and the signal region

4. Determine the fraction of QCD, W+jets, and $t\bar{t}$ events in the signal region.

To help in the understanding of the measurement regions, a table listing the enriched background and targeted final state along with the cuts and the anti-isolation requirement (the non-orthogonal condition in the ABCD method) will be presented. As indicated in the tables below, $\mu\tau$ and $e\tau$ measured states share the same categories. For the $\tau\tau$ state, only the QCD "enriched" background category is considered. For the $\mu\mu e\mu$ final state in the application, the fake rate measurements from the $\mu\tau$ and $e\tau$ measured states are used for the corresponding lepton (μ or e).

Fake factor measurements for the $\mu\tau$ measured state in the QCD region for 2017 is included in figure 6.1. Only plots pertaining to the $e\tau$ and $\mu\tau$ states were created, but all states measured. The rest of the measurements are included in Appendix B. For a more detailed description of the data driven background method along with the measurements for the closure and extra correction terms regard reference [34].

Table 6.1: Jet "enriched" background categories with cuts for each measured state, $\mu\tau$ and $e\tau$ have the same so they will be combined with the assumption that l demarcates the muon or electron. Baseline selection cuts for events are made by default as listed in section 5 without the signal extraction cuts.

background	measured state	cuts	anti-isolation
QCD	$\mu\tau/e\tau$	SS leptons Isolation $l \in (0.02, 0.15)$	τ VVVLoose DNN but fails Med. DNN
	$\tau\tau$	SS leptons subleading τ pass Med. DNN and leading VVVLoose DNN	leading τ fails Med DNN
W+Jets	$\mu\tau/e\tau$	SS leptons m_T between l and $p_T^{\text{miss}} > 70$ GeV	τ VVVLoose DNN but fails Med. DNN
$t\bar{t}$	$\mu\tau/e\tau$	SS leptons number of b-tag jets ≥ 1	τ VVVLoose DNN but fails Med. DNN
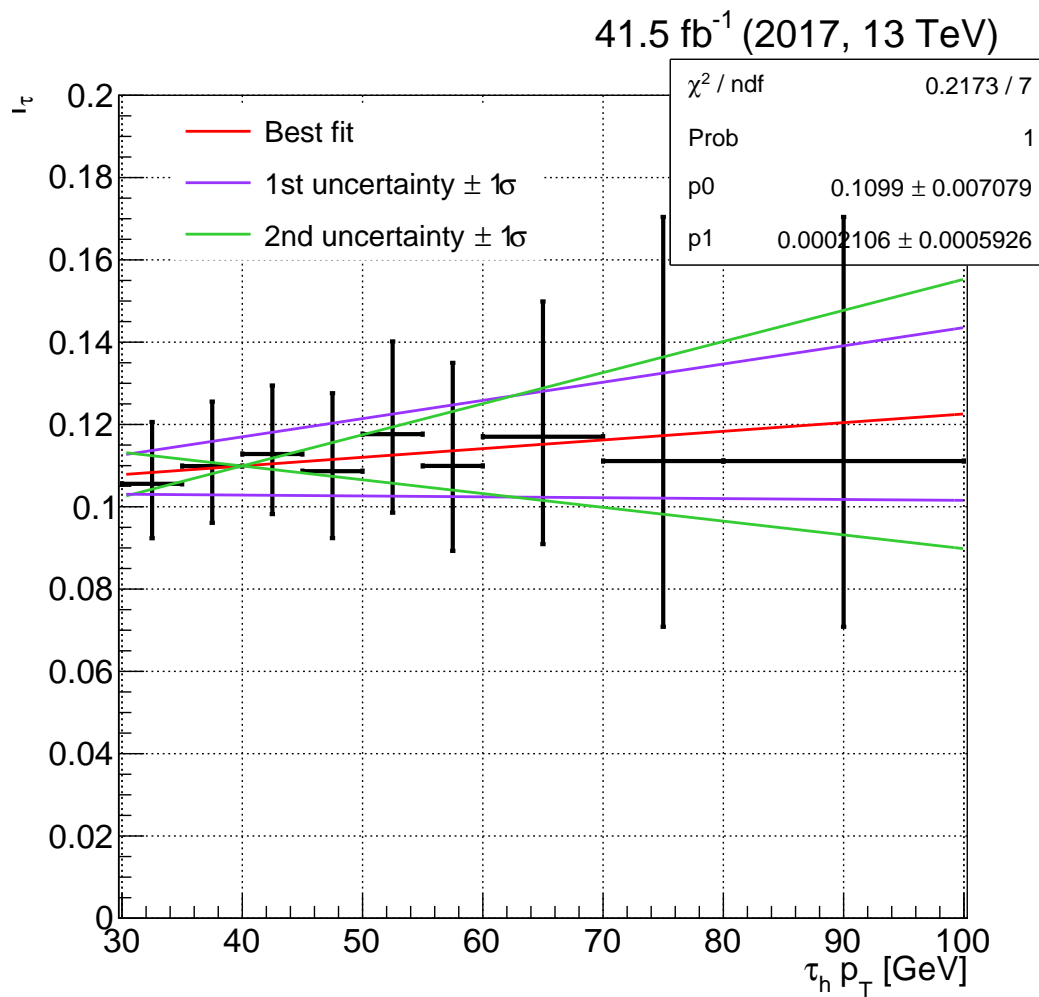
Figure 6.1: Fake factors determined in the QCD multijet determination region with 0 jets in the $\mu\tau$ measured state in 2017. They are fitted with linear functions as a function of the τ_h p_T. The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.

6.3 Application of the fake rate method

After the jet faking tau rate is measured, it is then applied to events that are identified as loose and not tight. Since the final state involves two tau leptons, this procedure is applied to each tau lepton in the final state, thus requiring application of the fake rate in three different scenarios. The final weight is then applied depending on the pass and fail criteria of each lepton candidate. In the scenario where the event fails both candidate requirements, then a minus sign is included, to avoid the case of double counting. The weight is effectively a transfer factor that is created using the fake rate measured earlier. The transfer factor has its form because the weight is the ratio of tight to loose—tight excluded—instead of tight to loose—tight included.

- If event fails identification for τ 1:

$$w_1(p_T) = \frac{f_1(p_T)}{1 - f_1(p_T)} \tag{6.4}$$

- If event fails identification for τ 2:

$$w_2(p_T) = \frac{f_2(p_T)}{1 - f_2(p_T)} \tag{6.5}$$

- If event fails identification for both:

$$w_{12}(p_T) = -\frac{f_1(p_T)}{1 - f_1(p_T)} \cdot \frac{f_2(p_T)}{1 - f_2(p_T)} \tag{6.6}$$

To illustrate the different regions in the ABCD method along with each tau candidate, a diagram was drawn depicting the scenarios and is shown in figure 6.3.

This fake factor methodology has been used by other analyses such as the SM Higgs measurement with an associated Z boson [35] .

For a closure test, the same criteria are 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.
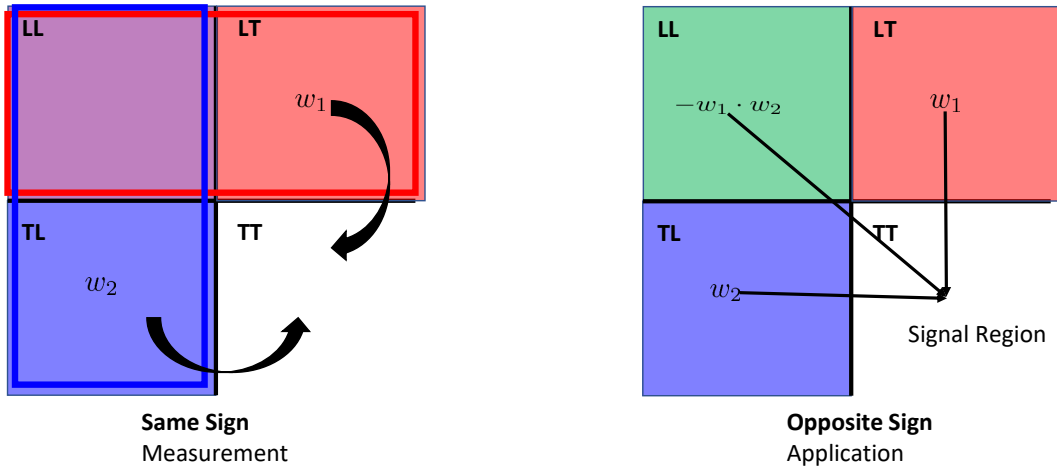
Figure 6.2: ABCD method diagram depicting the measurement and application regions for each τ_h lepton in the final sate.
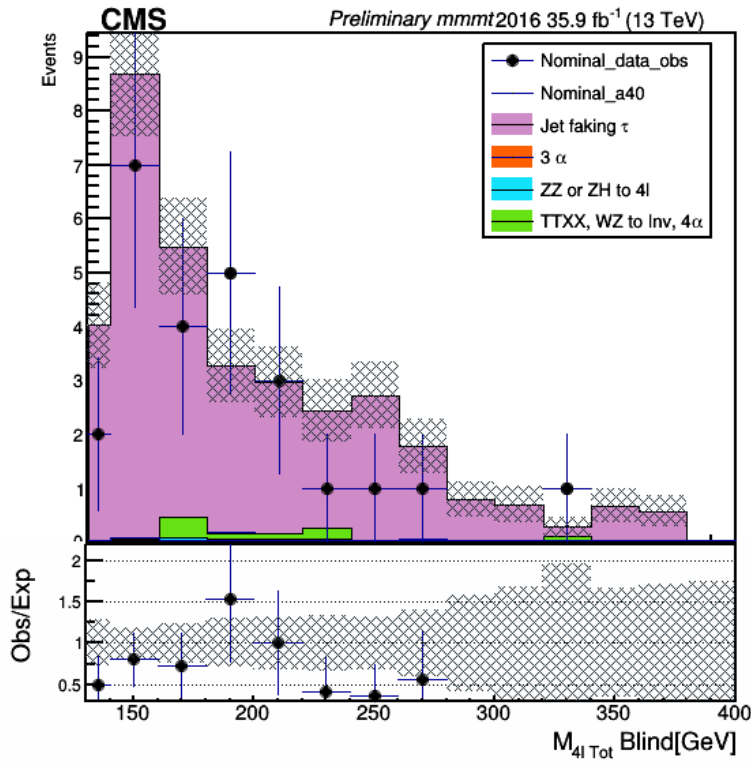
Figure 6.3: Validation of the fake factor method. Fake factors are applied to the same sign tight region.

Chapter 7

Control Plots

7.1   Data MC control plots

After object selection, event selection, background estimation, and corrections are conducted, the data and MC can be compared in the control region. Plots were created for each year and final state. To present the results, the years were combined and so were the final states. The combined result for all the years in the invariant mass of the four lepton system is shown in figure 7.1. The combination of all the years and all the final states in several kinematic variables can be found in figure 7.2. The composition of the histograms and distributions are as follows. Data is always in black points. Signal is outlined in blue. Background coming from the data driven method is in purple (or in the case of 2018 $\mu\mu e\mu$ Z+Jets and W+Jets MC) and is listed as "jet faking tau". All other backgrounds are from MC and are listed in the legend for the corresponding plot. Later in the fit model located in chapter 8, the two background processes considered are those involved from two Z bosons—also denoted irreducible—and from the data driven background—also denoted reducible.
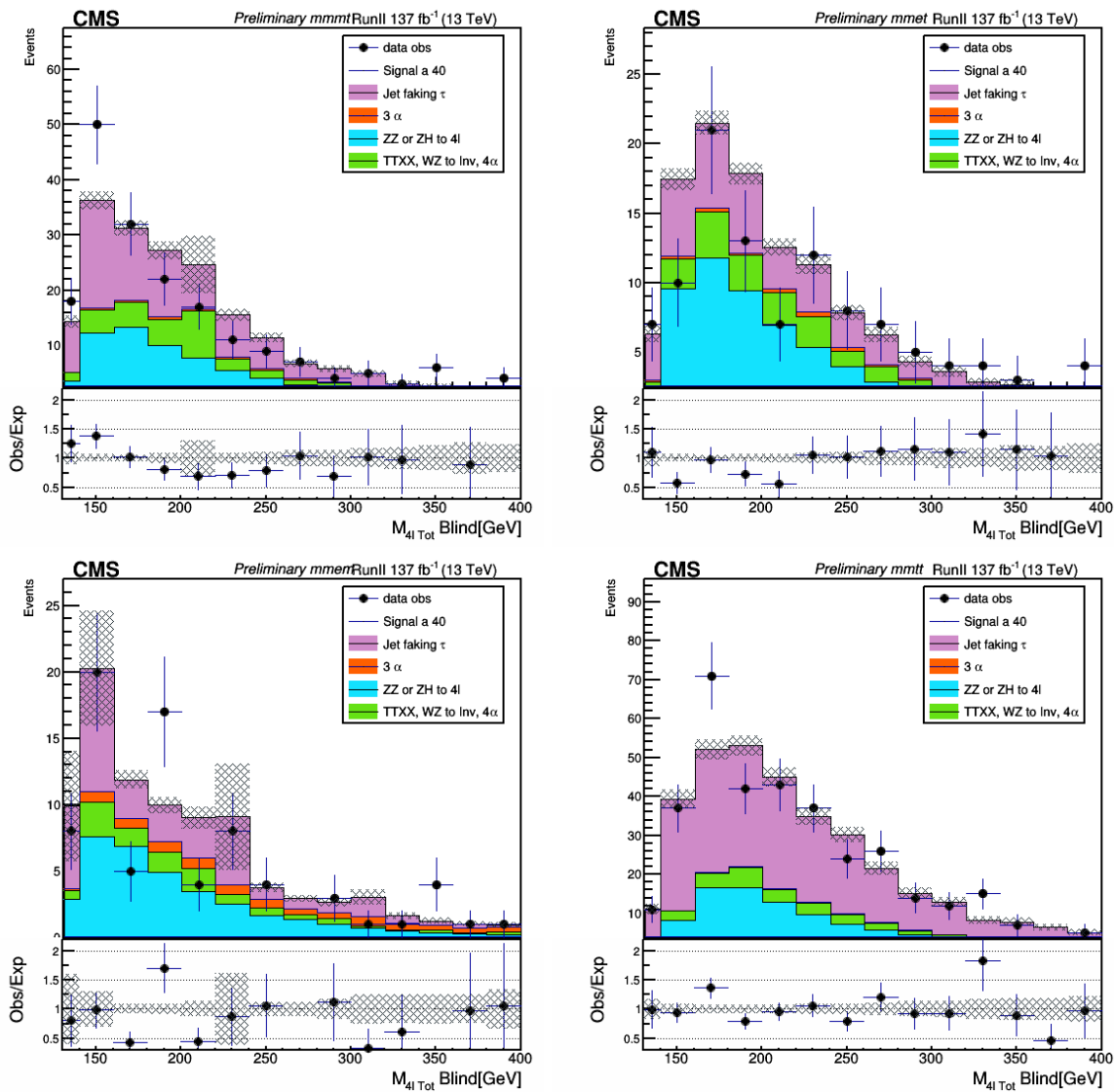
Figure 7.1: Invariant mass of the four lepton system using the full Run II dataset for each final state
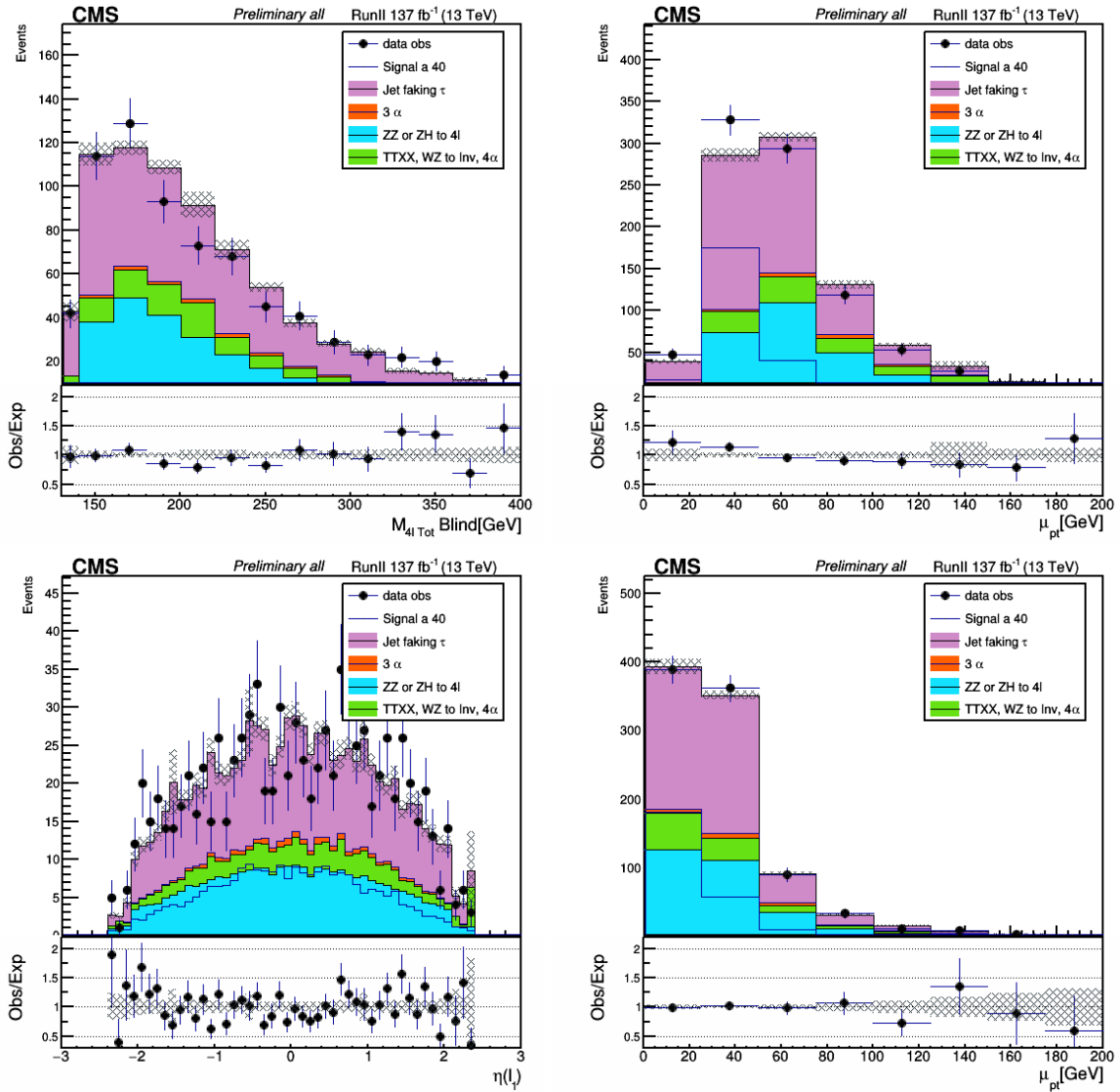
Figure 7.2: Several data MC control plots for the combination of the full Run II dataset and all final states.

Chapter 8

Statistical Inference Modeling

8.1 Hypothesis testing

In order to conduct a hypothesis test, a test statistic is needed. In high energy particle physics this is typically constructed through the *profile likelihood ratio* and then a confidence level is set using the test statistic [36]. To start, a binned histogram containing a distribution from a kinematic variable is chosen—like the mass of the parent particle—to be used in the hypothesis test. To construct the likelihood, the typical approach is to assume a Poisson distribution as the probability density function (PDF) for the events in the ith bin of the kinematic variable and then "smear" it by multiplying it with a Gaussian that is also dependent on the events. The Poisson distribution represents the true number of events one would expect and the Gaussian represents the systematic error—also denoted as nuisance parameters— that are endemic to the model. The events are split into signal and background by construction. The amount of signal events is not known and is the subject of the search, so they are allowed to vary by a coefficient denoted as μ - the signal strength. For an upper limit, the signal strength is changed until the cumulative distribution function of the test statistic reaches the desired confidence or p-value. For binned analyses, the product of Poisson distriubitons is used for the likelihood:

$$\mathcal{L}(\text{data}|\mu, \theta) = \text{Poisson}(\text{data}, \mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \tag{8.1}$$

For n_i events in the ith bin:

$$\text{Poisson}(\text{data}, \mu \cdot s(\theta) + b(\theta)) = \prod_i \frac{(\mu s_i + b_i)^n}{n_i!} e^{-\mu s_i - b_i} \tag{8.2}$$

For reference: s and b demarcate the signal and background events respectively, i the bin number, n the total number of expected events, and θ the nuisance parameter of $p(\tilde{\theta}|\theta)$ the Gaussian.

In the case of this analysis, a slightly different likelihood is considered. An unbinned parametric shape-based approach is studied, which uses probability density functions over the entirety of the fit variable. The fit variable is often called the 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. The construction of the likelihood for the unbinned parametric shape is then:

$$\mathcal{L}(\text{data}, \mu \cdot s(\theta) + b(\theta)) = k^{-1} \prod_i (\mu S f_s(x_i) + B f_b(x_i)) e^{-\mu S - B} \tag{8.3}$$

Where k is the normalization, S and B the event rate for signal and background, $f_s(x_i)$ and $f_b(x_i)$ the probability density functions for signal and background, and i the number of different categories. A similar approach can be found in reference [37]. A Bernstein polynomial would capture the slow changing background and the Voigtian captures the sharp peaking dimuon mass signal. Therefore in the pseudoscalar analysis, a product of Bernstein polynomials for $f_b(x_i)$ and Voigtian functions for $f_s(x_i)$ are used.

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

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\mathrm{D}|\mu, \hat{\hat{\theta}}_\mu)}{\mathcal{L}_{\max}(\mathrm{D}|\hat{\mu}, \hat{\theta})} \quad 0 \leq \hat{\mu} \leq \mu \tag{8.4}$$

   Where $\hat{\mu}$ and $\hat{\theta}$ are the maximum likelihood estimators, $\hat{\hat{\theta}}$ the optimized value of the estimator for the nuissance parameters, and D is the input dataset typically chosen as simiulation for expected limits and real data for the observation.

2. Find the *observed* value of the test statistic $\tilde{q}_\mu^{obs}$ for given signal strength μ.

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 PDFs 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 test statistic's PDFs under the assumption of a signal strength.

5. Define p-values to be associated with the actual observation for both $s + b$ and b only hypotheses

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | s + b) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \qquad (8.5)$$

$$1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_b^{obs} | b) = \int_{\tilde{q}_b^{obs}}^{\infty} f(\tilde{q}_\mu | \mu = 0, \hat{\theta}_b^{obs}) d\tilde{q}_\mu \qquad (8.6)$$

. Then take the ratio to form the confidence levels (CLs)

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \qquad (8.7)$$

.

6. Let α be the measure of confidence, then for $\mu = 1$ if $CL_s \leq \alpha$ then the signal hypothesis is rejected in favor of the background only hypothesis

7. Further, to quote a 95% confidence level on μ, we need to adjust the signal strength (μ) , until $CL_s = 0.05$

8.2   Worked example of an upper limit with low statistics

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

$$N = B \cdot \sigma \cdot A \cdot \mathcal{L}, \qquad (8.8)$$

where B is the branching fraction for the physics process, A is the signal acceptance, σ is the cross section, and $\mathcal{L}$ is the luminosity. As outlined in the previous section, suppose that the number of expected events follows a Poisson distribution. In the case of a background model predicting zero average events, the upper 95% bound would then be 3.7 events.

Inverting the relation

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}}, \qquad (8.9)$$

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

$$A = \frac{\text{events pass all cuts}}{\text{starting events}} = \frac{1293}{250000} \approx 0.005, \qquad (8.10)$$

and taking the cross section for gluon gluon fusion production of the Higgs $\sigma = 48$pb, along with the luminosity for 2016 $\mathcal{L} = 35,900$pb^{-1}. Then the upper 95% CL limit on the branching fraction is

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}} = 0.00043 = 4.3 \times 10^{-4}. \tag{8.11}$$

8.3   Fit model for the 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 8.2. The Voigtian shape was chosen to reflect the narrow simulated peak that is statistically smeared by experimental measurement. The Voigtian function is a Gaussian convoluted with a Lorentizian function, so in addition to the Gaussian paramters there is one extra degree of freedom which is assiocieted with the Lorentzian. The Lorentizian parameter controls the sharpness the distriubiton. For the signal MC, the standard deviation of the distributions tends to increase as the mass approaches 60 GeV. To compare the signal MC distributions, they are all plotted in figure 8.1.

Shapes from the signal samples in intervals of 5 GeV across the whole fit range 20-60 GeV are interpolated using spline functions for the fit parameters, thus precise limits can be obtained at the 1 GeV granularity. The interpolated model describes the signal samples well and produces similar results for the distributions at the 5 GeV granularity. A spline function is created for the mean, standard deviation, normalization,and the Lorentzian. A first order polynomial is used to fit the mean and a third order polynomial is used to fit the standard deviation, normalization, and Lorentzian parameters. Examples of such functions are shown in figure 8.3. The bands that envelope the spline indicate the spread and the accepted error on the spline in the statistical inference model.

