A Search for Decays of the 125 GeV Higgs Boson in the $bb\tau\tau$ Final State with the ATLAS Detector using p-p Collisions at $\sqrt{s} = 13$ TeV

by

Kevin M. Nelson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2023

Doctoral Committee:

Professor Dante Amidei, Chair
Professor Lu Li
Professor Ryan McBride
Professor James Wells
Professor Bing Zhou

Kevin M. Nelson

kevnels@umich.edu

ORCID iD: 0000-0003-4194-1790

DEDICATION

For Andrew, Mom, and Dad.

ACKNOWLEDGEMENTS

First and foremost, I must bring attention to the guidance I received from my doctoral advisor, Professor Dante Amidei. Over the last five years I have learned so much from you and enjoyed your unwavering support. You advised me with character, integrity, and a practical mindset for physics. I conclude my graduate studies a better scientist and communicator for having worked with you.

This thesis would also not have been possible without the sometimes daily assistance and advice from Dr. Chris Hayes. It has been a wonderful experience to work alongside you and watch your family grow with the arrival of Andrew. You are an invaluable friend and colleague, and I'm sure even long after we both leave Michigan I will still be bouncing ideas off of you.

Additionally, the advice and consultation from my colleagues on the $bb\tau\tau$ search team is greatly appreciated. Thank you to Verena Outschoorn, Rafael Lopes De Sa, Cooper Wagner and Yuan-Tang Chou.

I must also acknowledge the hard work and dedication of my colleagues on the sMDT upgrade project. Almost nothing in experimental physics is realized alone, and this is especially true for large-scale instrumentation projects. I thank Professor Bing Zhou, Dr. Claudio Ferretti, Dr. Edward Diehl, Dr. Taisheng Dai, Curtis Weaverdyck, Andy Chen, Liana Simpson, Zichen Wang, Man Yuan, Chaunshun Wei, Dr. Yuxiang Guo, Jing Li, Dr. Xiangting Meng, Leah Cooperrider, Vee Pillsbury, Emmett Salzer, and Evan Carpenter for their hard work in the construction and testing of precision muon chambers.

Immeasurable support has come from my friends and family over the last five years. Thank you to so many cherished friends: Chami, Avik, Thomas, Aiden, Andy, Maya, Alec, Haley, Joe, Christina, Tessa, Ariana, Matt, David, Liana, Ashley, Dan, Amanda, Jack, Emily, and so many more. If grad school wasn't already hard enough, we survived a global pandemic together, and you kept me sane through the most uncertain times. And to Mom, Dad, Andrew, Ben, Mike, Anne, Greg, Grant, Matthew, Julia, Josh, Ashley, Henry, Ian and Daniel: you made me who I am today, and for that I am forever thankful. To Mia, thank you for believing in me, even when I did not.

Finally, I acknowledge a community at-large of physicists and engineers both at CERN and around the world who made the collection of the data presented in this thesis possible. The LHC project takes tens of thousands of extraordinarily talented individuals, and represents the pinnacle of human achievement.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vii

LIST OF TABLES xx

LIST OF APPENDICES xxiv

LIST OF ABBREVIATIONS xxv

ABSTRACT xxix

CHAPTER

I. Introduction and Theoretical Motivation 1

1.1 Introduction 1
1.2 The Standard Model of Particle Physics 2
1.3 Beyond the Standard Model 7
1.4 $bb\tau\tau$ Search Strategy 12

II. Experimental Apparatus 17

2.1 Particle Accelerators 17
2.2 The Large Hadron Collider 20
2.2.1 LHC Run 2 22
2.3 The ATLAS Detector 23
2.3.1 Overview 23
2.3.2 Magnet System 25
2.3.3 Inner Detector 26
2.3.4 Electromagnetic Calorimeter 28
2.3.5 Hadronic Calorimeter 30
2.3.6 Muon Spectrometer 33