A RooPlot of "m$_{\mu\mu}$ Total"
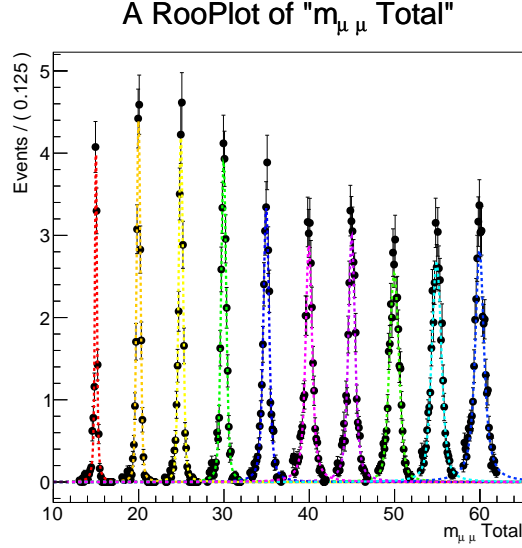
Figure 8.1: Signal fit using a Voigtian function for all MC simulated mass points

For the irreducible background coming from $ZZ \to 4l$, a Bernstein polynomial is used to fit the shape over the entire a mass range in figure 8.7. Depending on the final state and shape, the degree of the polynmoial is chosen by best fit. A Fischer F-test was conducted and there is not enough statistics in the bins to provide an accurate difference in the log-likelihood in order to recommend a particularly higher order polynomial over other orders. Thus for $\mu\mu\tau\tau$ and $\mu\mu e\tau$, a 1st order polynomial 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 quadratic function is used. The true values of the error estimation on the parameters are taken from the fit itself and can be seen in the plots like in figure 8.7. The error on these shape parameters are shown in the impacts, which demonstrate how the fit parameters affect the overall statistical inference model 8.9.

For the jet faking tau background, a Bernstein polynomial is also used to fit the shape over the entire a mass range indicated in figure 8.8. Similar to the ZZ (irreducible) background, the jet faking tau background polynomial degree is chosen by best fit. The rest of the channels and years are shown in the appendix D.

8.4 Systematic uncertainties

In order to measure the systematic effects on the final fit distributions, changes in the fit templates are done and propagated to the fit model in the form of rate parameters. These rate parameters
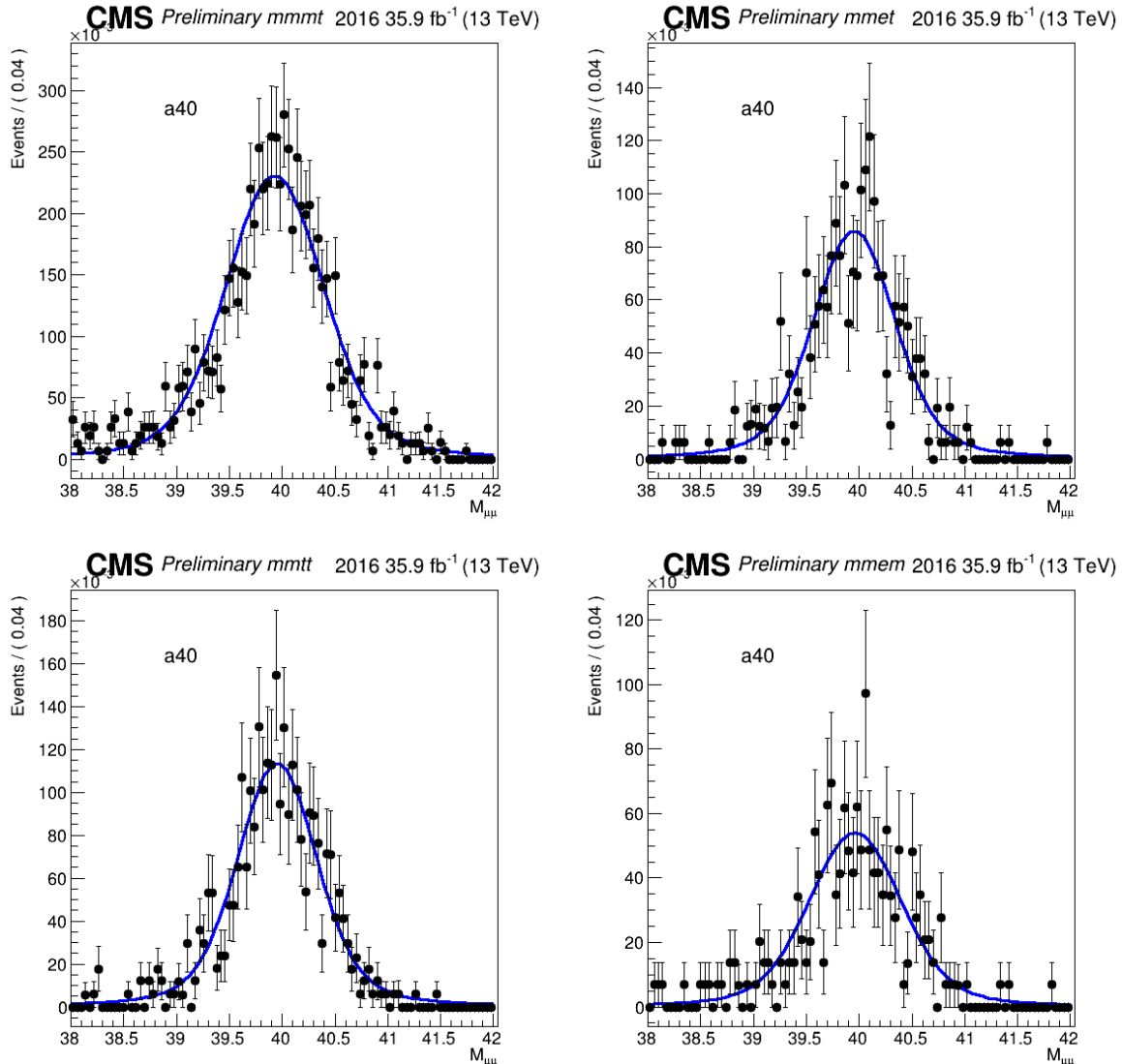
Figure 8.2: Signal fits using a Voigtian function for a-mass at 40GeV

differ slightly between the signal and background distributions. For background, the error in the fit parameters is directly included in the uncertainty model.

The uncertainty of the spline function affects the uncertainty for the signal, so it is included in the fit model. As mentioned in the fit model section 8.3 and shown in figure 8.3, 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 number. Regarding the section highlighting the corrections 4.4,
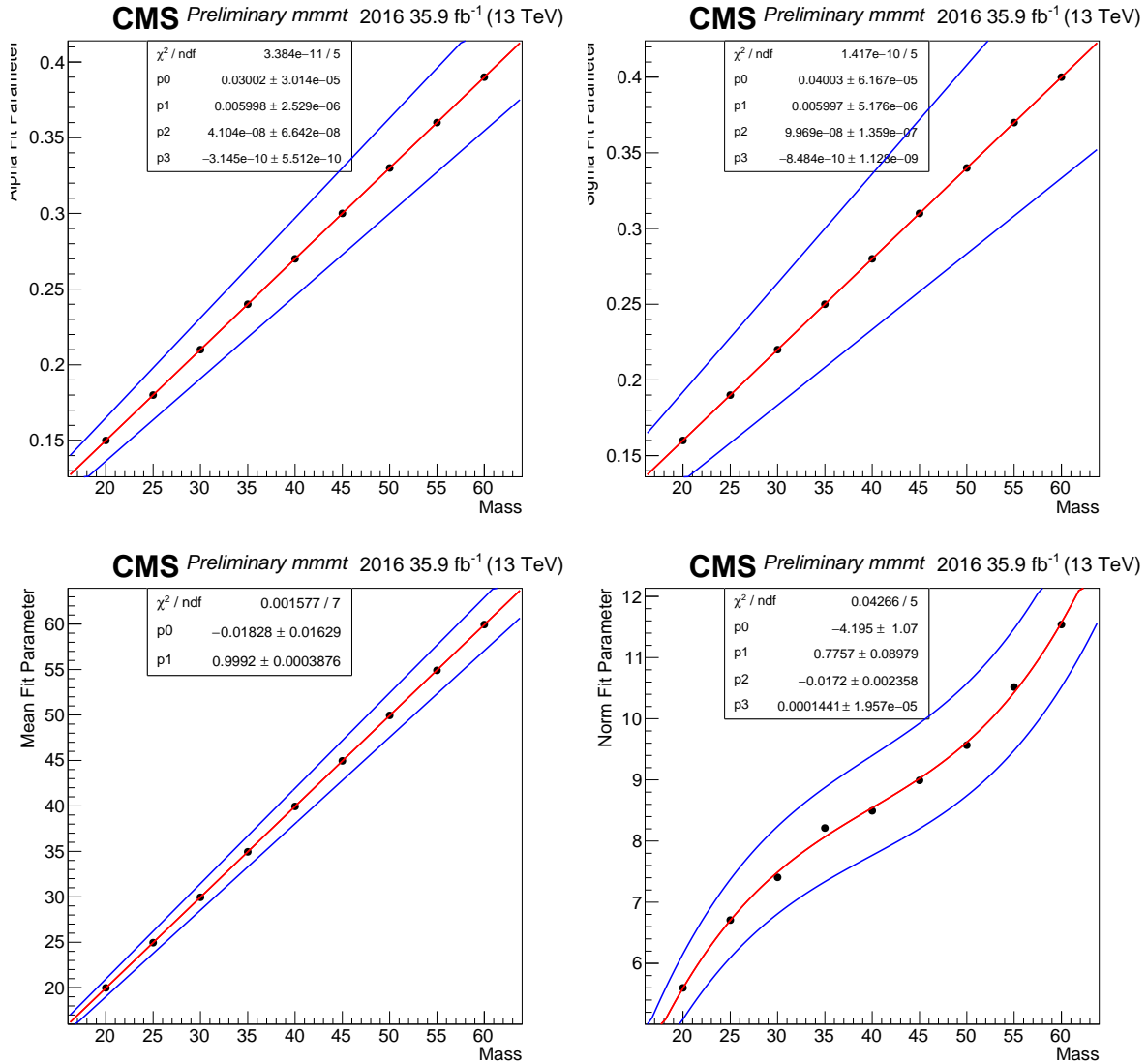
Figure 8.3: Spline functions for 2016 mmmt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean

one can see the bin-shift from the energy scale. The bin-shift indicates the amount the mean of the distribution is affected from the energy scale shift. The shift should fit within the envelope of the percentage on the parameter for it to be modeled correctly.

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 used.

The extent to which these systematics effect the search is calculated through the concept of the "impact". An impact is a way to see how that systematic uncertainty impacts the overall statistical model. To measure an impact for a particular systematic uncertainty, it is allowed to
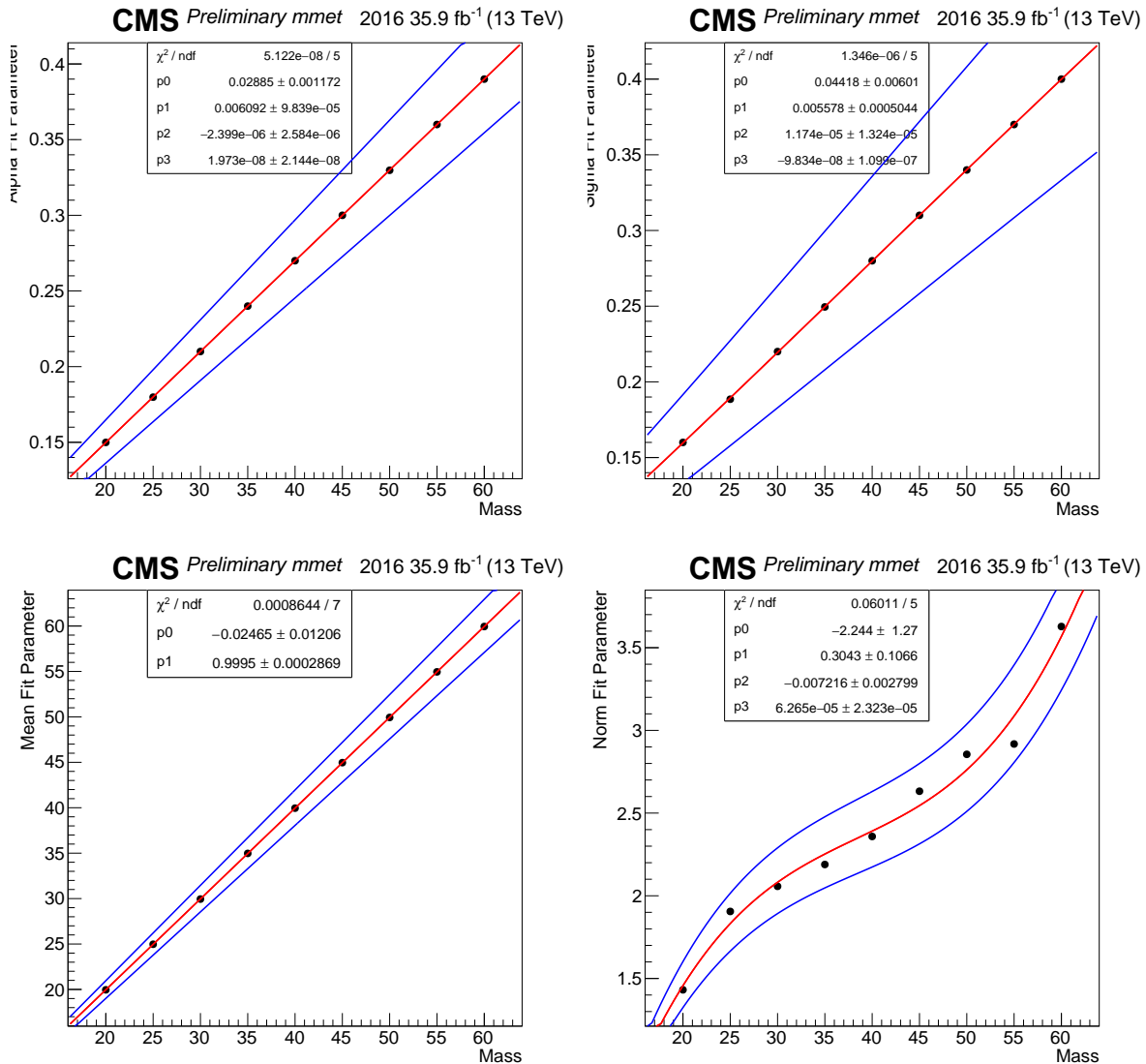
Figure 8.4: Spline functions for 2016 mmet a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean

vary within the fit range while the rest of the parameters in the likelihood function are frozen. The corresponding difference in the signal strength is measured. In order to read the impact plots and to understand what the impacts represent in the fit model please look at table 8.1. Each impact has a corresponding nomenclature, range of variation, and effect on the POI. The nomenclature is listed in 8.1 and the variation and effect on the POI is seen in the plots 8.9.
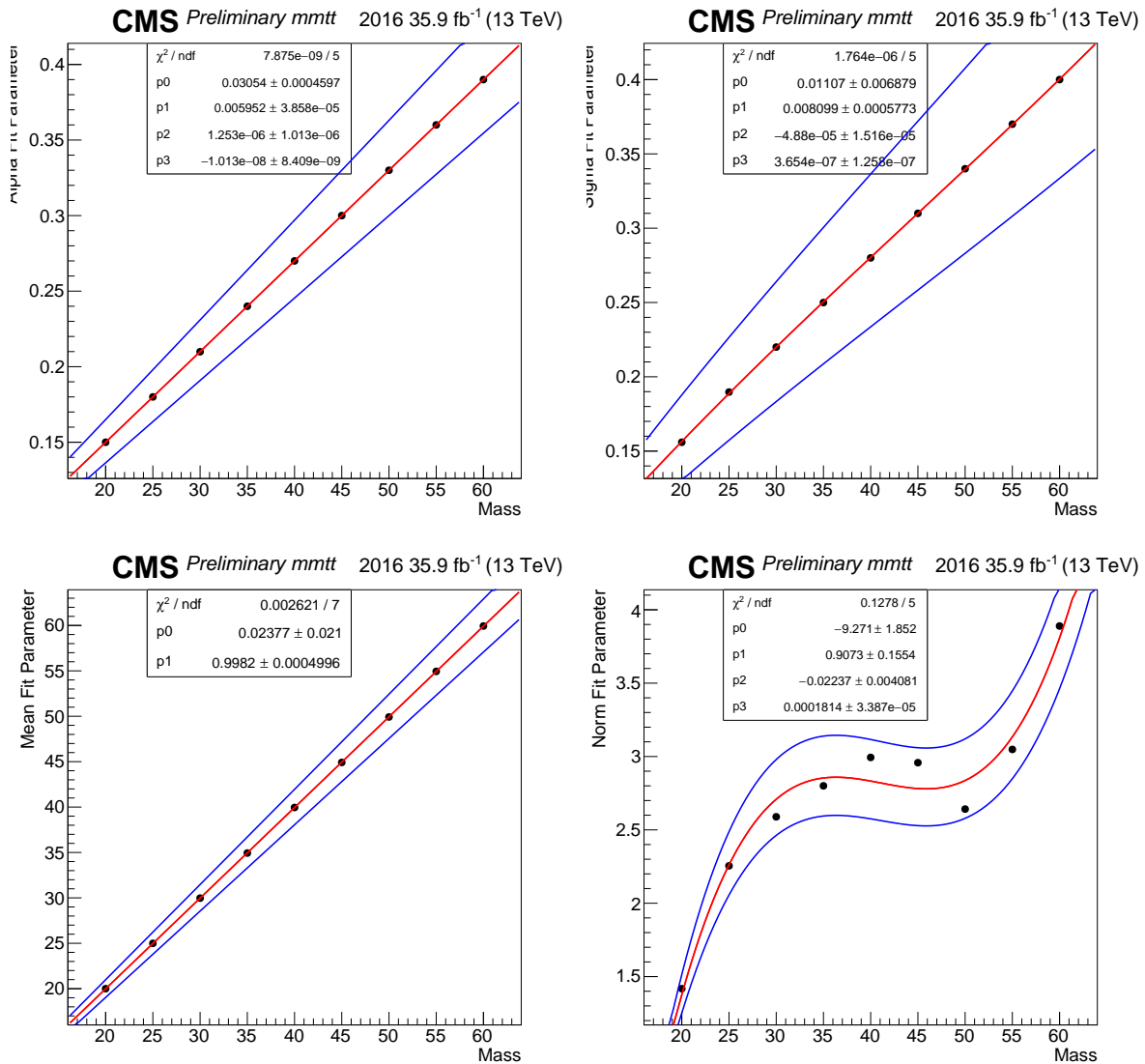
Figure 8.5: Spline functions for 2016 mmtt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
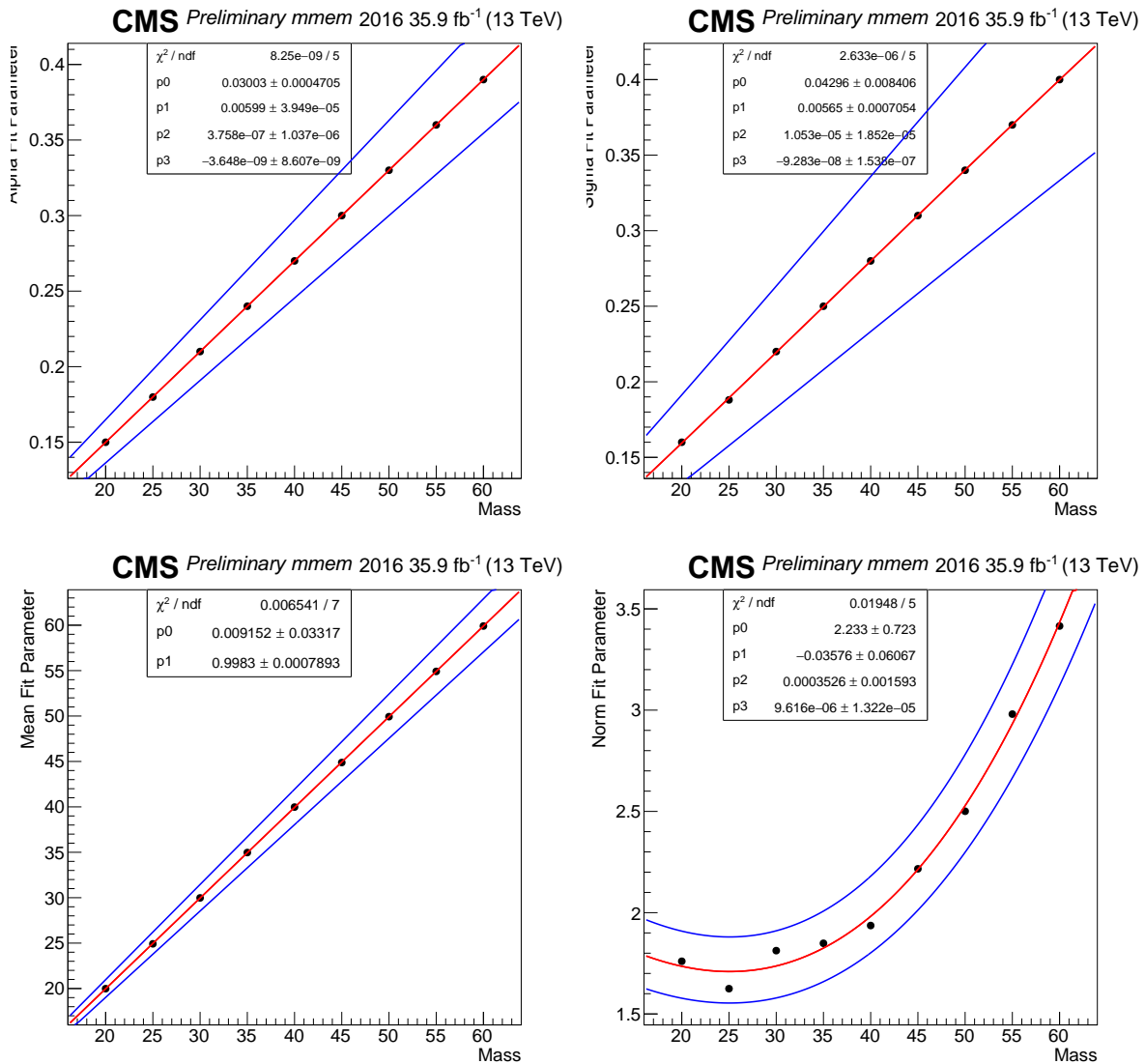
Figure 8.6: Spline functions for 2016 mmem a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
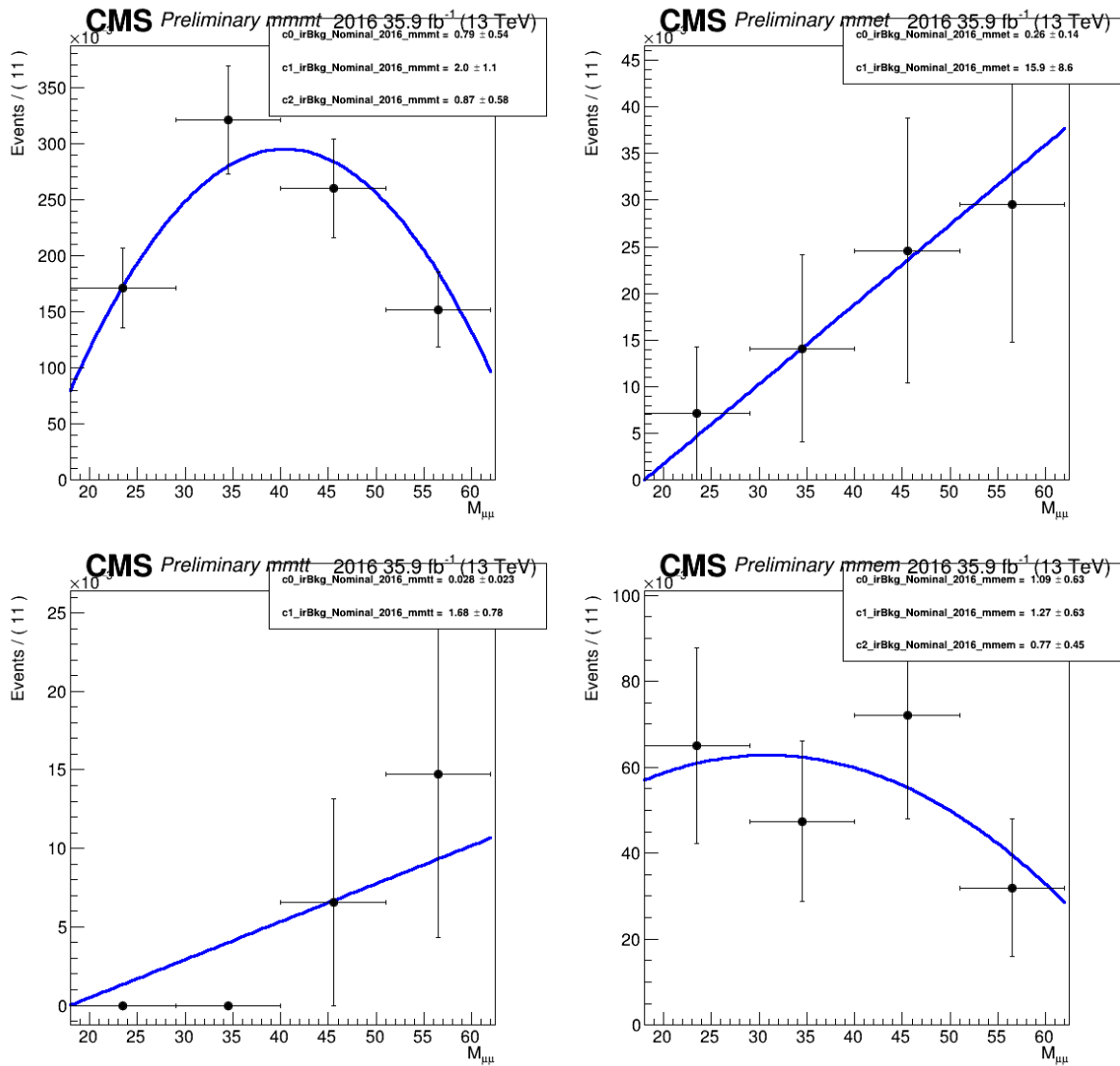
Figure 8.7: Irreducible background fit using Bernstein polynomials
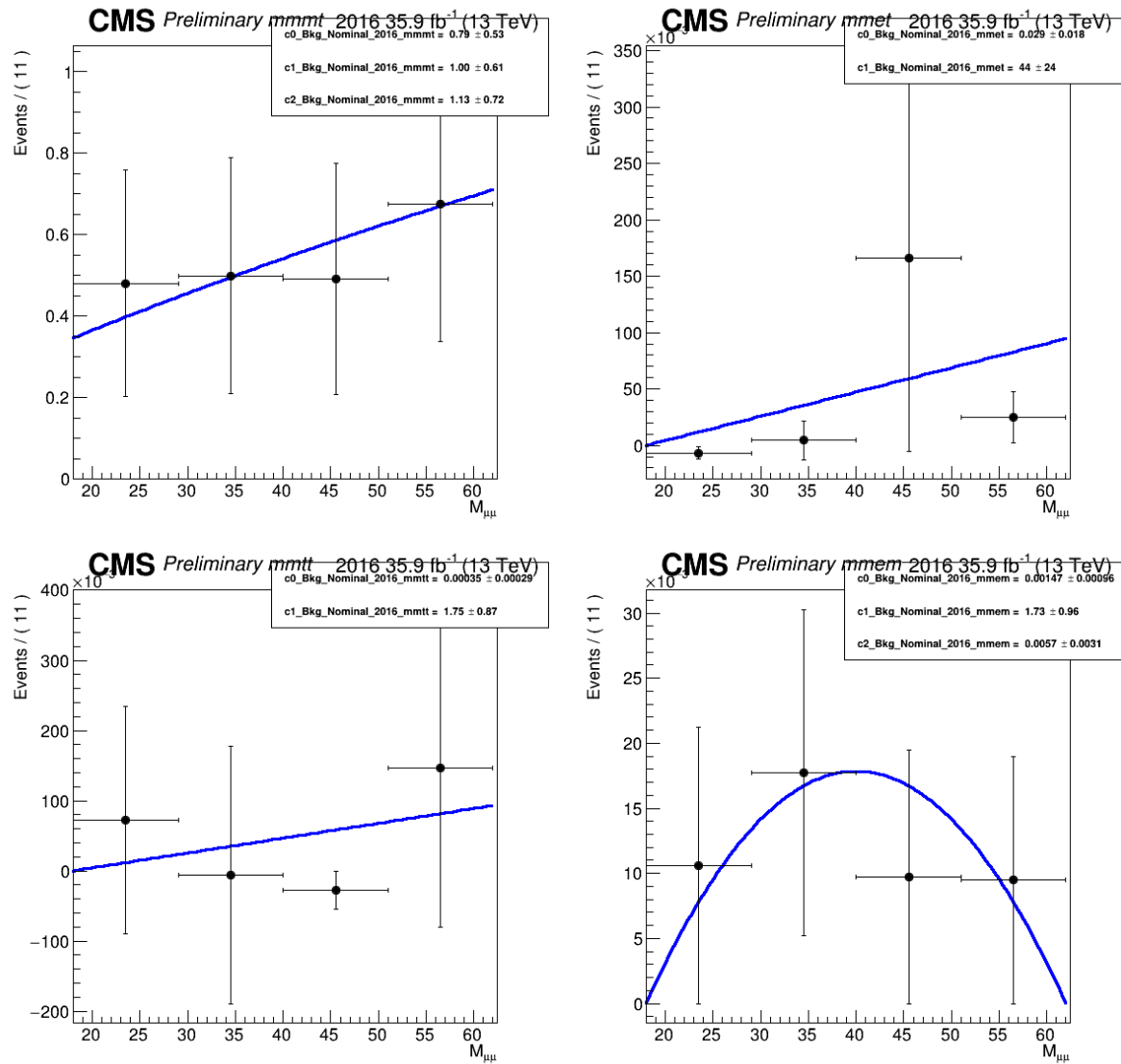
Figure 8.8: Reducible background fit using Bernstein polynomials

Table 8.1: List of uncertainties with the corresponding name and description. The name directly refers to the variable in the impact plots 8.9.