2.3.7 Trigger and Data Acquisition . . . 35

III. Testing of sMDT Chambers for the ATLAS Phase-II Upgrade . . . 37

3.1 Introduction . . . 37
3.2 sMDT Construction . . . 39
3.2.1 Tube construction . . . 39
3.2.2 Chamber construction . . . 42
3.2.3 Electronics installation . . . 48
3.3 Chamber Quality Control Tests . . . 52
3.3.1 Data Acquisition . . . 52
3.3.2 Data Analysis . . . 54
3.3.3 Resolution . . . 61
3.3.4 Efficiency . . . 62
3.4 Multiple Coulomb Scattering . . . 63
3.4.1 Geant4 Simulation . . . 63
3.4.2 Simulation of the Cosmic-Ray Spectrum . . . 64
3.4.3 Deconvolution of Multiple Scattering Effects . . . 66
3.5 Systematic Uncertainty Estimation . . . 67
3.5.1 Auto-calibration and r(t) parameterization . . . 67
3.5.2 Track Fitting . . . 67
3.5.3 Soft-Electron Component of Cosmic-Ray Spectrum . . . 68
3.5.4 Time Slew Correction . . . 69
3.5.5 Signal Propagation Time . . . 71
3.5.6 Systematic uncertainties on efficiency measurement . . . 72
3.6 Results . . . 72
3.6.1 Noise Rate . . . 72
3.6.2 Single-Hit Resolution . . . 73
3.6.3 Efficiency . . . 75
3.7 Conclusions . . . 76

IV. Physics Object Reconstruction . . . 79

4.1 Overview . . . 79
4.2 Inner Detector Tracking . . . 81
4.3 Primary Vertex . . . 81
4.4 Electrons . . . 81
4.5 Muons . . . 83
4.6 Jets . . . 85
4.7 Hadronic Taus . . . 88
4.8 Missing Transverse Energy . . . 90
4.9 Overlap Removal . . . 91

V. Data Analysis Strategy . . . 93

5.1 Data and Simulated Samples . . . 95
5.1.1 Data . . . 95
5.1.2 Monte Carlo Simulations . . . 96
5.2 Event Selection . . . 99
5.2.1 Trigger Selection . . . 100
5.2.2 $e\mu$ Channel . . . 103
5.2.3 $\ell\tau_{\text{had}}$ Channels . . . 106
5.2.4 Event Categorization . . . 108
5.3 Signal Modeling . . . 109
5.4 Background Estimation . . . 114
5.4.1 Overview . . . 115
5.4.2 Top . . . 116
5.4.3 Drell-Yan (+jets) . . . 126
5.4.4 Fake leptons in the $e\mu$ final state . . . 130
5.4.5 Jet$\rightarrow \tau_{\text{had}}$ fakes in the $\ell\tau_{\text{had}}$ final states . . . 138
5.4.6 Fake leptons in the $\ell\tau_{\text{had}}$ final states . . . 145
5.5 Neural Network Discriminant . . . 147
5.6 Systematic Uncertainties . . . 154
5.6.1 Experimental Uncertainties . . . 154
5.6.2 Fake lepton and tau estimates . . . 159
5.6.3 Theoretical Uncertainties . . . 161
5.7 Measurement Technique . . . 164

VI. Results . . . 173

VII. Conclusions . . . 181

APPENDICES . . . 183

BIBLIOGRAPHY . . . 270

LIST OF FIGURES

Figure

1.1 Branching ratios of the light psuedoscalar a in the 2HDM+S model [1]. (a) Type I, (b) Type II $\tan\beta = 0.5$, (c) Type II $\tan\beta = 5$. 10

1.2 Branching ratios of the light psuedoscalar a in the 2HDM+S model [1]. (a) Type III $\tan\beta = 0.5$, (b) Type III $\tan\beta = 5$, (c) Type IV $\tan\beta = 0.5$, (d) Type IV $\tan\beta = 5$. 11

1.3 Feynman diagram: a Higgs boson decays to two b quarks and two τ leptons. The hatched circle represents a "blob," which stands in for multiple possible mediators 13

1.4 Dominant Higgs production modes at the LHC. (a) gluon-gluon fusion (b) vector-boson fusion (c) associated production of vector and Higgs bosons. $t\bar{t}H$ is ommitted because it is not present in the simulated higgs boson production samples used for this work. For more information on simulated samples, see Section 5.1.2. 14