Name of Uncertainty	Description	Magnitude
scale	e, μ, and τ (split by decay mode) energy scales	% ch. dep.
c0_, c1_, ... cN_	Coefficents of the Bkg (datadriven) or irBkg (ZZ) parametric shape	% ch. dep.
_norm	normalization of the parametric shape for backgrounds	% ch. dep.
lumi	luminosity uncertainty	1.6%
intAlpha	alpha interpolated spline function shape uncertainty	10%
intSigma	sigma interpolated spline function shape uncertainty	20%
intMean	mean interpolated spline function shape uncertainty energy scale shift uncertainties for signal included	5%

A systematic impact plot, sometimes referred to as a pull distribution, is listed in figure 8.9. The rest of the channels and years are located in the appendix C.
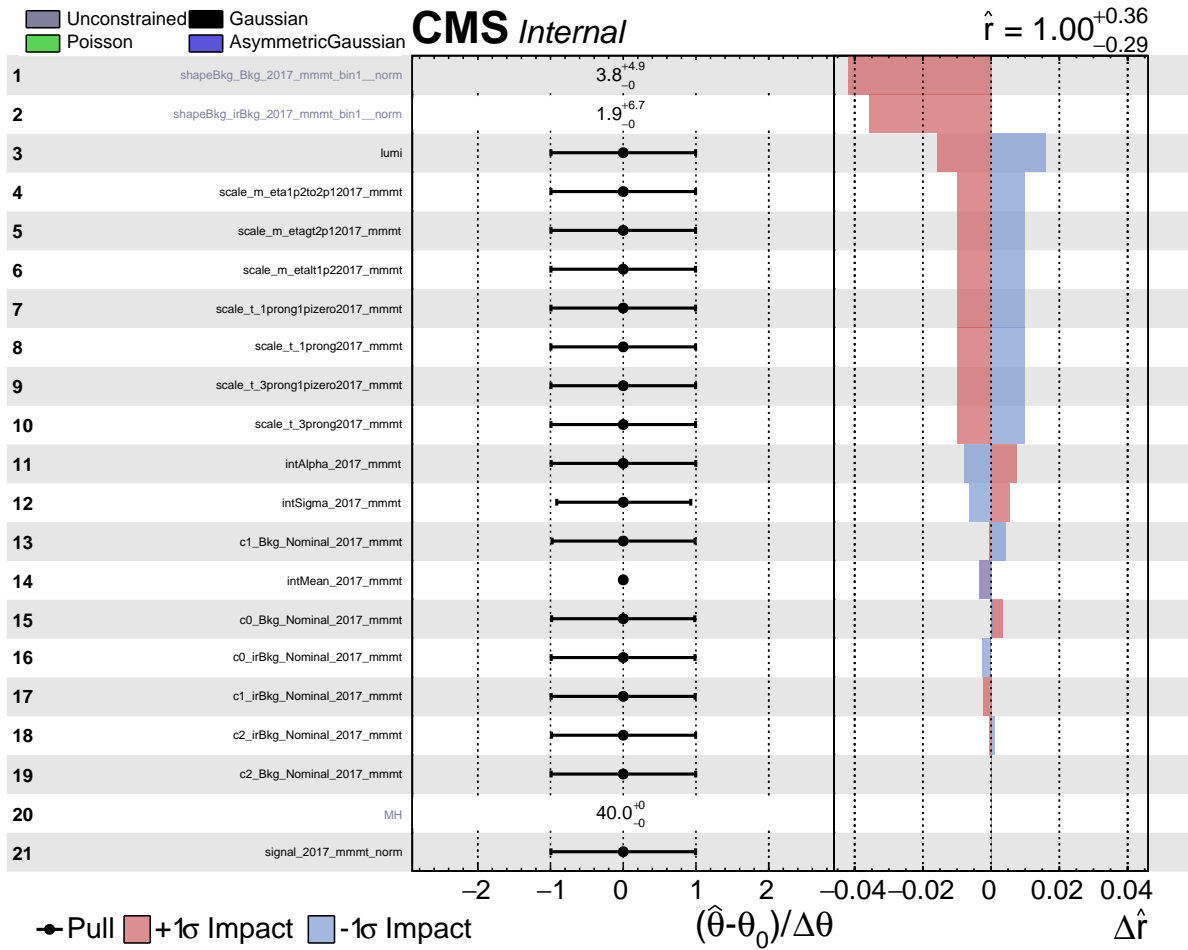
Figure 8.9: Expected systematic impacts for the fit model in the $\mu\mu\mu\tau$ final state, for normalization impacts (shapeBkg_Bkg and shapeBkg_irBkg), they are expected to be one-sided, and —as mentioned in the fit model 8.1—the mean is precisely measured and is expected to be constrained less than one for most channels.

Chapter 9

Conclusions

9.1 Results

After the final event selection, including the signal extraction cuts listed in section 5.2.2, the statistical hypothesis test can be made. The final number of events listed in each category for the full Run II dataset is shown in the table 9.1 below.

Year	Signal Total	Background Total		
	$m_a = 40$ GeV	Data Driven (FF)	Irreducible (ZZ)	Total
2016	16.15	3.21	1.43	4.64
2017	19.49	6.63	3.26	9.89
2018	27.45	14.93	2.79	17.72
Run II	63.09	24.77	7.48	32.25

Table 9.1: Expected event yields of signal and background categories across all years with 137 fb^{-1} of data. Signal normalized to .01% of the SM Higgs Branching Fraction.

.

As discussed in the previous section 8.3, the shapes that were created are used in an upper limit for each mass point. Initial values of the signal distributions are selected to make sure that the signal strength modifier (μ) in the limit is of order unity. The range of masses in the limit is between 20 GeV and 60 GeV to ensure compatibility with $h \rightarrow aa$ combination limits for additional 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 statistically limited search, we can estimate the background yield as no events. The estimated upper limit on the branching

fraction calculated earlier is:

$$B = \frac{N}{\sigma \cdot A \cdot \mathcal{L}} = 0.00043. \tag{9.1}$$

This limit is set 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 branching fraction, which was assumed to be $\sigma_{SM}(h) \times 0.01\%$. Multiplying the CL by 0.01% returns the limit on $\frac{\sigma_h}{\sigma_{SM}} B(h \to aa \to 2\mu 2\tau)$. Preliminary limits are set using the asymptotic limit method [36] for each mass point.
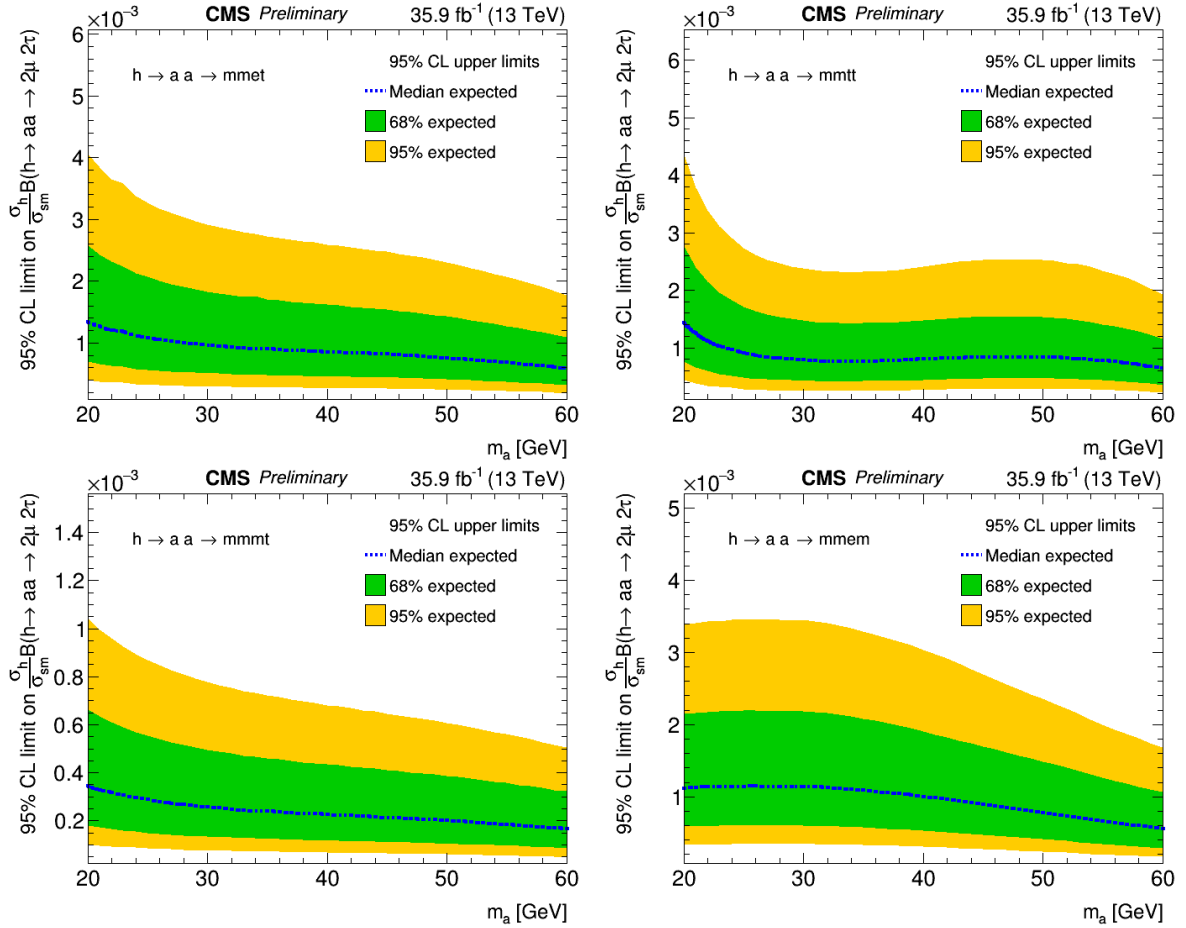
Figure 9.1: Asymptotic upper 95% CL Limits on the branching fraction times ratio of the SM cross section for 2016
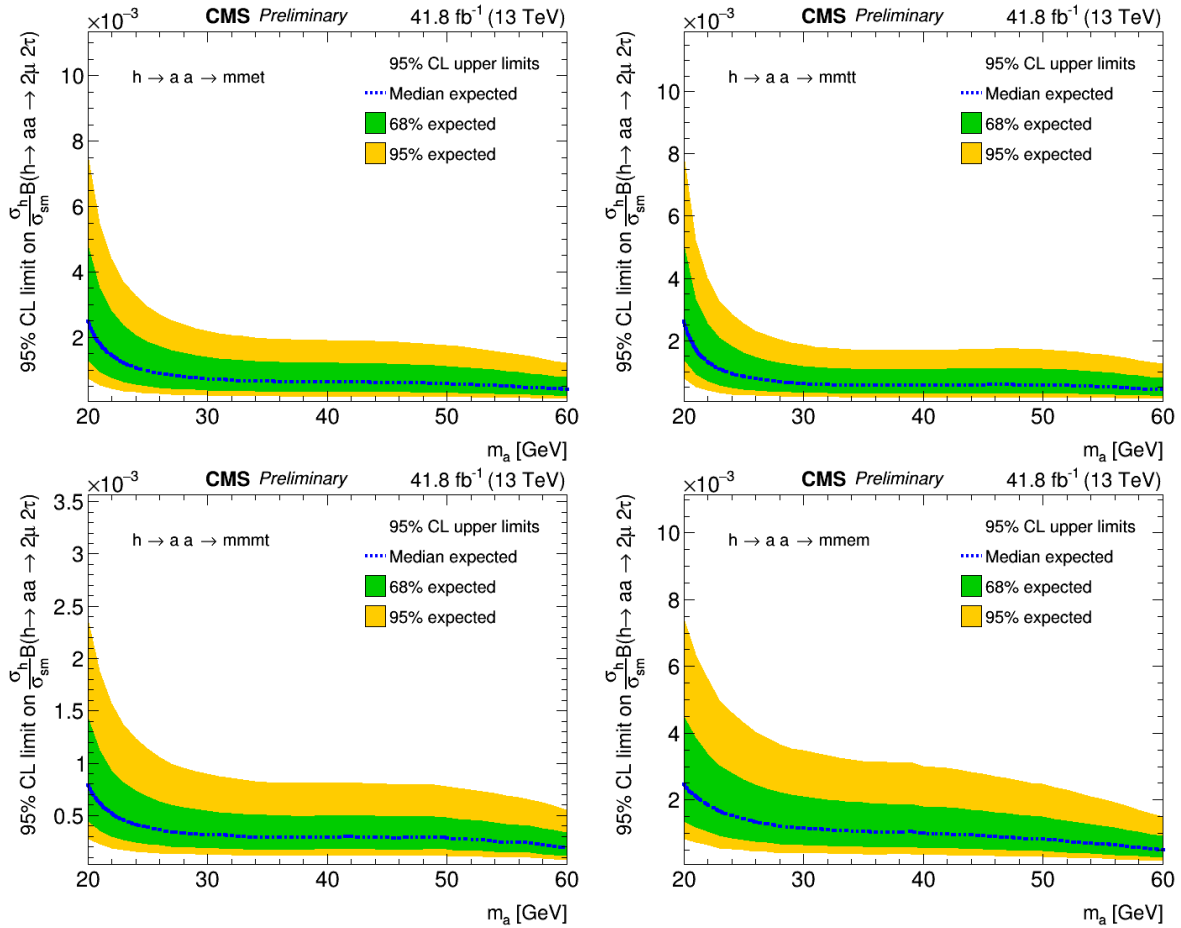
Figure 9.2:   Asymptotic upper 95% CL Limits on the branching fraction times ratio of the SM cross section for 2017
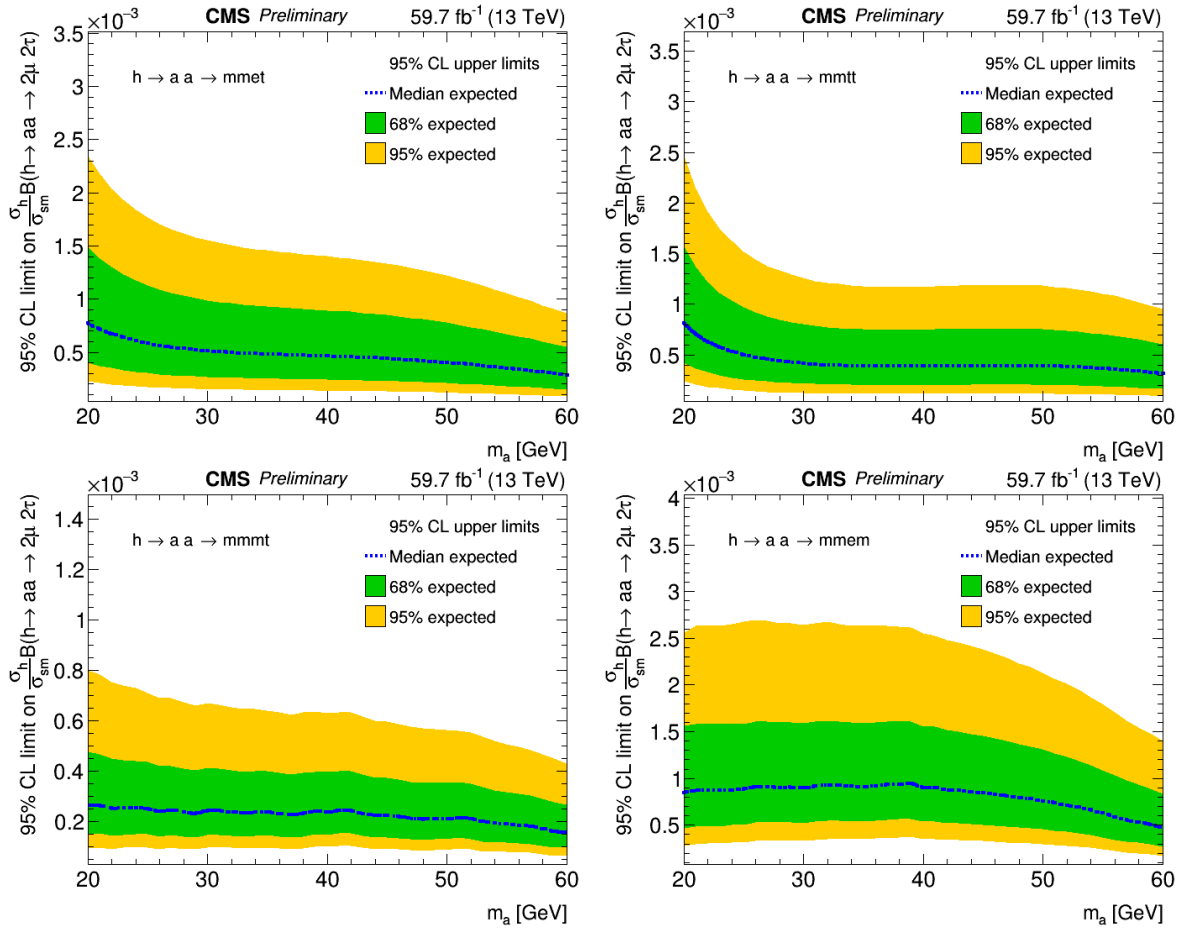
Figure 9.3: Asymptotic upper 95% CL Limits on the branching fraction times ratio of the SM cross section for 2018

All of the years and channels are then added together to form the combined result 9.1 and the model 2HDM+S interpretations for different scenarios. Type III, where coupling to τ leptons is favored, is expected to be the most stringent scenario for this analysis. More parameter space in theory is excluded at the upper 95% level in regions of lower values on the limit (regions in blue) in figure 9.5. Type I excludes mostly high mass particles and isn't depended on $\tan\beta$. Type II and III exclude more at the high $\tan\beta$ region as opposed to Type IV which excludes at the low $\tan\beta$ region.
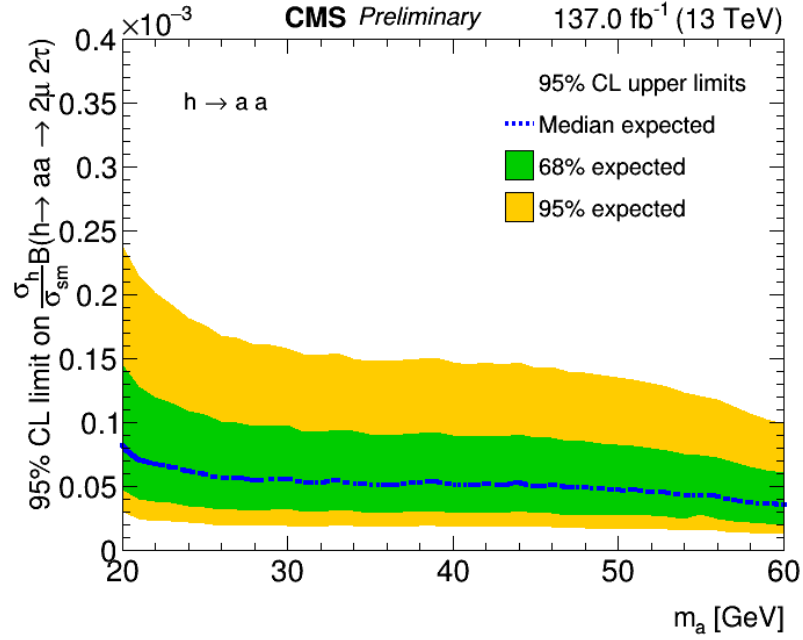
Figure 9.4: Asymptotic upper 95% CL Limits on the branching fraction times ratio of the SM cross section for the full Run II dataset (137fb^{-1})
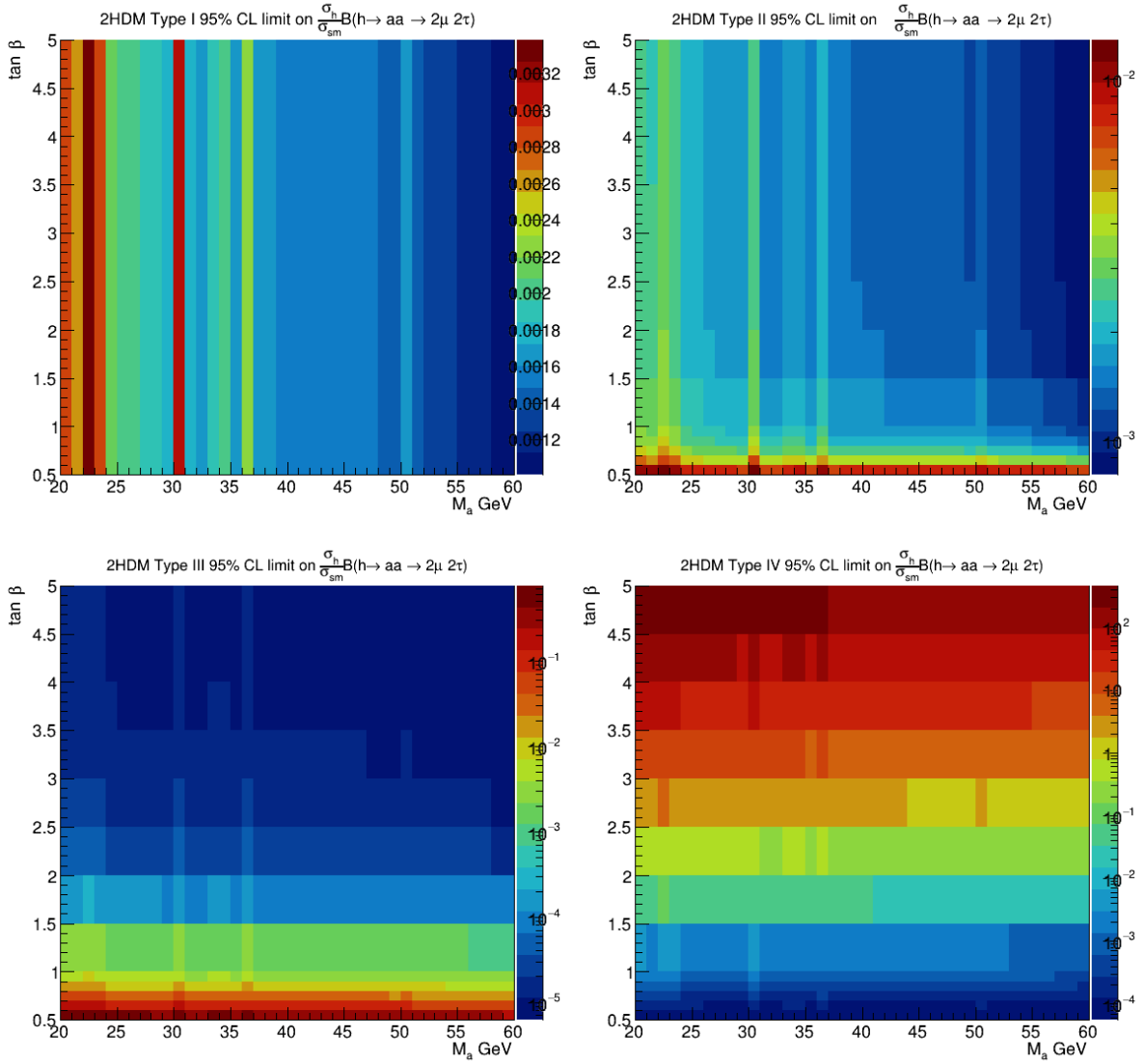
Figure 9.5: upper 95% CL limits on the branch fraction of $h \to aa$ times the ratio of the SM cross sections for the full Run II dataset (137fb^{-1}) for different 2HDM+S model specific scenarios plotted also as a function of $\tan \beta$.

9.2   Conclusion

An overview of the Large Hadron Collider, CERN, CMS, luminosity operations, and an analysis focusing on the search for a BSM processes involving an exotic Higgs-like particle was presented. Using the full Run II dataset collected at CMS corresponding to an integrated luminosity of 137fb^{-1}, the search for the SM Higgs Boson, h, decaying to a pair of pseudoscalars, a, which then decay to pairs of muons and tau leptons was completed. Expected upper 95% confidence level limits are set to about 10^{-4} after the addition of all final final states. These results are independent of separate 2HDM+S models and is considered a generic search that applies to multiple MSSM scenarios along with any BSM physics within the search window.

It has been an honor of a lifetime to work alongside CMS, Purdue University, and Princeton Univsersity to deliever this analysis and years of service work!

Appendix A

Data and Simulation Samples

A.1   Data and Simluation Used for Analysis

The full Run II dataset was used corresponding to 137fb^{-1}.

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

Data set
/DoubleMuon/Run2016C-02Apr2020-v1/NANOAOD
/DoubleMuon/Run2016D-02Apr2020-v1/NANOAOD
/DoubleMuon/Run2016E-02Apr2020-v1/NANOAOD
/DoubleMuon/Run2016F-02Apr2020-v1/NANOAOD
/DoubleMuon/Run2016G-02Apr2020-v1/NANOAOD
/DoubleMuon/Run2016H-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016C-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016D-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016E-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016F-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016G-02Apr2020-v1/NANOAOD
/SingleMuon/Run2016H-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016H-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016G-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016F-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016E-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016D-02Apr2020-v1/NANOAOD
/SingleElectron/Run2016C-02Apr2020-v1/NANOAOD
/MuonEG/Run2016H-02Apr2020-v1/NANOAOD
/MuonEG/Run2016G-02Apr2020-v1/NANOAOD
/MuonEG/Run2016F-02Apr2020-v1/NANOAOD
/MuonEG/Run2016D-02Apr2020-v1/NANOAOD
/MuonEG/Run2016C-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016H-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016G-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016F-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016E-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016D-02Apr2020-v1/NANOAOD
/DoubleEG/Run2016C-02Apr2020-v1/NANOAOD

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

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

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

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

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

Monte Carlo Datasets for 2016
/W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext1-v1/NANOAODSIM
/W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext1-v1/NANOAODSIM
/W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext1-v1/NANOAODSIM
/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext2-v1/NANOAODSIM
/W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext2-v1/NANOAODSIM
/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext1-v1/NANOAODSIM
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-ext2-v1/NANOAODSIM
/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ttZJets_13TeV_madgraphMLM-pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ttWJets_13TeV_madgraphMLM/RunIISummer16NanoAODv7-Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2e2nu_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2mu2nu_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/GluGluZH_HToWW_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/HZJ_HToWW_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ZZTo4L_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToLL_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToNuNu_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToQQ_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/ZHToTauTau_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WminusHToTauTau_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/WplusHToTauTau_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/HWminusJ_HToWW_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/HWplusJ_HToWW_M125_13TeV_powheg_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-15_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-20_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-25_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-30_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-35_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-40_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-45_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-50_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-55_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM
/SUSYGluGluToHToAA_AToMuMu_AToTauTau_M-60_TuneCUETP8M1_13TeV_madgraph_pythia8/RunIISummer16NanoAODv7-PUMoriond17_Nano02Apr2020_102X_mcRun2_asymptotic_v8-v1/NANOAODSIM

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

Monte Carlo Datasets for 2017
/DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_new_pmx_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_v3_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_v2_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017RECOSIMstep_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017RECOSIMstep_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/W1JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_ext_102X_mc2017_realistic_v8-v1/NANOAODSIM
/W2JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/W3JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_PU2017_102X_mc2017_realistic_v8-v1/NANOAODSIM
/W4JetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WZTo3LNu_13TeV-powheg-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/ttZJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/ttWJets_TuneCP5_13TeV_madgraphMLM_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT1_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT1_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT1_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext2-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext2-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_EXT_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/GluGluZH_HToWW_M125_13TeV-powheg-pythia8_TuneCP5/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/HZJ_HToWW_M125_13TeV_powheg_jhugen714_pythia8_TuneCP5/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ZZTo4L_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext1-v1/NANOAODSIM
/ZZTo4L_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8_ext2-v1/NANOAODSIM
/ZZTo4L_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_new_pmx_102X_mc2017_realistic_v8-v1/NANOAODSIM
/HWminusJ_HToWW_M125_13TeV_powheg_pythia8_TuneCP5/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/HWplusJ_HToWW_M125_13TeV_powheg_pythia8_TuneCP5/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToLL_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToNuNu_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ggZH_HToTauTau_ZToQQ_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/ZHToTauTau_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WminusHToTauTau_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/WplusHToTauTau_M125_13TeV_powheg_pythia8/RunIIFall17NanoAODv7-PU2017_12Apr2018_Nano02Apr2020_102X_mc2017_realistic_v8-v1/NANOAODSIM
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M15/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M20/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M25/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M30/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M35/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M40/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M45/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M50/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M55/
/eos/home-s/shigginb/HAA_ntuples/ggha01a01Tomumutautau_2017_M60/

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-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_ext1-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
/WJetsToLNu_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
/WWW_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
/GluGluHToTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo2e2mu_13TeV_TuneCP5_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo2e2tau_13TeV_TuneCP5_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_EXT_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21_ext1-v1/NANOAODSIM
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_EXT_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/GluGluToContinToZZTo4tau_13TeV_MCFM701_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/HZJ_HToWW_M125_13TeV_powheg_jhugen714_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_ZToLL_M125_13TeV_powheg_pythia8/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
/WminusHToTauTau_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
/ZHToTauTau_M125_13TeV_powheg_pythia8/RunIIAutumn18NanoAODv7-Nano02Apr2020_102X_upgrade2018_realistic_v21-v1/NANOAODSIM
/HWminusJ_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 data driven background estimatation of the rate at which jets fake tau leptons in the QCD, ttbar, and W+jet regions. The y-axis can be interpreted as the percent fake rate.
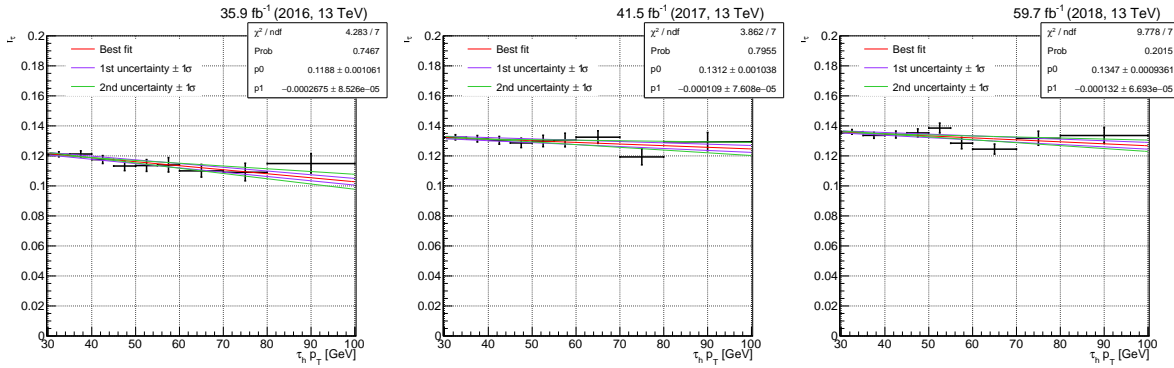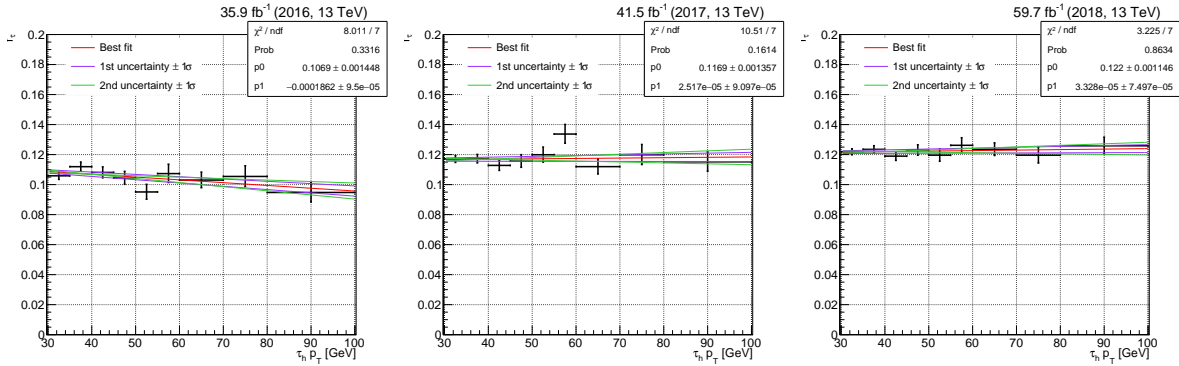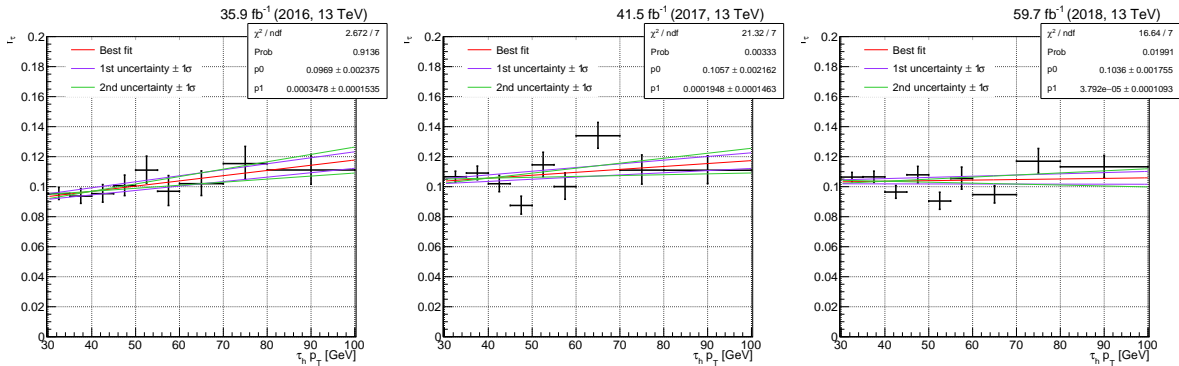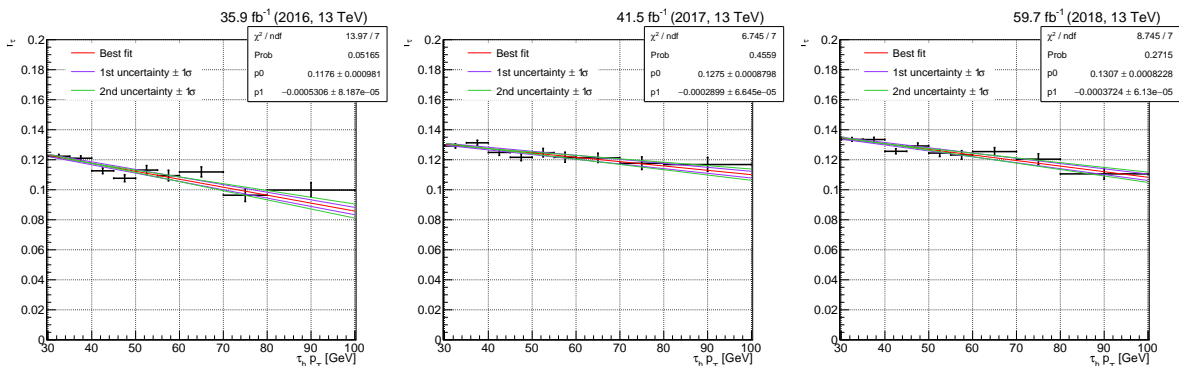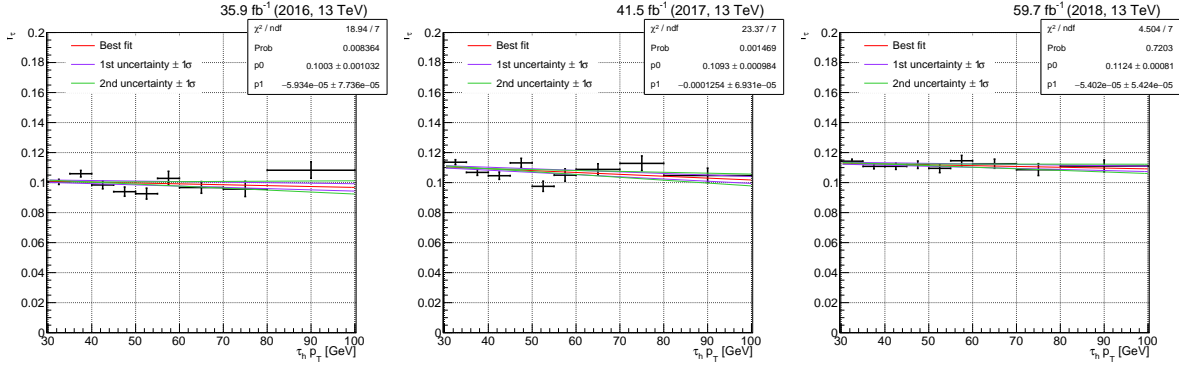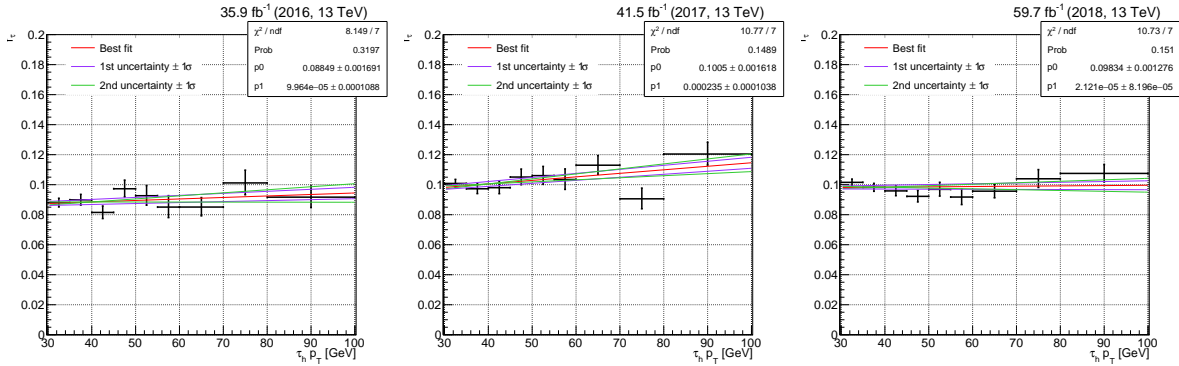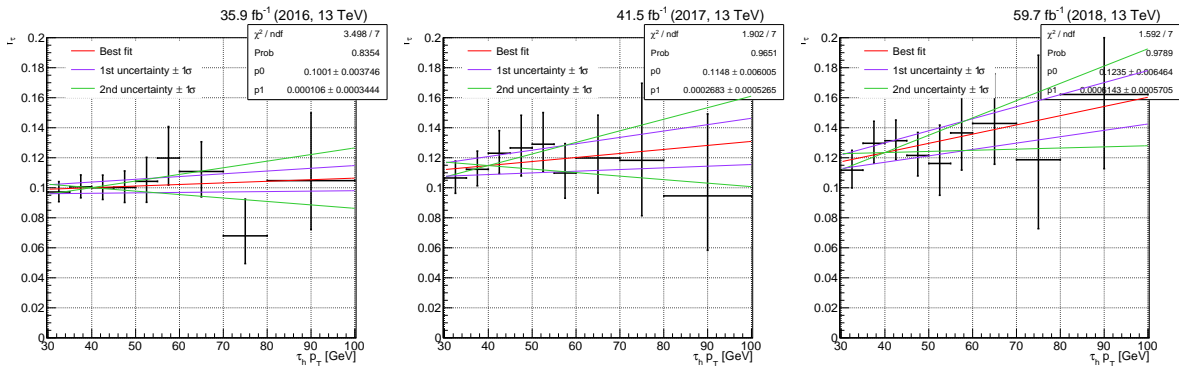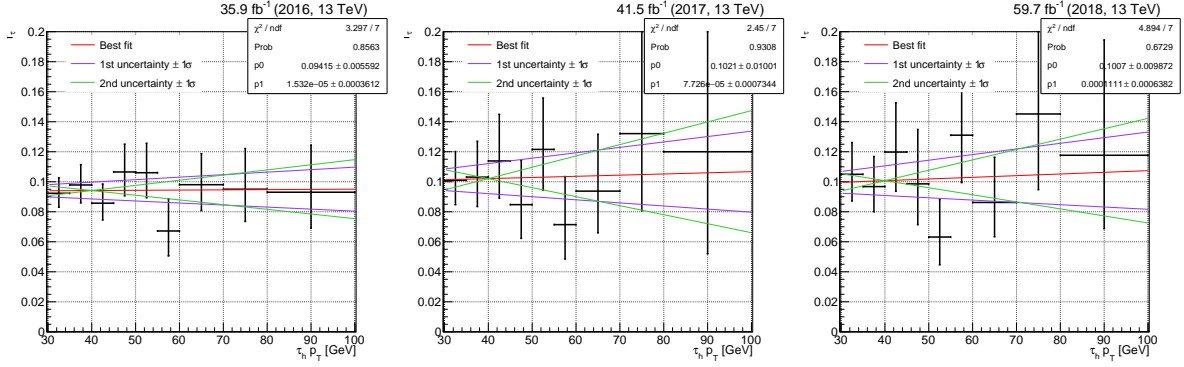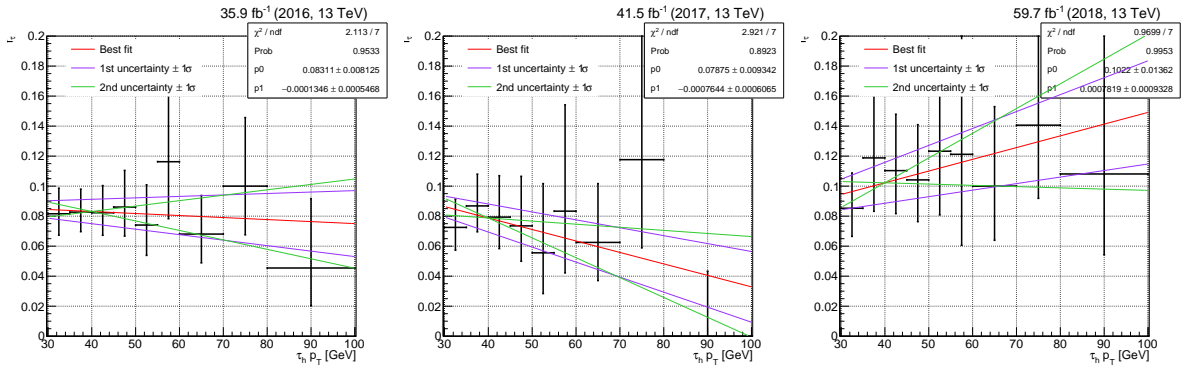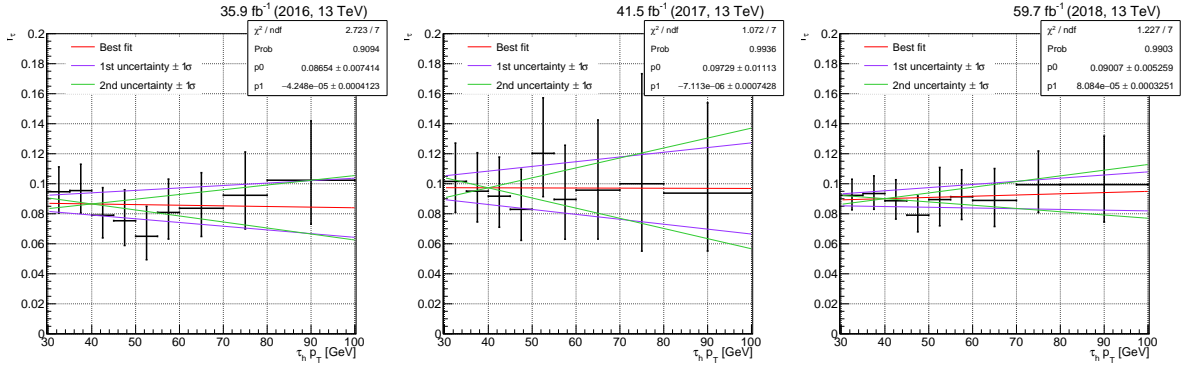

Figure B.1: Fake factors determined in the W+jets determination region with zero jets in the $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_h p_T. The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.
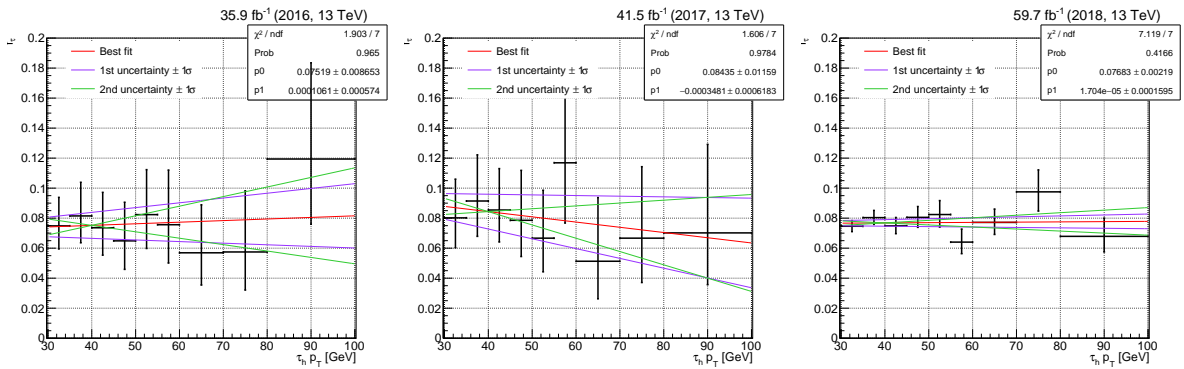
Figure B.2: Fake factors determined in the W+jets determination region with one jet in the $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h\ p_T$. The green and purple lines indicate the shape systematics obtained by uncorrelating the uncertainties in the two fit parameters returned by the fit.
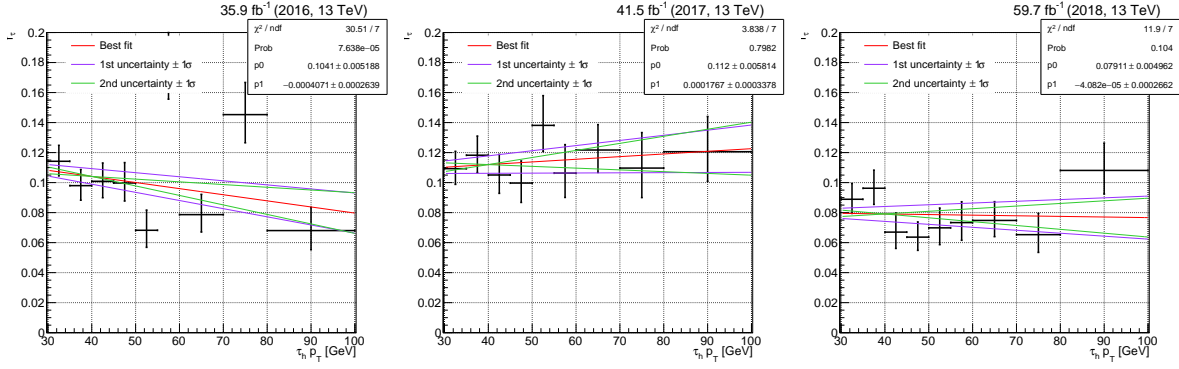