1.5 Comparison of initial and final state particles for (a) exotic Higgs decay signal and (b) $t\bar{t}$ background. The (visible) incident and outgoing particles are the same, so a detector that can measure the energy and direction (4-vector) of each particle is necessary. 15

2.1 Phase space diagram for transverse beam dynamics in one dimension [2]. The x axis is translation and the y axis is momentum. The area of the ellipse is the beam emittance. α, β, and γ are the usual Courant-Snyder parameters. 19

2.2 The CERN accelerator complex [3]. In this diagram, Linac2 has been replaced by Linac4, but the data collected for this thesis used the Linac2 accelerator. The four beam crossing sites are shown with the experiment name and a yellow circle on the LHC ring. 20

2.3 Integrated luminosity for the Run 2 data taking period delivered by the LHC (green), recorded by ATLAS (yellow) and good for ATLAS physics (blue) [4]. 21

2.4 Mean number of interactions per bunch crossing for the Run 2 data taking period [4]. 22

2.5 The ATLAS detector at the CERN LHC [5]. The detector is 25 m tall and 44 m long and weighs 7000 tonnes. 24

2.6 The geometry of the ATLAS magnet system. Magnet windings are shown in red, including the inner solenoid (central red cylinder) and three toroids (one barrel and two endcaps), which each have eight coils. The various layers of the tile calorimeter, including the return yoke, are shown in blue, yellow and green between the solenoid and toroid magnets [6]. 26
2.7 The Inner Detector [6]. Not shown is the Insertable B-Layer: an additional pixel layer added in May 2014 at R=33 mm from the beam axis after a smaller beam pipe was installed [7, 8, 9]. 27
2.8 Diagram of cells in the ATLAS EM calorimeter [6]. There are 3 layers of cells with varying sizes. Charge is collected on electrodes lining the lead accordion. 30
2.9 Amount of material in front of the EM calorimeter [6]. The crack between 1.37 and 1.52 is vetoed due to a significant amount of material from a support structure for other detector components. 31
2.10 A single segment of the ATLAS tile calorimeter [6]. There are 64 azimuthal segments for full coverage. 32
2.11 The ATLAS muon spectrometer [10]. (a) Barrel Resistive Place Chambers (RPCs) and Monitored Drift Tube (MDT) chambers depicted with the toroid magnets and calorimeters. Alternating large and small MDT chambers are shown. (b) Muon spectrometer layout in the (z, y) plane for a small azimuthal sector. 33
2.12 Definition of the sagitta [11]. For a negatively charged muon traversing the inner, middle, and outer stations (in that order), a magnetic field into the page will produce the observed deflection. 34
2.13 Trigger efficiency for single electron triggers used in this search [12]. 36
2.14 Muon trigger efficiency measured in the barrel with RPCs [10] with respect to (a) offline muon p_T and (b) mean number of interactions per bunch crossing (pileup). In the endcap ($1.05 < |\eta| < 2.4$) the trigger efficiency plateau is approximately 90%. 36
3.1 (a) Exploded diagram of endplug assembly. (b) Cross-section of assembled tube with signal-cap (yellow), gas manifold (gray) and twister (blue) [16]. . 40
3.2 An assembled sMDT chamber (a) top view with platforms, kinematic mounts, sensor cables, HV distribution boxes and Faraday cages. (b) side with kinematic mounts attached to crane (blue) and spacer cross bars visible between the two MLs. 43
3.3 Automatic glue-machine mounted on the granite table. A fully-glued sMDT chamber sits in the precision jigging. 44
3.4 (a) The precision comb is assembled on the granite table. (b) The comb is shown with a partially assembled sMDT chamber. The comb constrains the horizontal position to ±4 µm and the vertical position to ≈10 µm RMS. . . 45
3.5 (a) The height gauge in use on the granite table. (b) Results from height gauge measurements from 30 sMDT chambers. The RMS width is below the specification of 20 µm. 45
3.6 Diagram of a single RASNIK line. The RASNIK system is sensitive to movement of the mask, lens, and CCD sensor to measure chamber deformation. 46