Figure B.3: Fake factors determined in the W+jets determination region with at least two jets in the $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the $\tau_h\ p_T$. The green and purple lines indicate the shape systematics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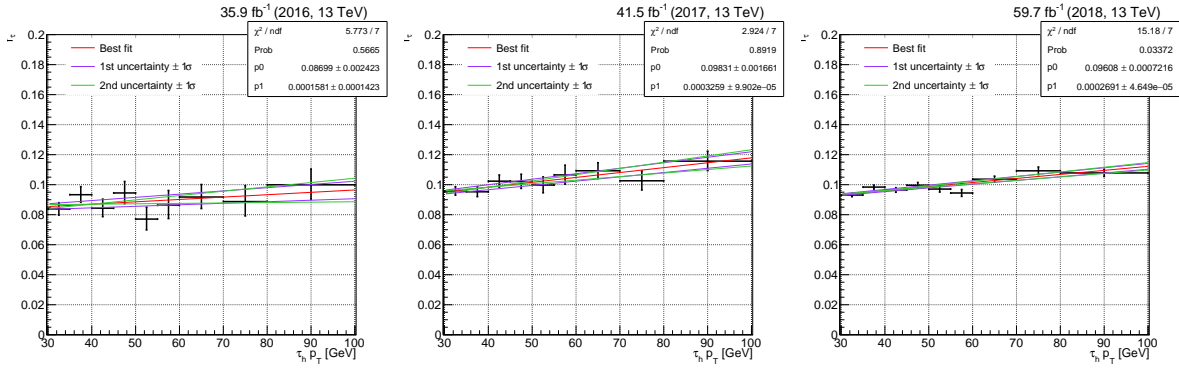
Figure B.5: Fake factors determined in the W+jets determination region with one jet in the τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_h p_T. The green and purple lines indicate the shape systematics obtained by uncorrelating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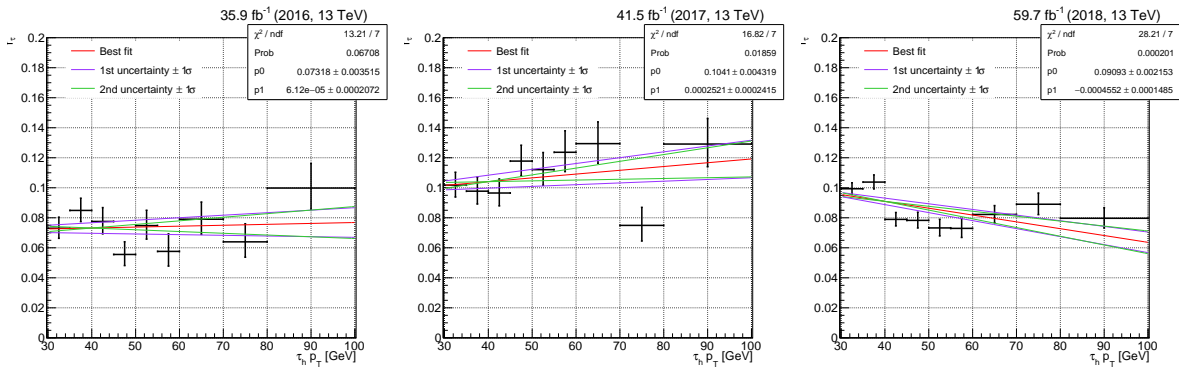

Figure B.6: Fake factors determined in the W+jets determination region with at least two jets in the τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_h p_T. The green and purple lines indicate the shape systematics obtained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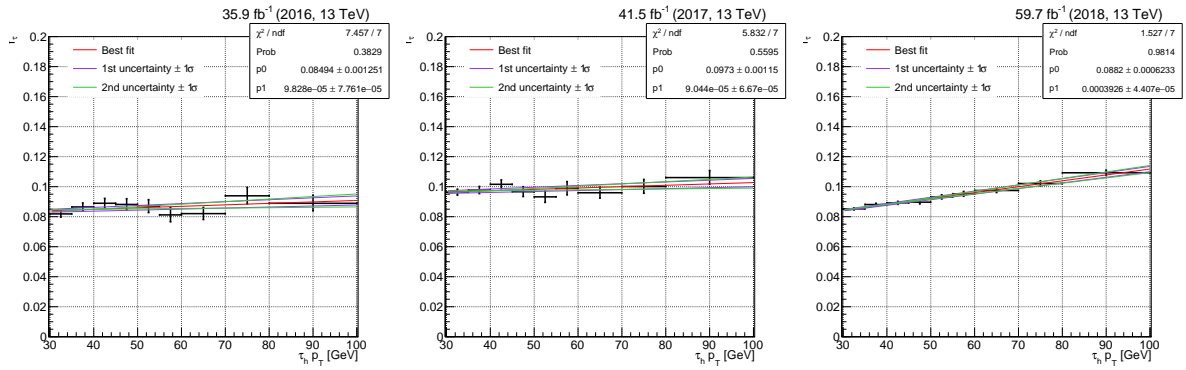

Figure B.7: Fake factors determined in the QCD multijet determination region with zero jets in the $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_h p_T. The green and purple lines indicate the shape systematics obtained 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 $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 $e\tau_h$ measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 τ_h measured state in 2016 (left), 2017 (center), and 2018 (right). They are fitted with linear functions as a function of the τ_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 $\mu\mu\mu\tau$ 2017 result shown in figure 8.9.
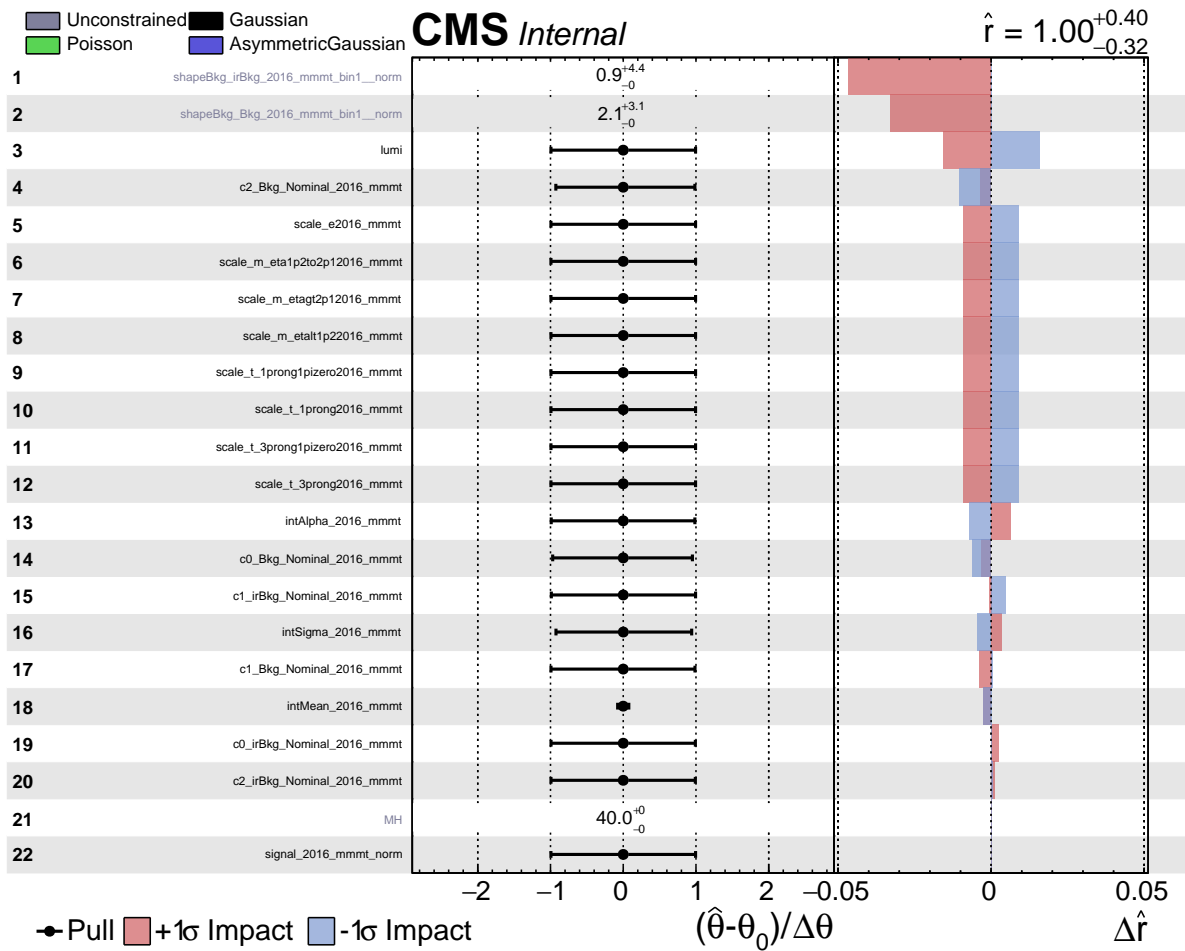
Figure C.1: Expected systematic impacts for the fit model for 2016 $\mu\mu\mu\tau$

Unconstrained Gaussian
Poisson AsymmetricGaussian

CMS *Internal*

$\hat{r} = 1.0^{+0.8}_{-0.5}$

1	shapeBkg_Bkg_2016_mmet_bin1__norm	$0.2^{+1.1}_{-0}$
2	scale_e2016_mmet	
3	scale_t_1prong1pizero2016_mmet	
4	lumi	
5	scale_m_eta1p2to2p12016_mmet	
6	scale_m_etagt2p12016_mmet	
7	scale_m_etalt1p22016_mmet	
8	scale_t_1prong2016_mmet	
9	scale_t_3prong1pizero2016_mmet	
10	scale_t_3prong2016_mmet	
11	intMean_2016_mmet	
12	intAlpha_2016_mmet	
13	intSigma_2016_mmet	
14	shapeBkg_irBkg_2016_mmet_bin1__norm	0.1^{+0}_{-0}
15	c1_irBkg_Nominal_2016_mmet	
16	c0_irBkg_Nominal_2016_mmet	
17	MH	40.0^{+0}_{-0}
18	c0_Bkg_Nominal_2016_mmet	
19	c1_Bkg_Nominal_2016_mmet	
20	signal_2016_mmet_norm	

$-2 \quad -1 \quad 0 \quad 1 \quad 2 \qquad -0.05 \quad 0 \quad 0.05$

Pull +1σ Impact −1σ Impact

$(\hat{\theta}-\theta_0)/\Delta\theta$ $\qquad\qquad \Delta\hat{r}$

Figure C.2: Expected systematic impacts for the fit model for 2016 $\mu\mu e\tau$
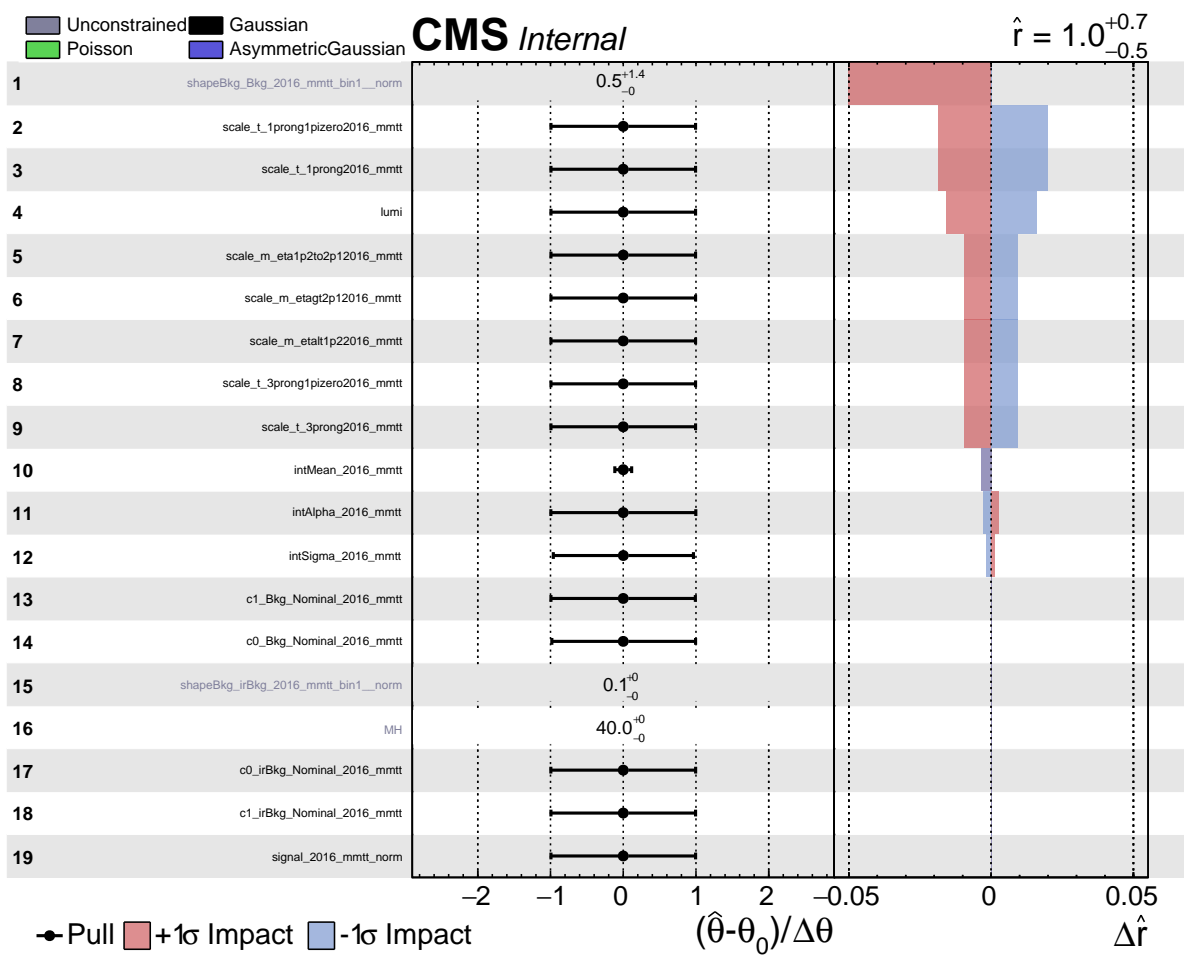
Figure C.3: Expected systematic impacts for the fit model for 2016 $\mu\mu\tau\tau$
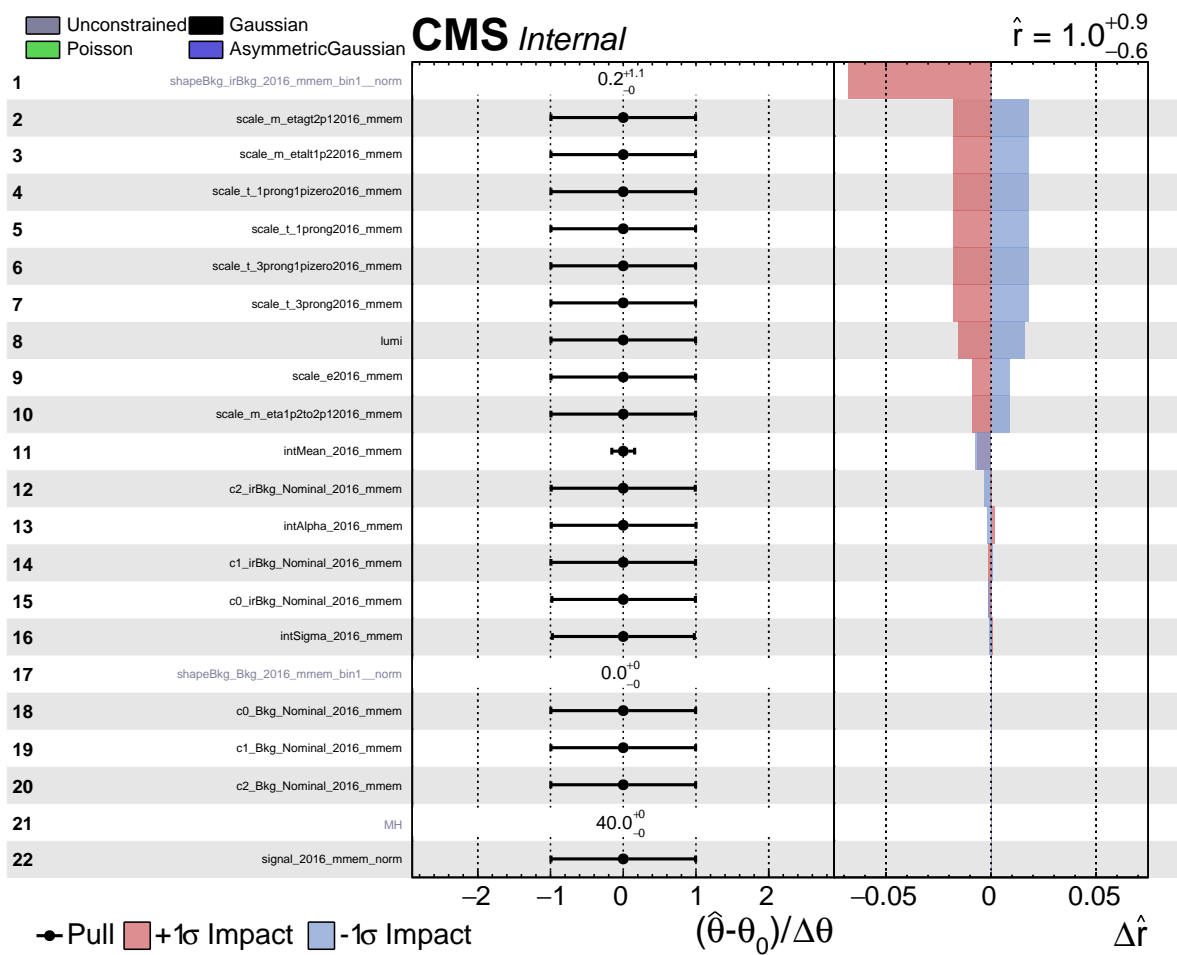
Figure C.4: Expected systematic impacts for the fit model for 2016 $\mu\mu e\mu$
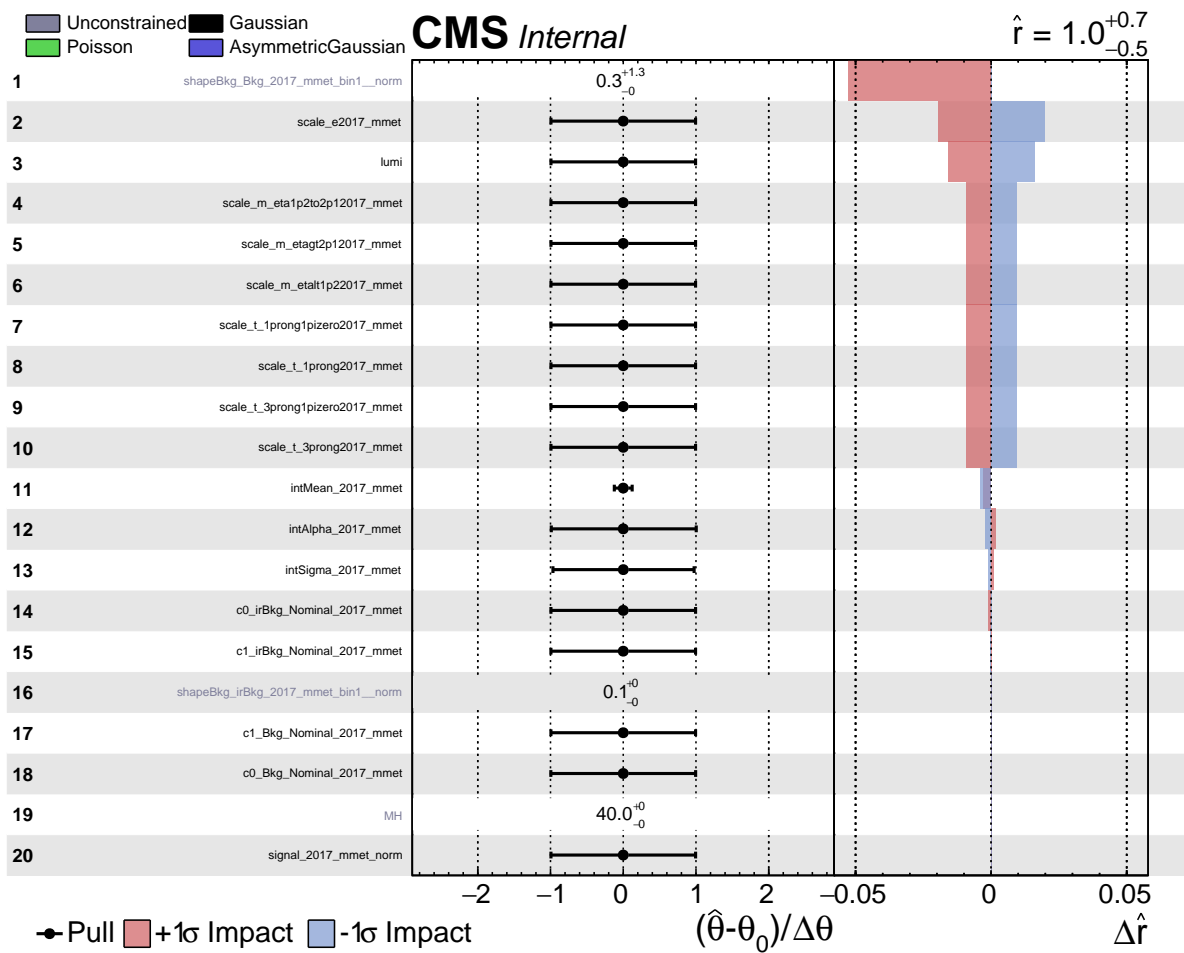
Figure C.5: Expected systematic impacts for the fit model for 2017 $\mu\mu e\tau$
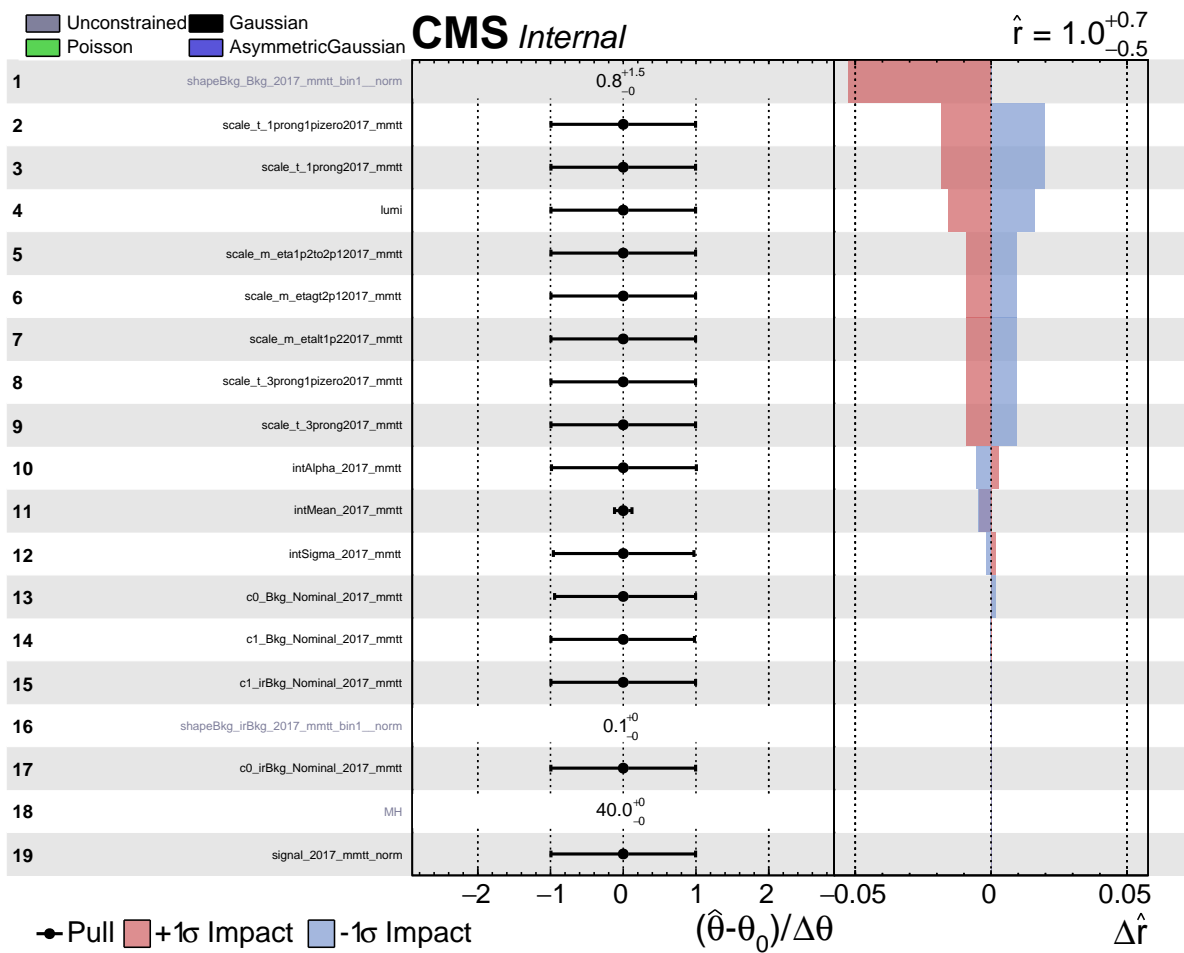
Figure C.6: Expected systematic impacts for the fit model for 2017 $\mu\mu\tau\tau$
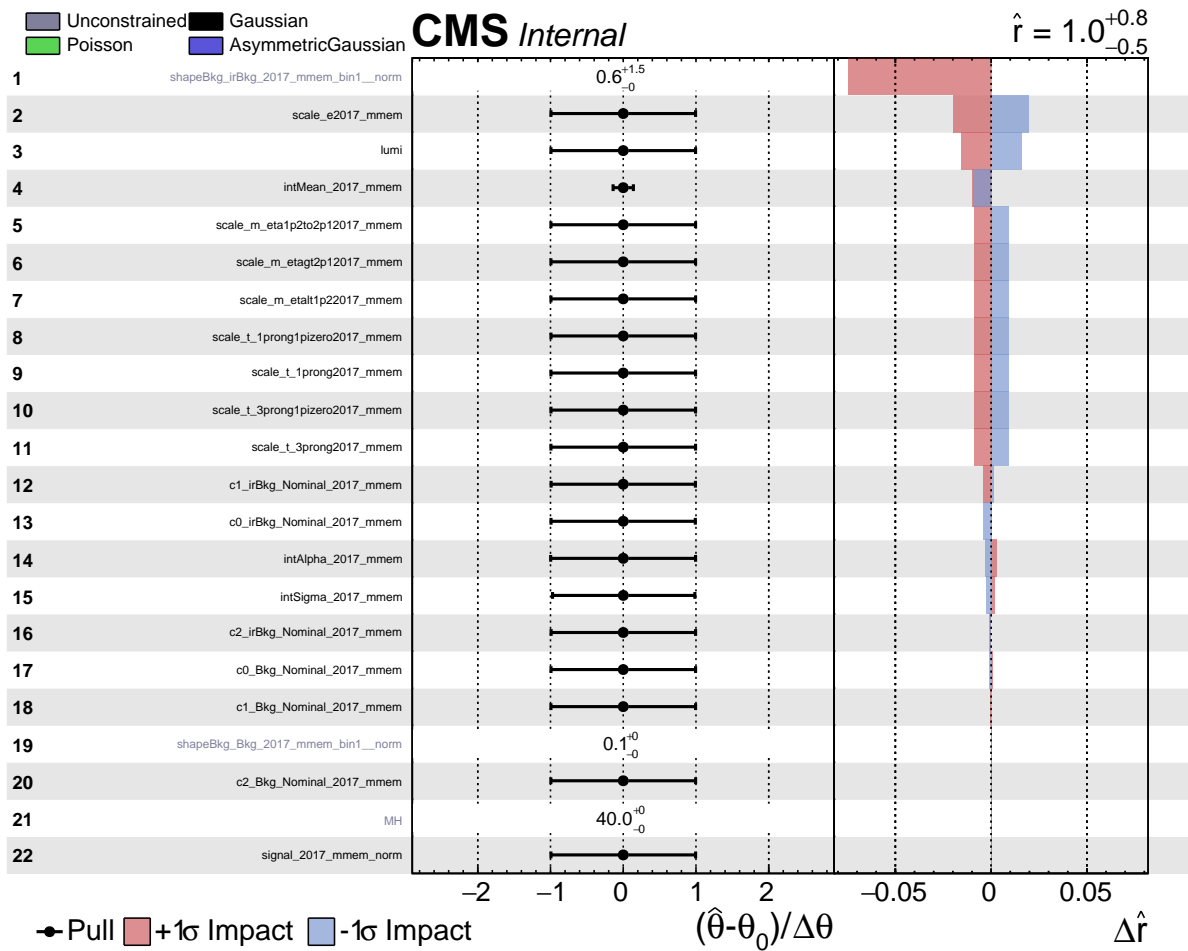
Figure C.7: Expected systematic impacts for the fit model for 2017 $\mu\mu e\mu$
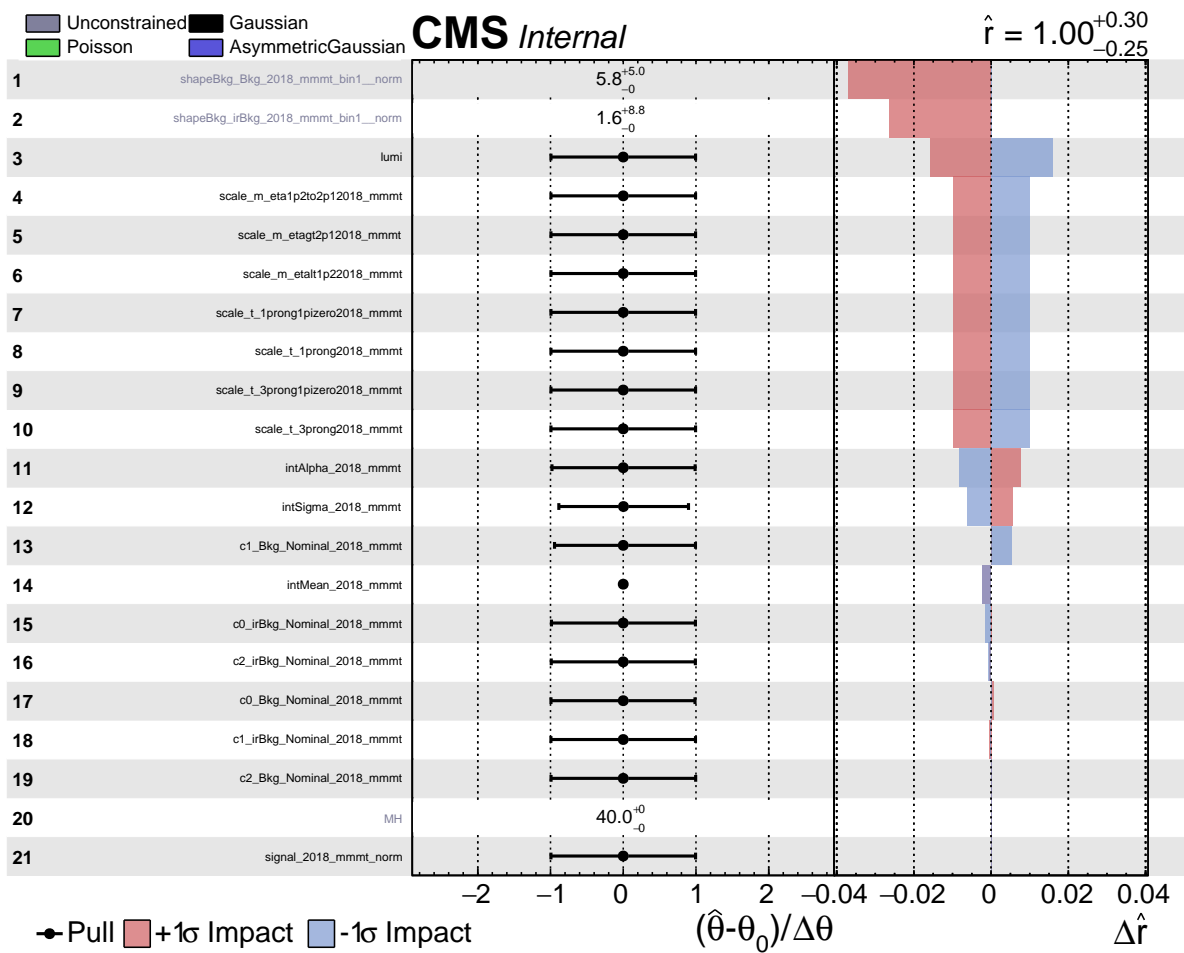
Figure C.8: Expected systematic impacts for the fit model for 2018 $\mu\mu\mu\tau$
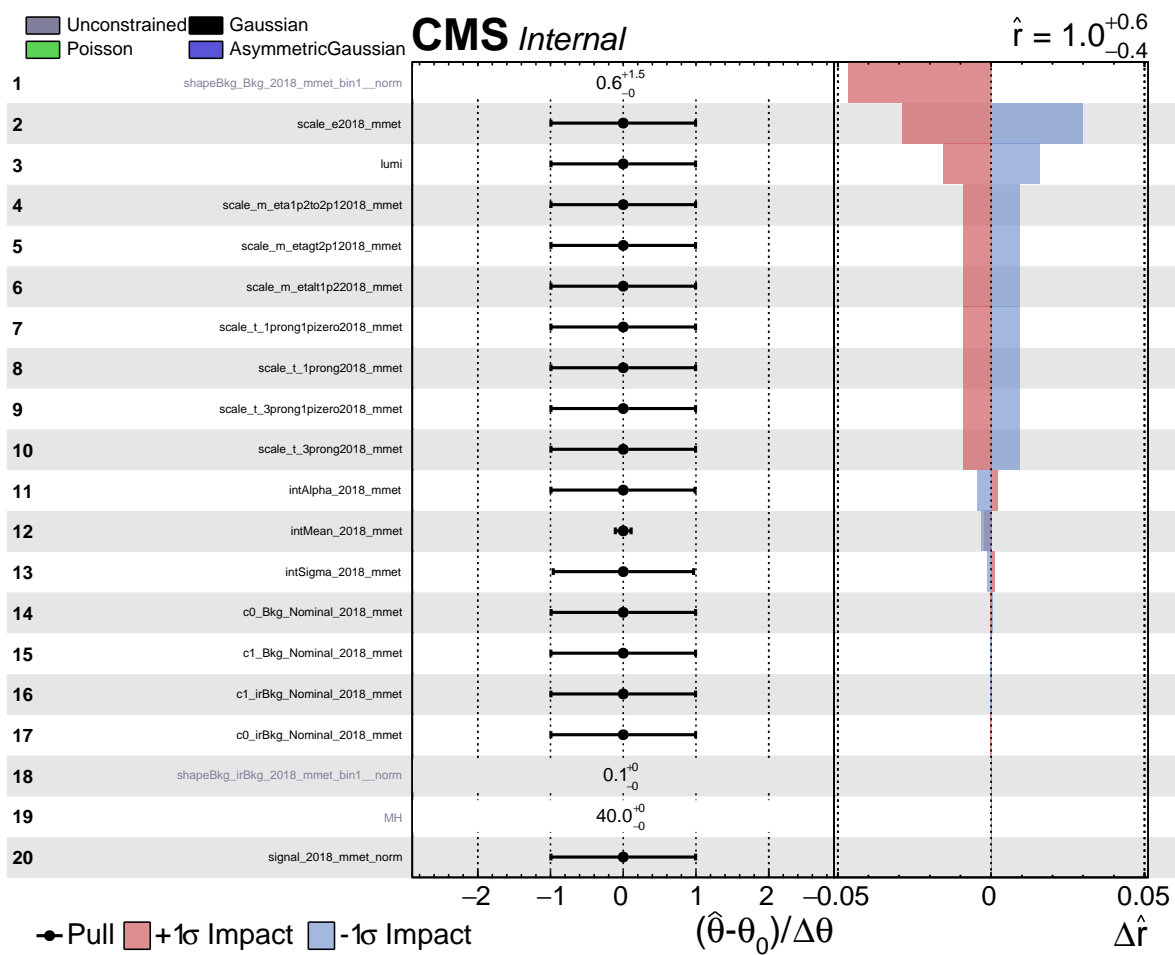
Figure C.9: Expected systematic impacts for the fit model for 2018 $\mu\mu e\tau$
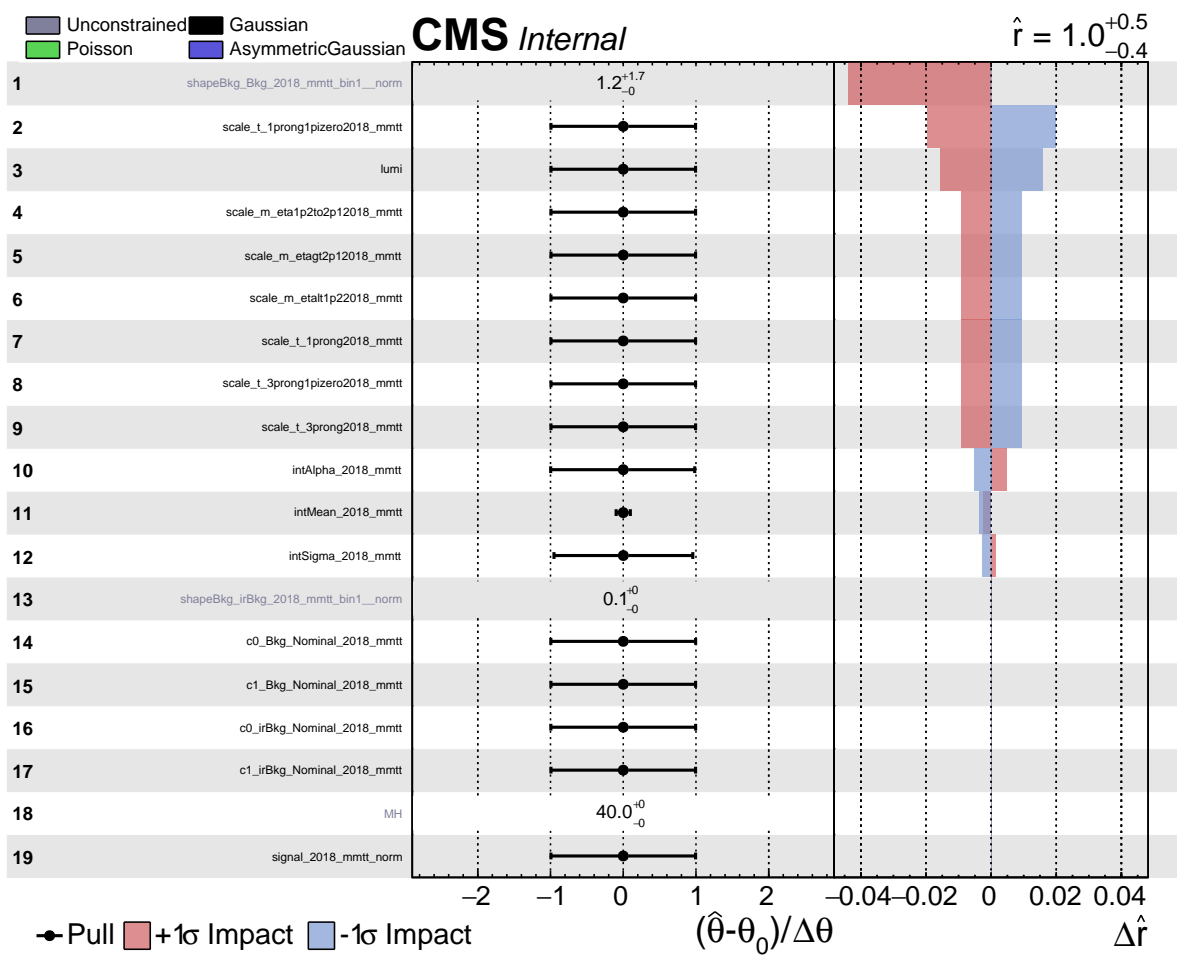
Figure C.10: Expected systematic impacts for the fit model for 2018 $\mu\mu\tau\tau$
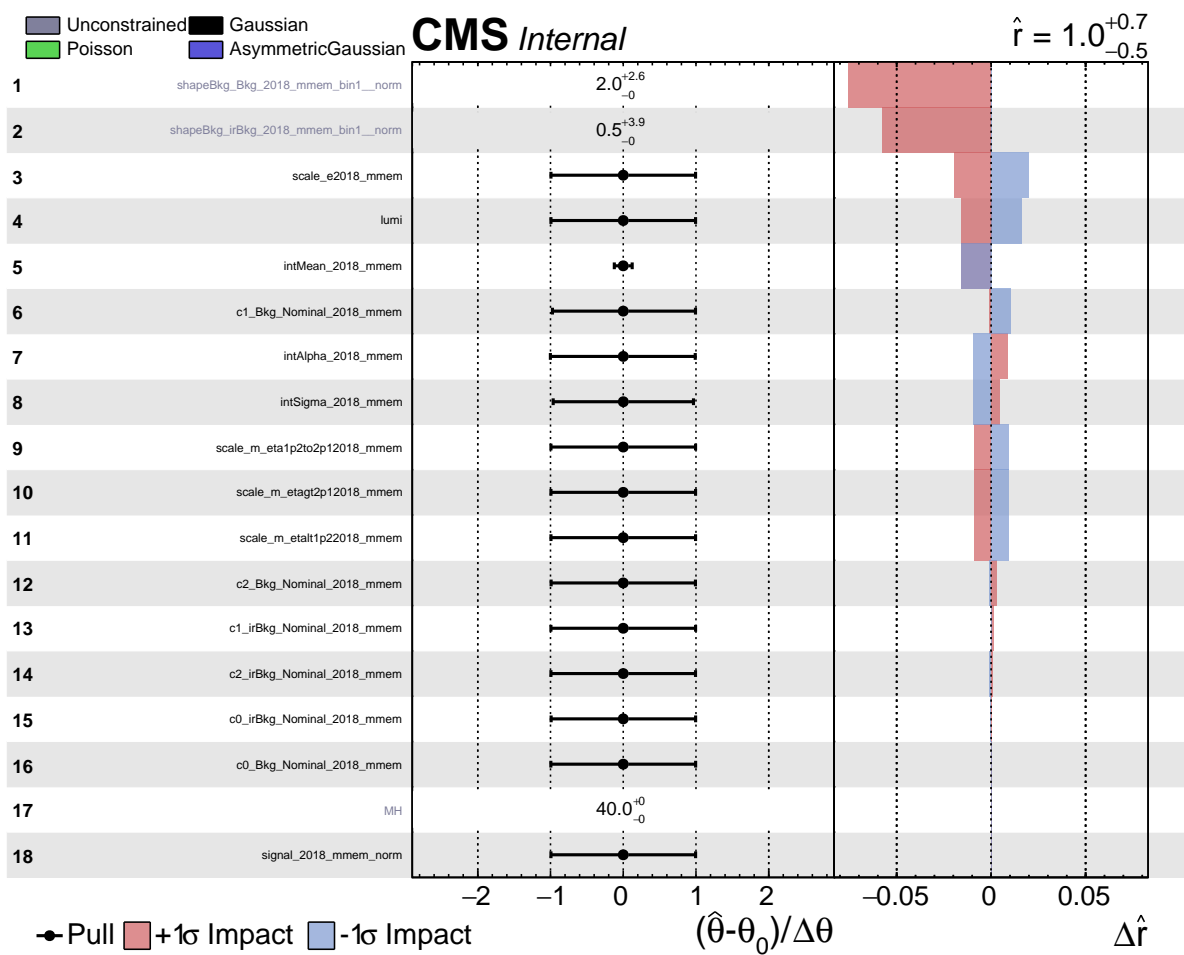
Figure C.11: Expected systematic impacts for the fit model for 2018 $\mu\mu e\mu$

Appendix D

Fit Models

This section contains the remaining parametric fit models for the rest of the years 2017, 2018 and the channels referenced originally in chapter 8.
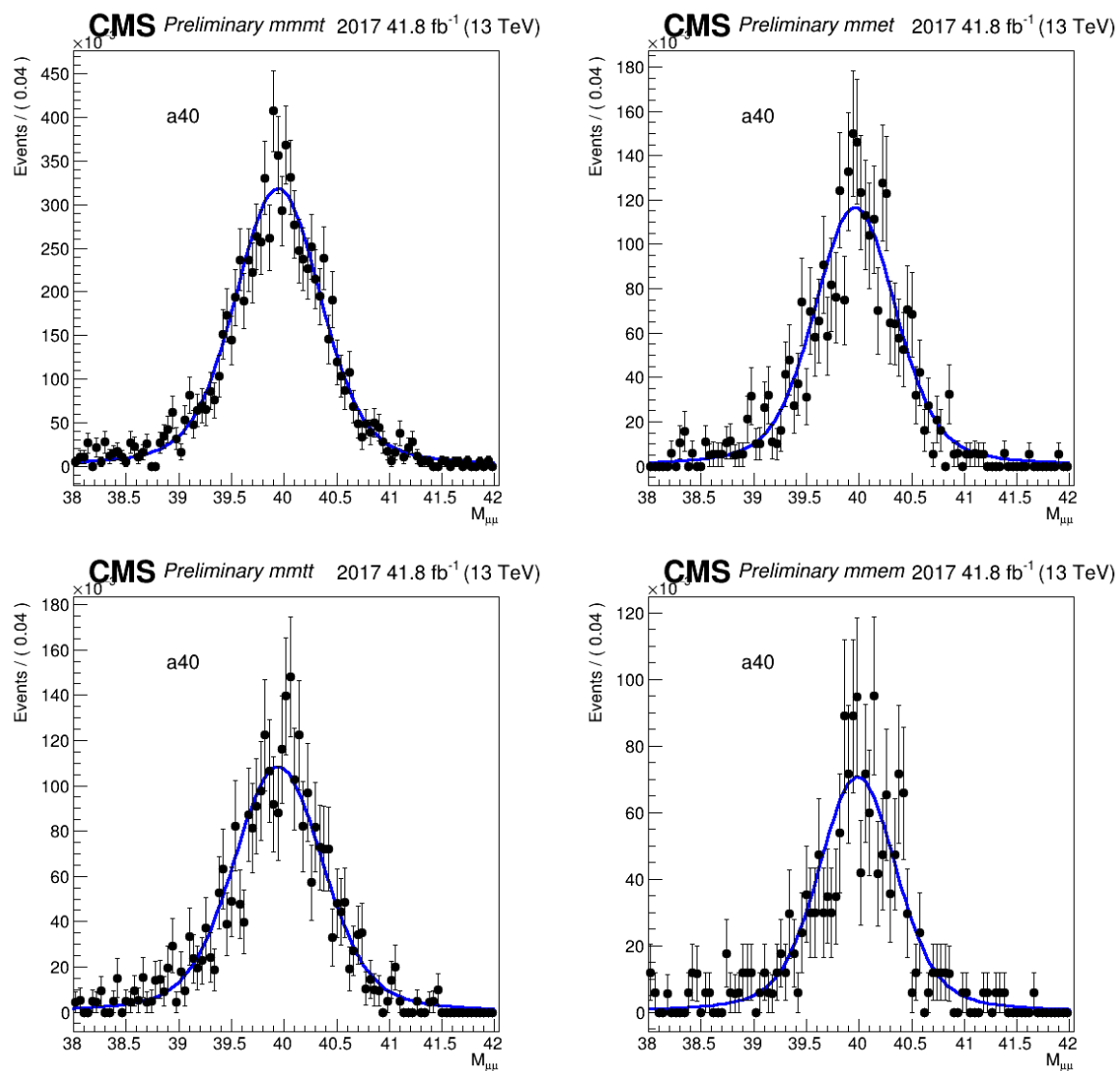
Figure D.1:   2017 Signal fit using a Voigtian function
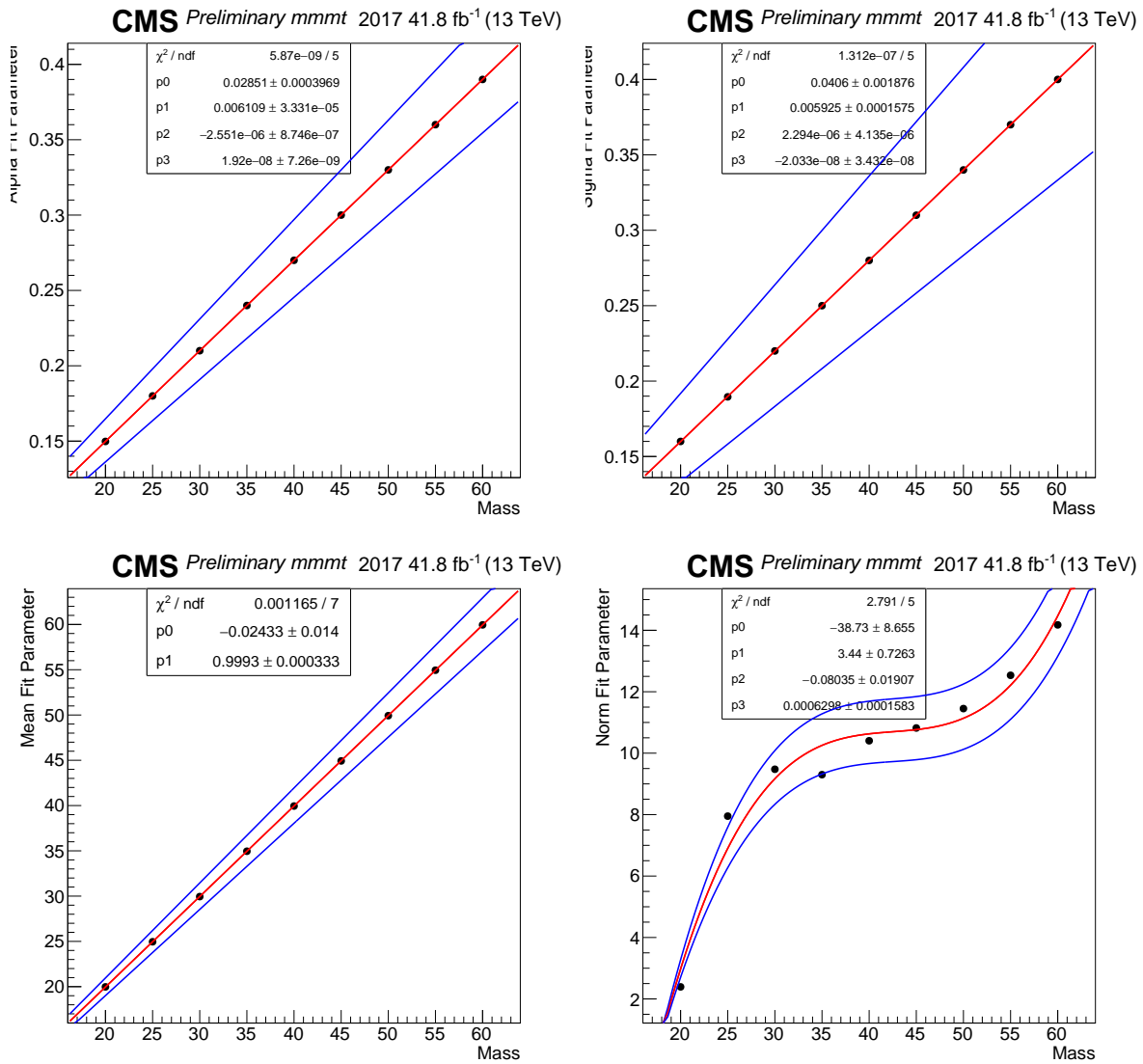
Figure D.2: Spline functions for 2017 mmmt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
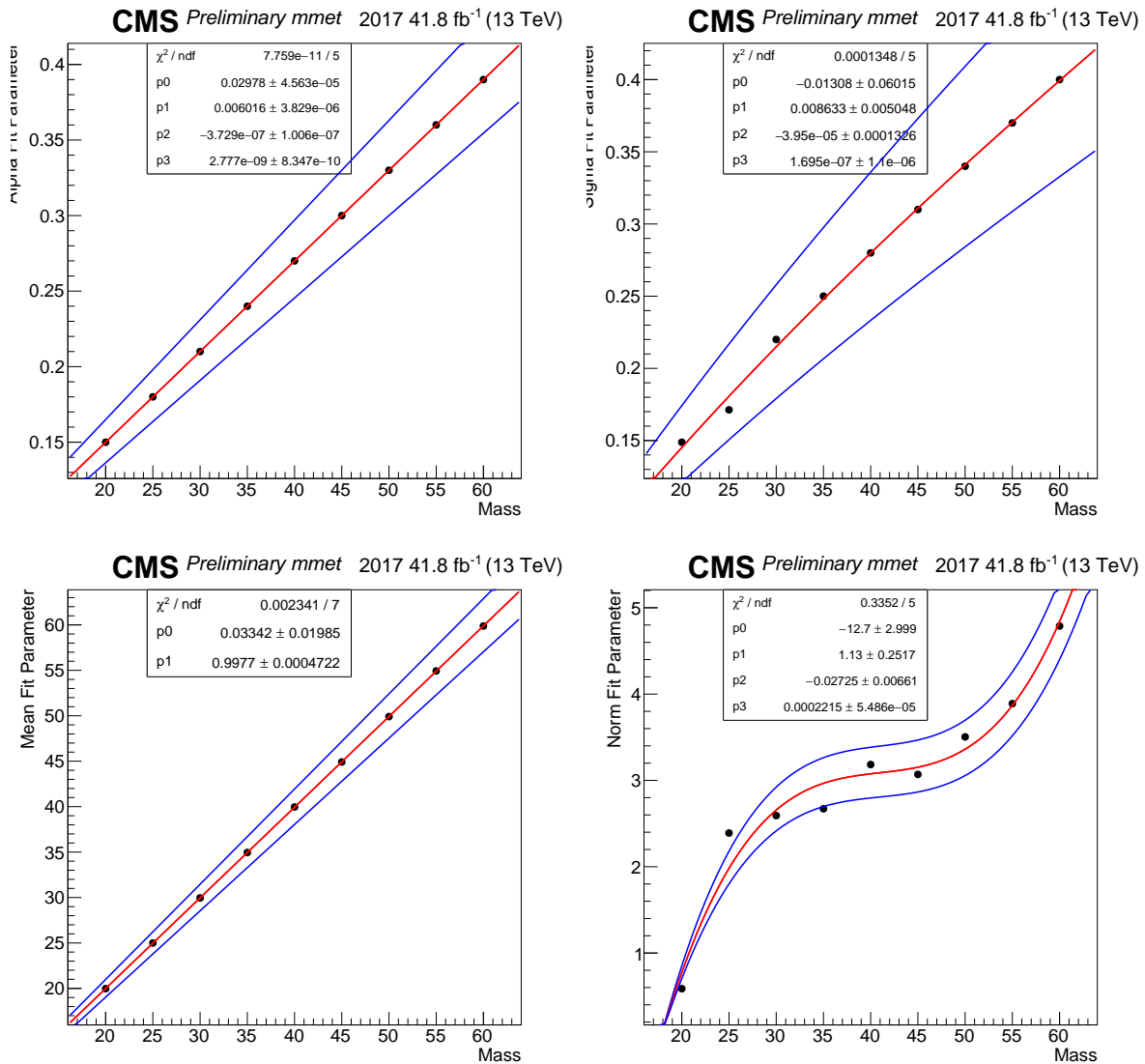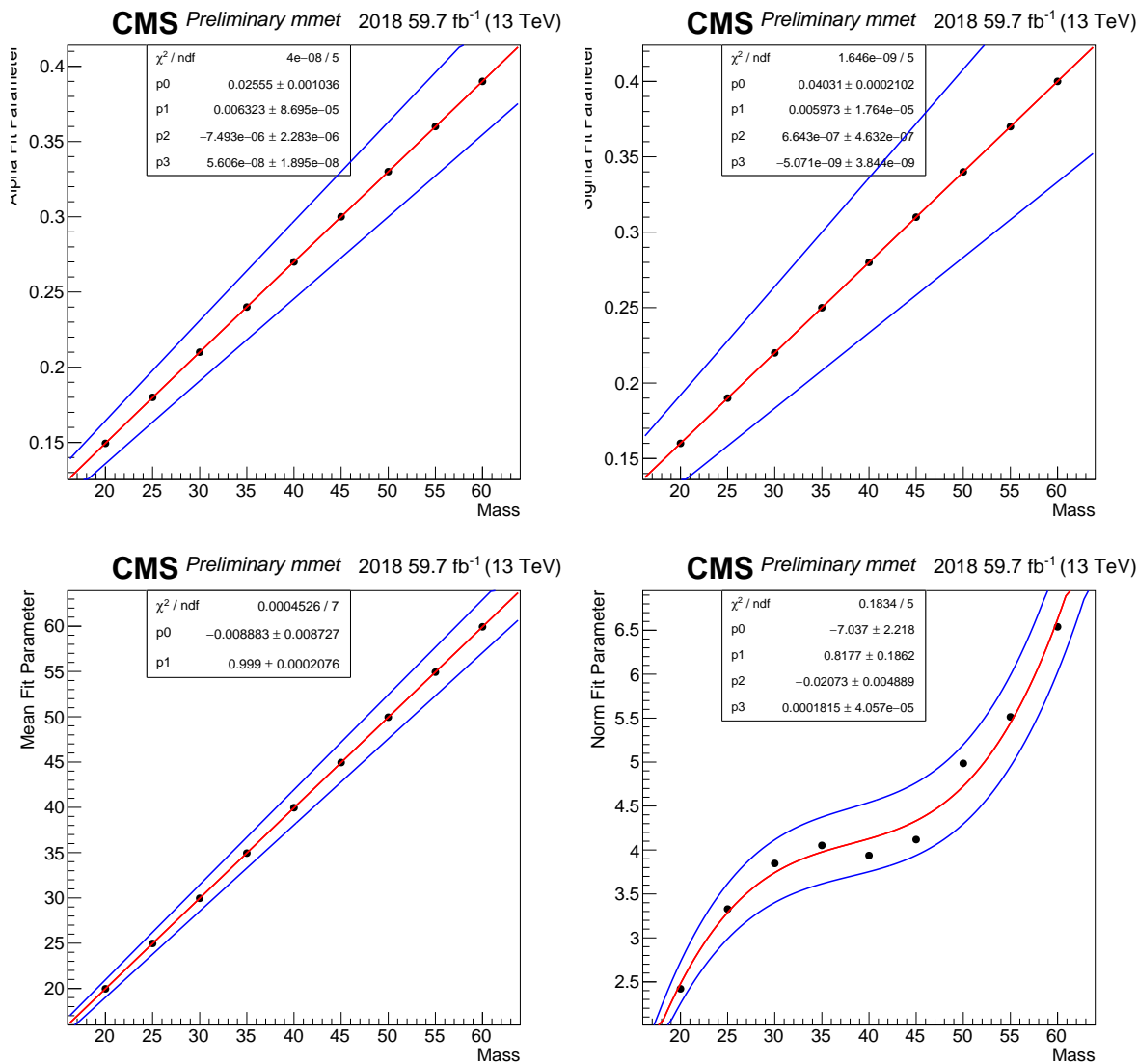
Figure D.3:   Spline functions for 2017 mmet a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
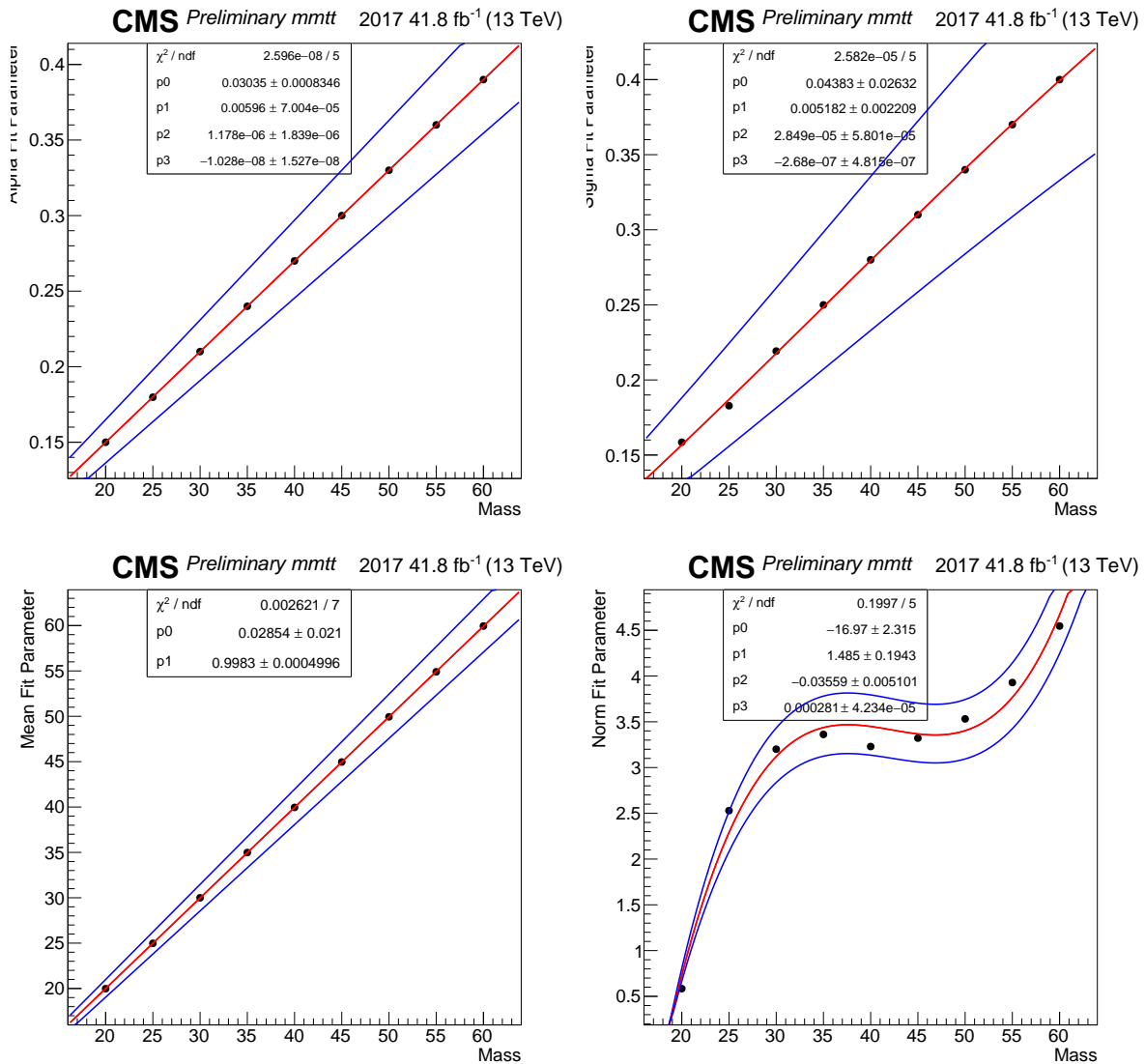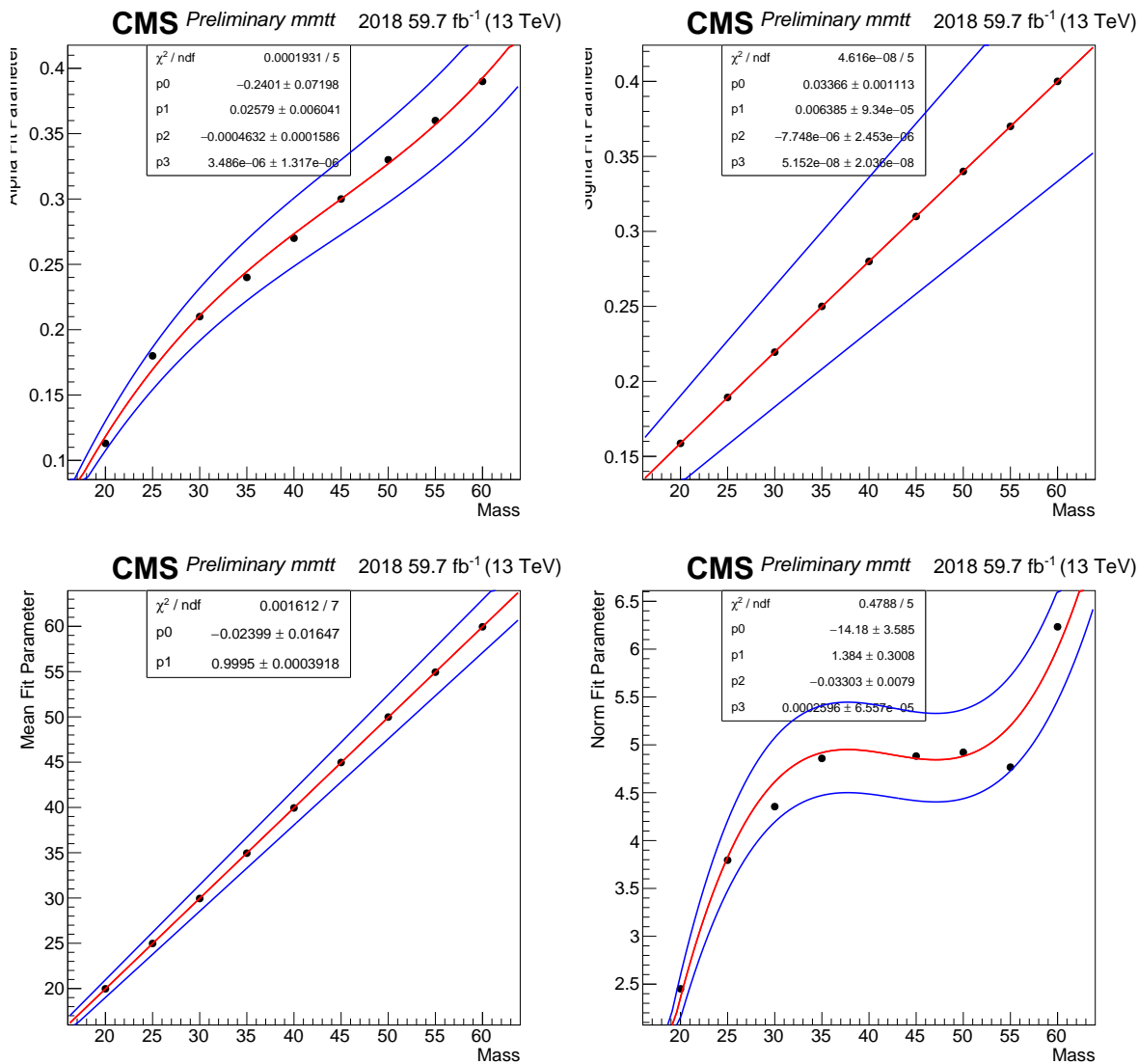
Figure D.4: Spline functions for 2017 mmtt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
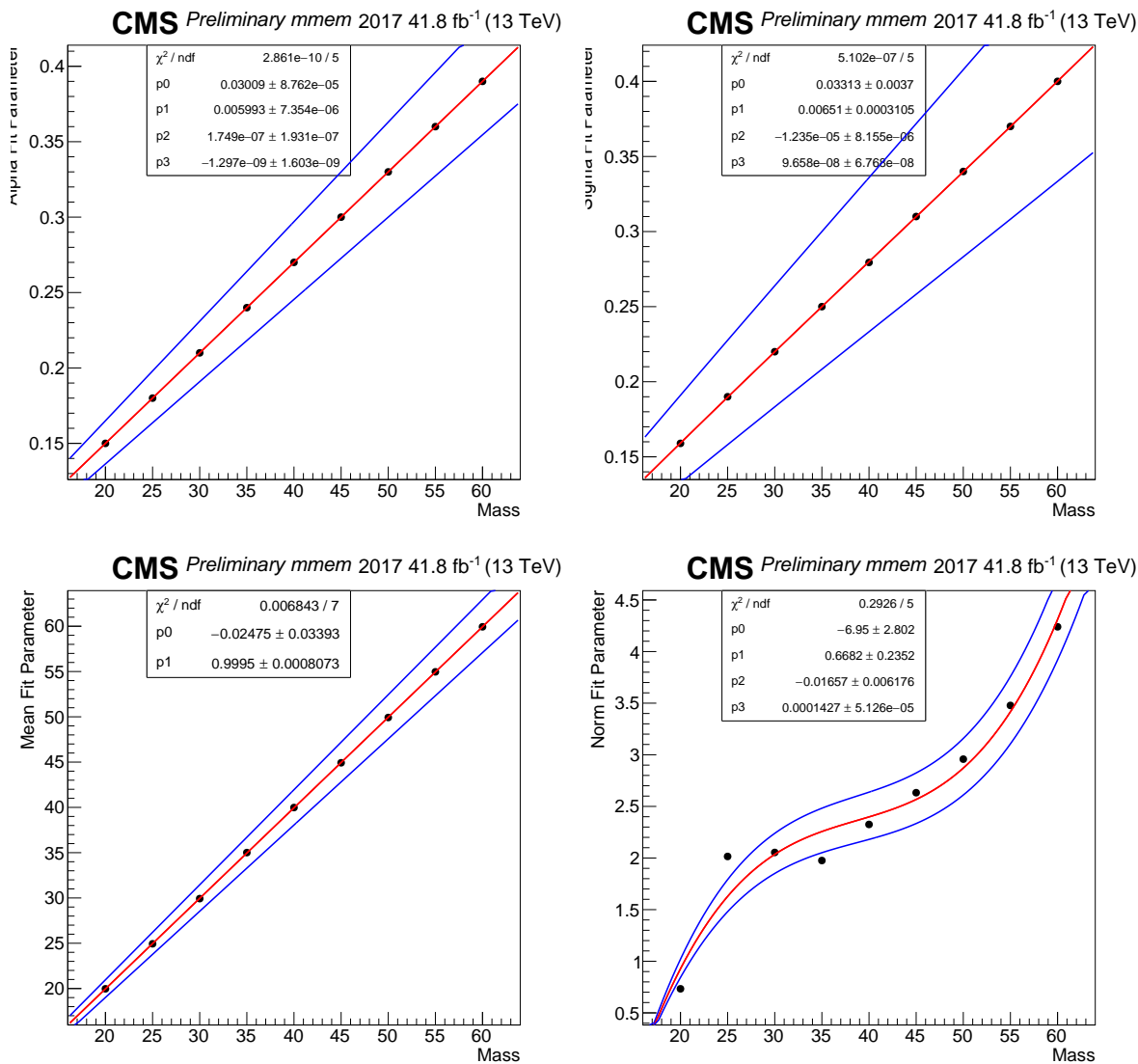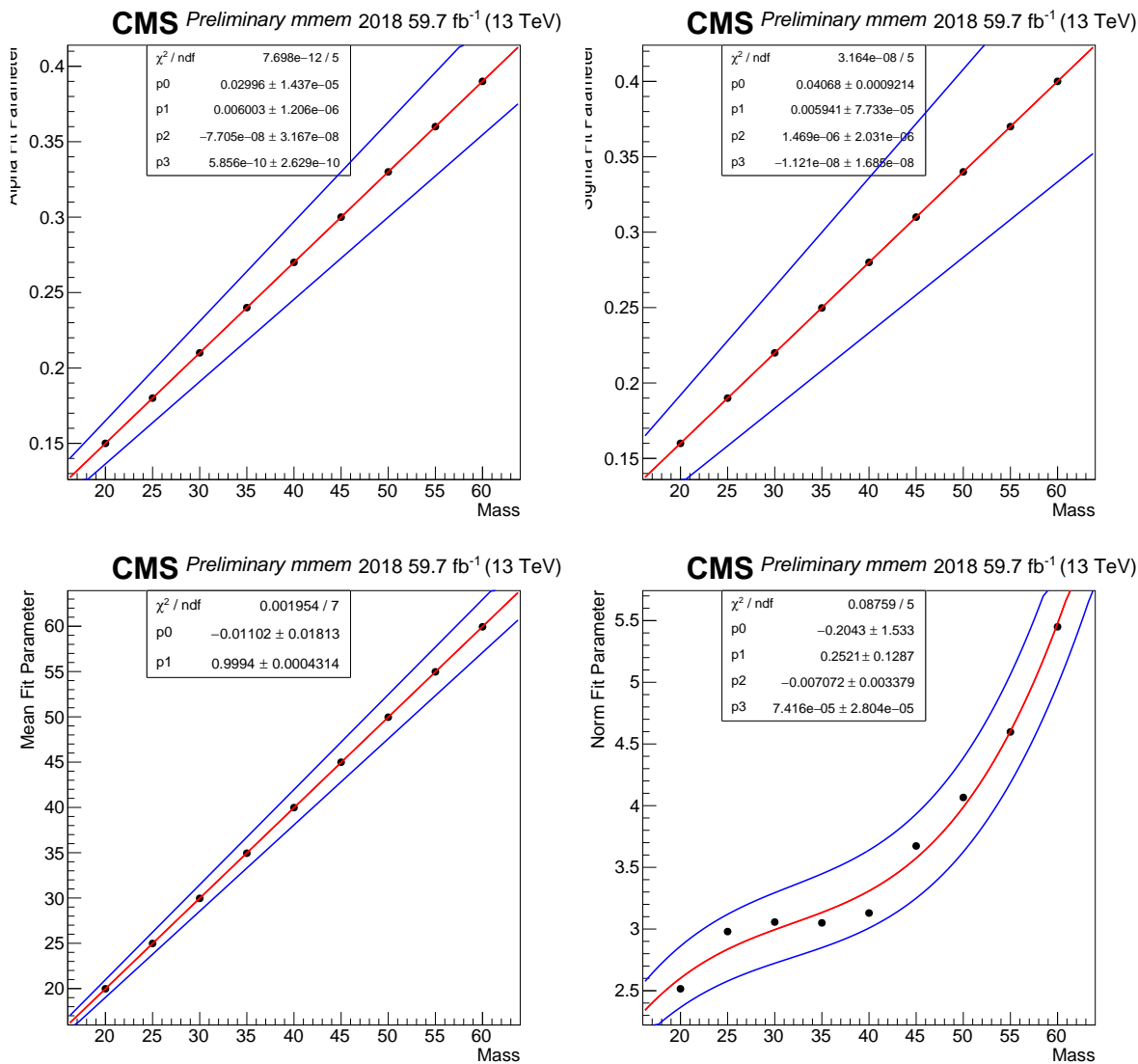
Figure D.5: Spline functions for 2017 mmem a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
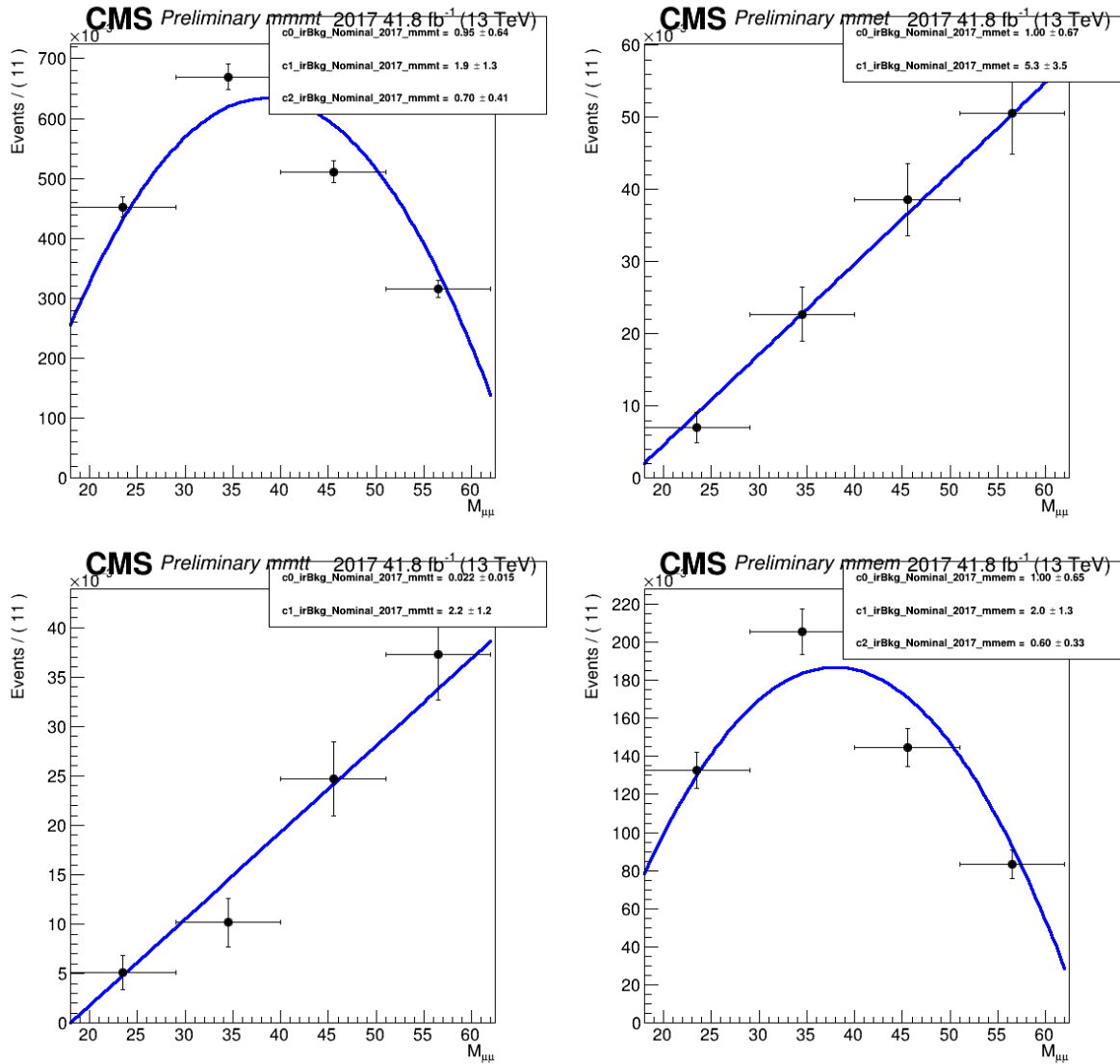
Figure D.6:   2017 irreducible background fit using Bernstein polynomials
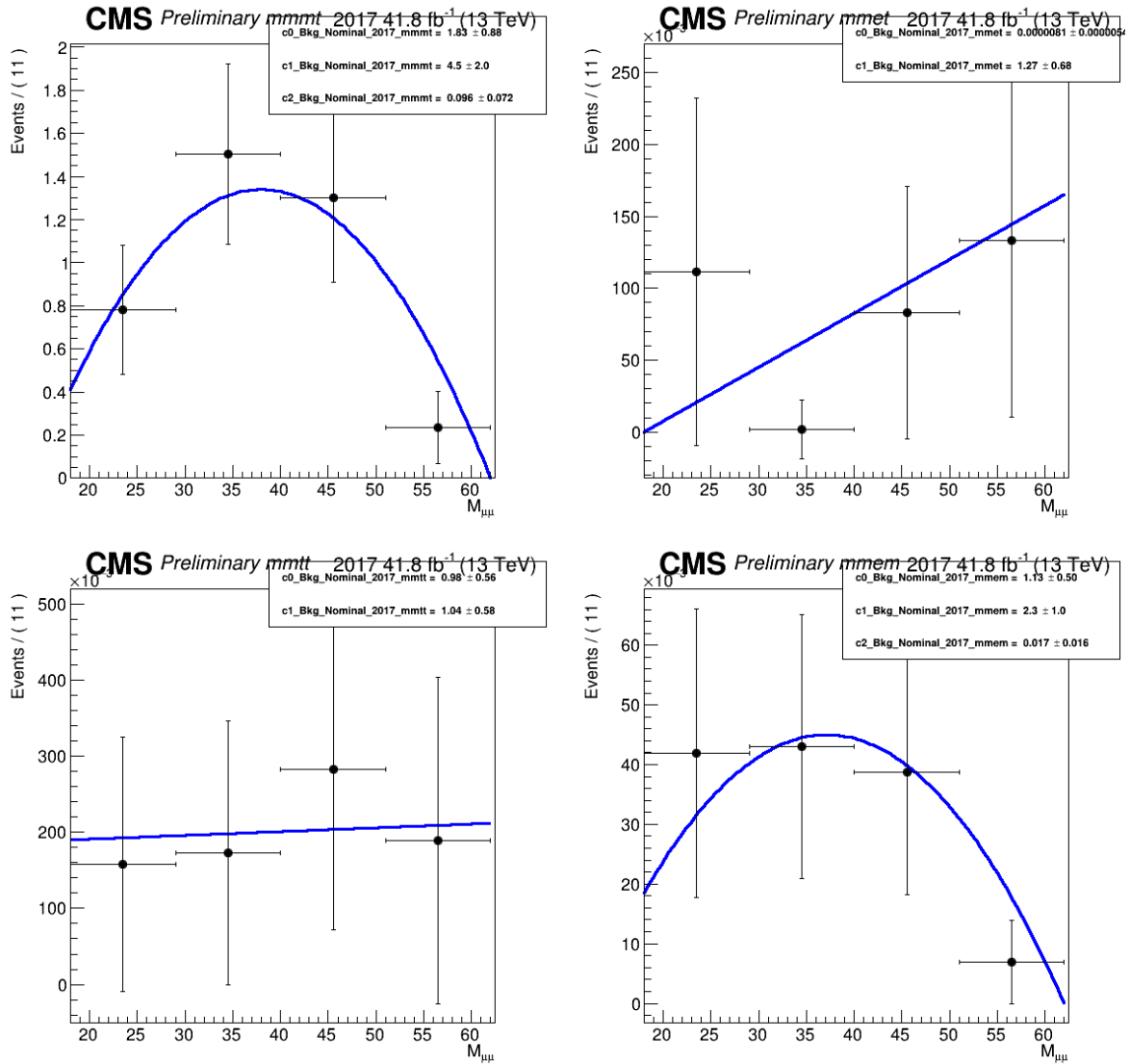
Figure D.7: 2017 reducible background fit using Bernstein polynomials
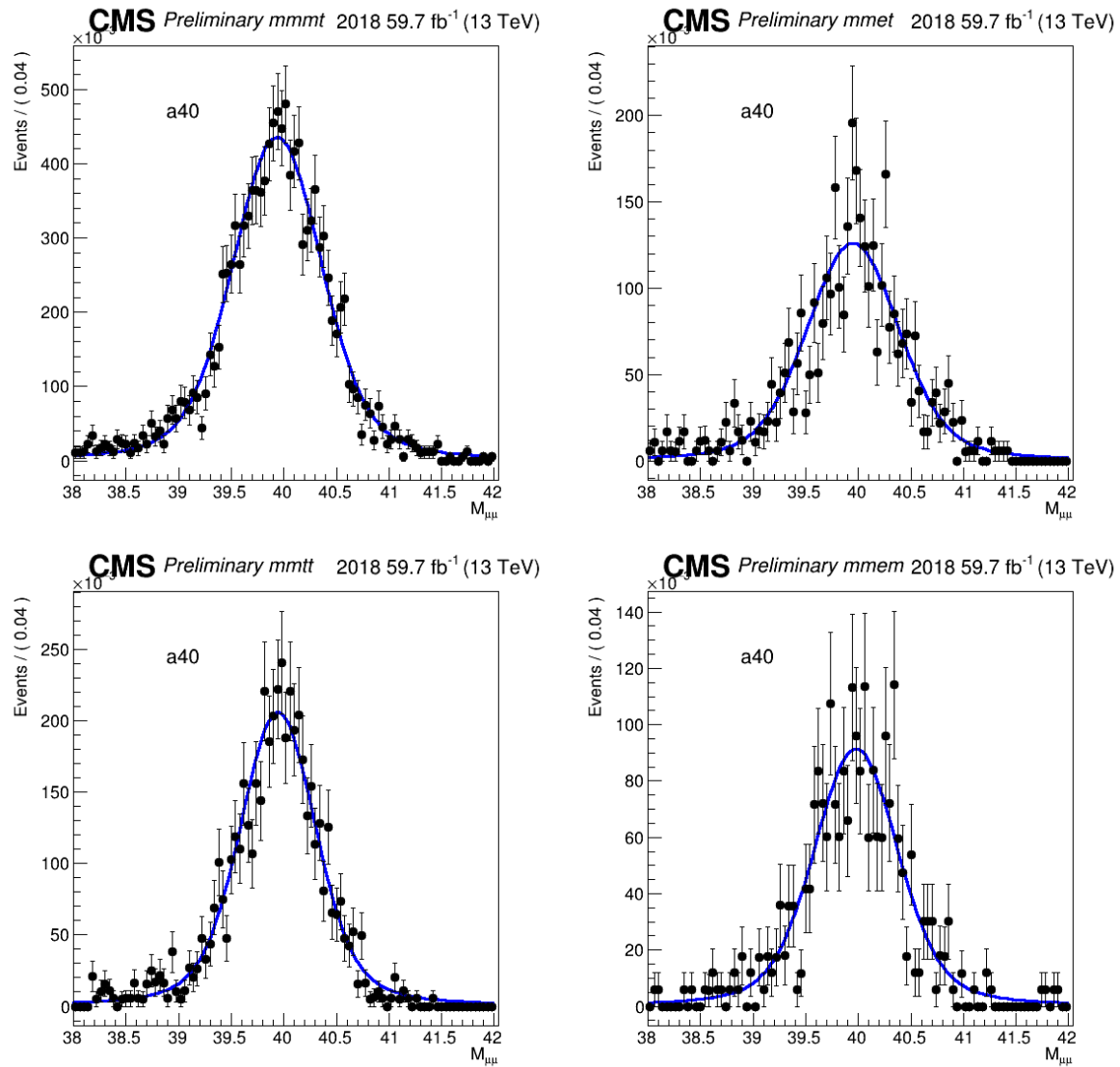
Figure D.8: 2018 Signal fit using a Voigtian function
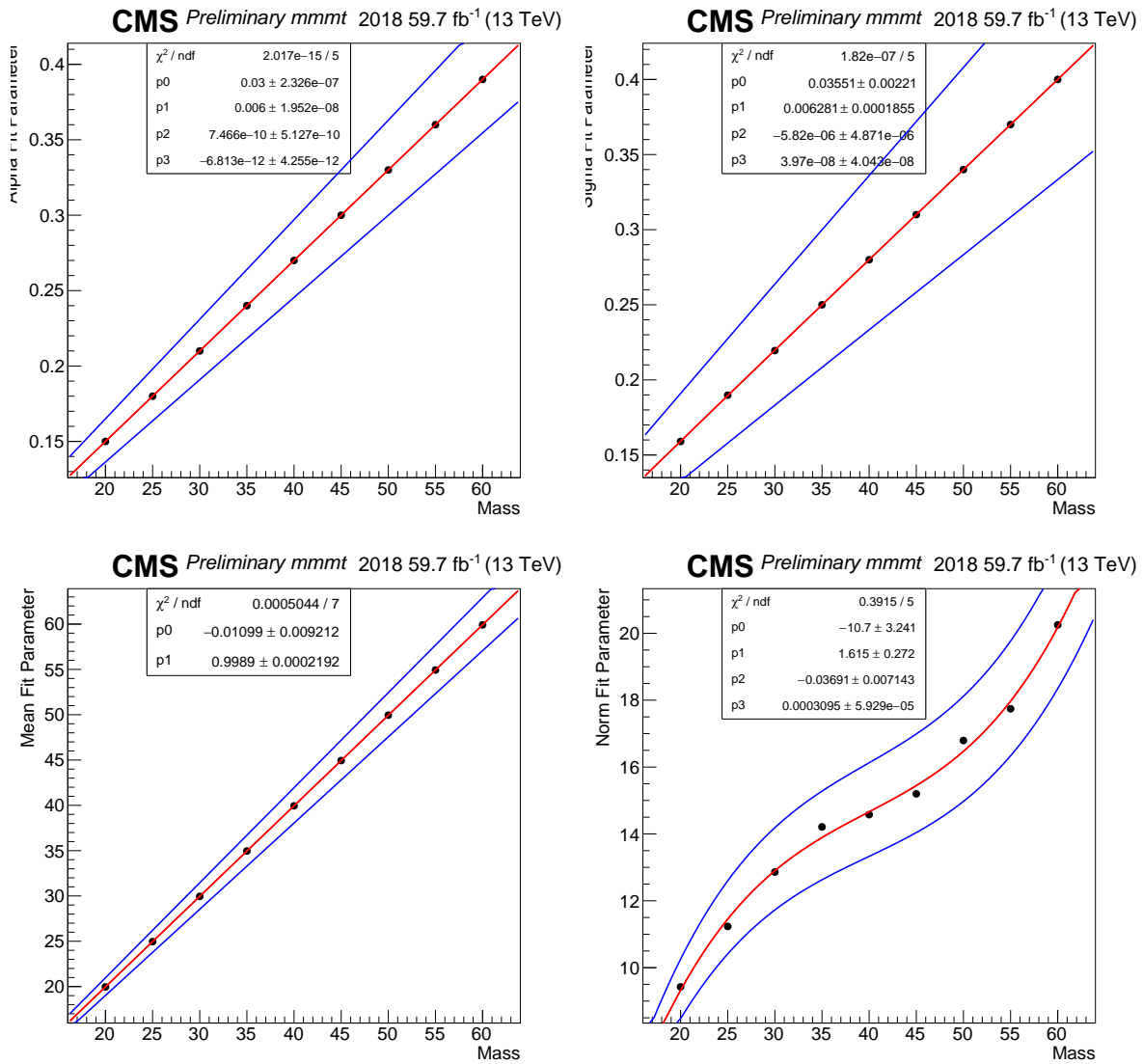
Figure D.9: Spline functions for 2018 mmmt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean

Figure D.10: Spline functions for 2018 mmet a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean
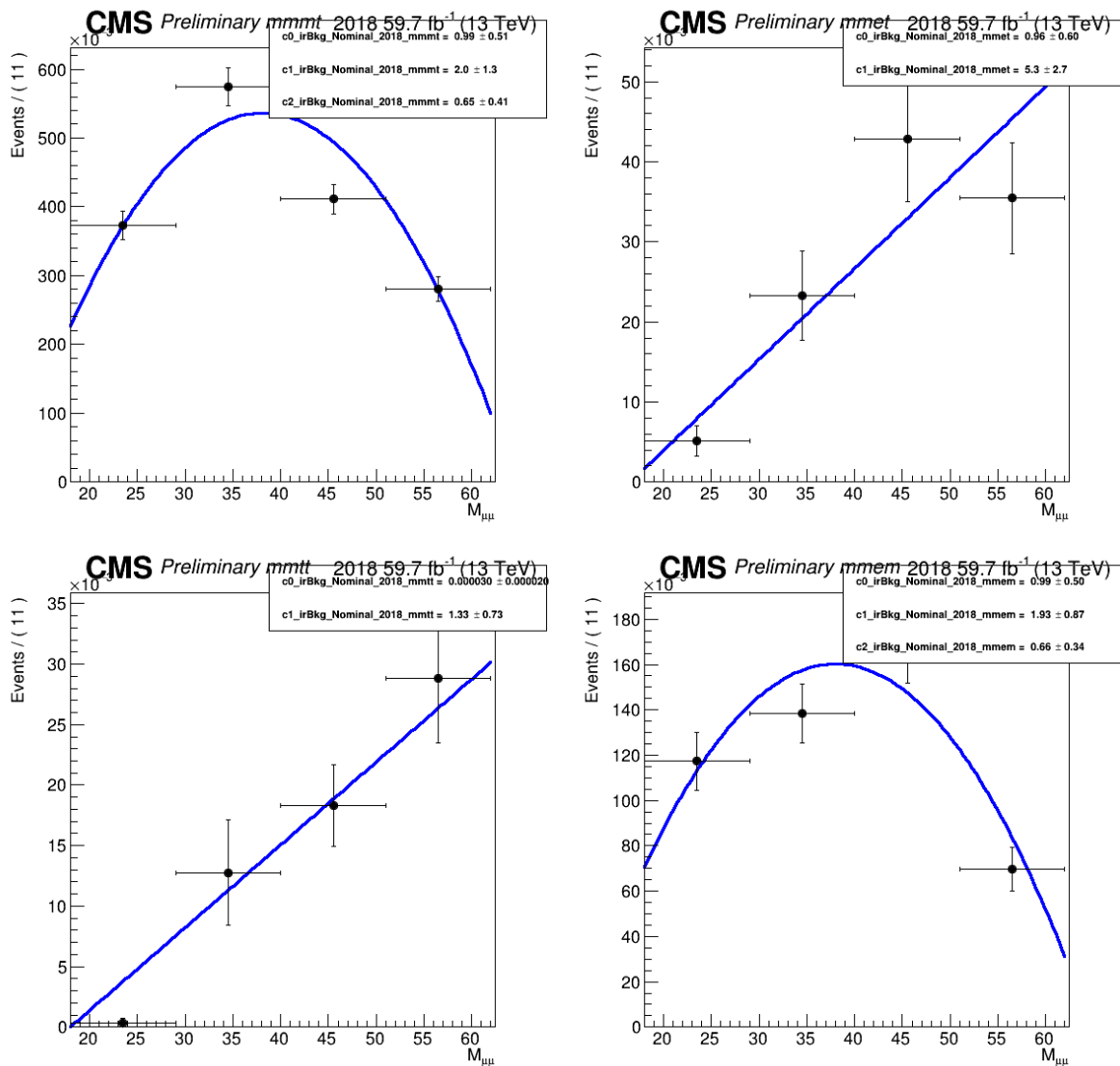
Figure D.11: Spline functions for 2018 mmtt a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean

Figure D.12: Spline functions for 2018 mmem a 3rd order polynomial is used for for Alpha, Sigma, and Normalization, a 1st order polynomial is used for the Mean

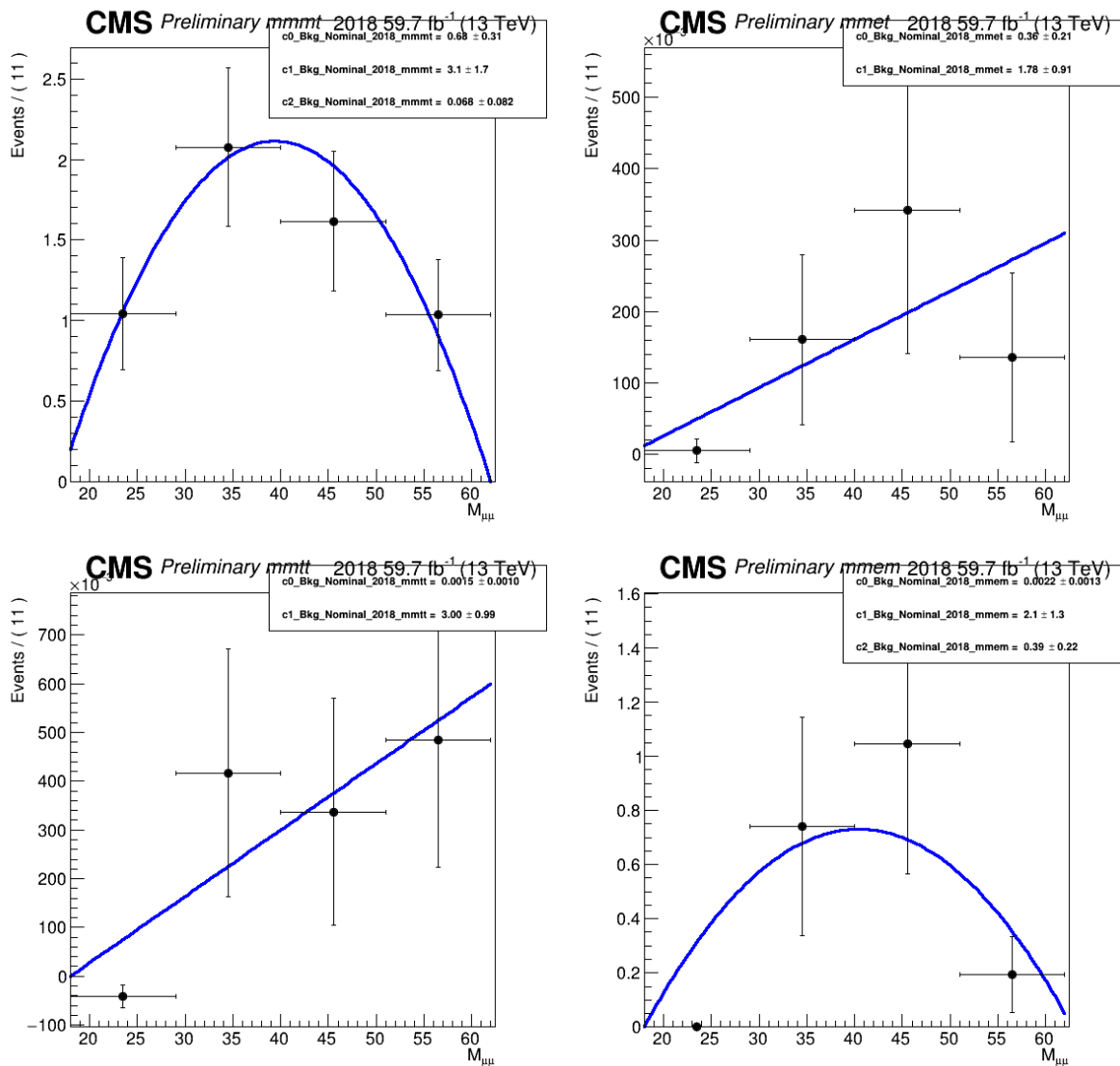Figure D.13: 2018 irreducible background fit using Bernstein polynomials

Figure D.14: 2018 reducible background fit using Bernstein polynomials

References

[1] E. Noether, "Invariant variation problems," *Transport Theory and Statistical Physics*, vol. 1, pp. 186–207, jan 1971.

[2] C. G. Tully, *Elementary particle physics in a nutshell.* Princeton, NJ: Princeton Univ. Press, 2011.

[3] T. Lancaster and S. J. Blundell, *Quantum field theory for the gifted amateur.* Oxford: Oxford University Press, Apr 2014.

[4] 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.

[5] 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.

[6] 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 τ leptons in proton-proton collisions at $\sqrt{s} = 13$tev leptons with the CMS detector," *Journal of High Energy Physics*, p. 283, 2017.

[7] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report.* CERN Yellow Reports: Monographs, Geneva: CERN, 2004.

[8] M. B. Harriet Jarlett, "25 years of large hadron collider experimental programme," *CERN News Topic: At CERN*, 2017.

[9] S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST*, vol. 03, p. S08004, 2008.

[10] T. Eichhorn, "Development of Silicon Detectors for the High Luminosity LHC," 2015. presented 8 Jul 2015.

[11] S. Chatrchyan, G. Hmayakyan, V. Khachatryan, A. M. Sirunyan, R. Adolphi, G. Anagnostou, S. Brauer, ... Hansel, T. Nobauer, M. Droge, U. Roser, A. Schoning, 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.

[12] G. Abbiendi, N. Amapane, C. Battilana, C. Fernandez Bedoya, R. Bellan, M. Bellato, A. Benvenuti, P. Biallas, S. Bolognesi, S. Braibant, F. Cavallo, R. Carlin, G. Cerminara, M. Chamizo, E. Conti, G. Dallavalle, D. Dattola, G. Dellacasa, M. Fouz, J. Frangenheim, P. Giacomelli, M. Giunta, F. Gonella, A. Gresele, L. Guiducci, K. Hoepfner, E. Jacobi, S. Marcellini, C. Mariotti, S. Maselli, G. Masetti, A. Meneguzzo, G. Mila, V. Monaco, A. Montanari, F. Montecassiano, F. Navarria, M. Nervo, F. Odorici, M. Passaseo, A. Parenti, A. Perrotta, J. Puerta, H. Reithler, P. Ronchese, T. Rovelli, R. Sacchi, G. Siroli, M. Sowa, A. Staiano, D. Teyssier, E. Torassa, R. Travaglini, A. Triossi, O. Tsigenov, S. Vanini, S. Ventura, C. Villanueva, M. Zanetti, and P. Zotto, "The cms muon barrel drift tubes system commissioning," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 598, no. 1, pp. 192–195, 2009. Instrumentation for Collding Beam Physics.

[13] C. Foudas, "The CMS Level-1 Trigger at LHC and Super-LHC. The CMS Level-1 Trigger at LHC and Super-LHC," tech. rep., 2008.

[14] L. Thomas, "CMS High Level Trigger performance at 13 TeV," *PoS*, vol. ICHEP2018, p. 226. 4 p, 2019.

[15] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector," CMS Physics Analysis Summary CMS-PAS-PFT-10-001, CERN, 2010.

[16] D. Hufnagel, "The architecture and operation of the CMS Tier-0," tech. rep., CERN, Geneva, Nov 2010.

[17] J. Knolle, "Precision luminosity measurement with proton-proton collisions at the CMS experiment in Run 2," tech. rep., CERN, Geneva, Oct 2020.

[18] CMS Collaboration, "Particle–flow event reconstruction in CMS and performance for jets, taus, and E_T^{miss}," CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.

[19] I. Kratschmer, "Muon reconstruction and identification in CMS Run I and towards Run II," tech. rep., CERN, Geneva, Oct 2014.

[20] 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.

[21] R. Workman $et\ al.$, "Review of Particle Physics," to be published (2022).

[22] A. Sirunyan, A. Tumasyan, $et\ al.$, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} =$13 TeV," $Journal\ of\ Instrumentation$, vol. 13, pp. P10005–P10005, oct 2018.

[23] A. M. Sirunyan $et\ al.$, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV," $JINST$, vol. 13, p. P10005, 2018.

[24] M. H. Hassanshahi, "Tau identification in CMS during LHC Run 2," tech. rep., CERN, Geneva, Nov 2021.

[25] CMS Collaboration, "Identification of hadronic tau lepton decays using a deep neural network," 2022.

[26] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," $JHEP$, vol. 0605, p. 026, 2006.

[27] S. Agostinelli $et\ al.$, "GEANT4: A simulation toolkit," $Nucl.\ Instrum.\ Meth.\ A$, vol. 506, p. 250, 2003.

[28] A. Sirunyan, A. Tumasyan, $et\ al.$, "Particle-flow reconstruction and global event description with the CMS detector," $Journal\ of\ Instrumentation$, vol. 12, pp. P10003–P10003, oct 2017.

[29] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov, "Extracting muon momentum scale corrections for hadron collider experiments," *The European Physical Journal C*, vol. 72, oct 2012.

[30] E. P. O. Group, "Egamma miniaod v2 apply energy scale 13 tev," August 2022. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaMiniAODV2.

[31] T. P. O. Group, "Tauid recommendations 13 tev," August 2022. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendation13TeV.

[32] CMS Collaboration, "Measurements of higgs boson production in the decay channel with a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ tev," 2022.

[33] E. P. O. Group, "Electron mva recommendations 13 tev," August 2022. https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2.

[34] CMS, "Search for the standard model higgs boson decaying to a pair of tau leptons in 2016 data," CMS Note 2016/355, 2016.

[35] "Measurement of Higgs boson production in the decay channel with a pair of τ leptons," tech. rep., CERN, Geneva, 2020.

[36] 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.

[37] M. Aaboud, G. Aad, B. Abbott, and et. all, "Measurements of higgs boson properties in the diphoton decay channel with 36fb^{-1} of pp collision data at $\sqrt{s} = 13$ tev with the atlas detector," *Physical Review D*, vol. 98, sep 2018.