3.7 Measured chamber deformation as the chamber is rotated through 360°. The error bars are the standard deviation of the measurements from 30 chambers. 46
3.8 Assembly of the gas distribution manifold. (a) two gas stacks (with opposite chirality) shown before assembly and with o-rings. (b) Gas stack assembly in 3D printed jigging. (c) gas stacks being mounted on a gas bar with more jigging. (d) fully assembled gas bar. Four gas bars are used per chamber. . 47
3.9 ML gas leak rate for 30 sMDT chambers. The dashed line is the maximum acceptable gas leak rate. 48
3.10 Diagram of electrical connections for a single MDT tube [17]. Resistor values are in Ohms. For sMDTs, the 383 Ω resistor is replaced with a 330 Ω resistor. 49
3.11 Grounding cable connections. (a) Gas bar to the side panel on the HV side. (b) Side panel to the support structure on the HV side. (c) Gas bar and the support structure to the side panel on the RO side. (d) Side panel to the support structure on the RO side. 50
3.12 Front end electronics for read-out and high voltage distribution. 51
3.13 Block diagram of the sMDT readout electronics and mini-DAQ system. See text for explanation of components. 53
3.14 sMDT cosmic ray test station in a humidity-controlled room. The sMDT chamber sits on a cart with a 1.5m^2 scintillator mounted above. In this image the Faraday cage is not yet fully assembled and the HV HH cards are seen on the chamber. 53
3.15 (a) Example drift time spectrum for one channel from a cosmic-ray data run collected using the BMG prototype chamber. The red curve is the Fermi-Dirac fit of the rising edge to determine t_0. The green curve is a fit the falling edge to determine t_{max}. DTmax is $t_{max} - t_0$ (b) Example ADC spectrum for the same channel, where the red curve and text show the results of a skew normal fit. 55
3.16 Chi-square distribution for fitted straight line tracks. The tracks are fit in a 2-D plane perpendicular to the wires, as shown in Figure 3.20. The red curve shows the expected chi-square distribution, given the number of degrees of freedom, normalized to the same area as the histogram. 58
3.17 The angular distribution for all fitted tracks for 24 hour cosmic-ray run taken with the BMG prototype. 58
3.18 The auto-calibrated r(t) function. Note that the Chebyshev polynomials are defined on the range [-1,1] so this r(t) function is linearly scaled to $[t_0,t_{\max}]$ for each tube. Therefore, different tubes may have slightly different r(t) functions if they have a differing maximum drift time $t_{\max} - t_0$. The hatched region represents the resolution as a function of radius presented in Figure 3.30a. 59
3.19 Δt_0 distribution from fit residual calculation for cosmic-ray data taken with the BMG prototype. 60

3.20 A single cosmic ray muon event in a BIS chamber with a straight line track fitted. The blue circles show the drift radius of each hit, which is determined using the auto-calibrated r(t) function. The grey and white regions show the demarcations between different readout electronics cards. 60
3.21 (a) Biased residuals from a cosmic-ray run taken with the BMG prototype. The red curve shows the double Gaussian fit. The fit parameters shown in the box are: A_n, σ_n = Amplitude and width of the narrow Gaussian fit, respectively; A_w, σ_w = Amplitude and width of the wide Gaussian fit, respectively; μ = common mean of wide and narrow fits; σ = the biased residual distribution width from Equation 3.5. (b) Unbiased residuals from cosmic-ray data taken with the BMG chamber. The red curve shows the double Gaussian fit. The fit parameters are the same as shown in (a) except σ = the unbiased residual distribution width using Equation 3.5. 62
3.22 3D event display of a BIS1 sMDT chamber in Geant4. 63
3.23 Residual distribution obtained from simulating multiple scattering of cosmic-rays passing through the test chamber. There is a sharp central peak and non-Gaussian tails. 64
3.24 Probability distribution for cosmic-ray muon flux at sea level. The sum of two exponential curves are fit to cosmic-ray data. For functional form, see Equation (3.7). For fit parameters, see Table 3.2. 65
3.25 Biased residual distribution of the BMG chamber from Figure 3.21a deconvoluted with the simulated multiple scatter residual distribution in Figure 3.23. The red curve is a double Gaussian fit with fit parameters as described in Figure 3.21a shown in box. 66
3.26 (a) Difference in r(t) function calculated on independent data sets. Differences are generally on the order of 5-10 microns. (b) Maximum difference between any two data partitions. Differences are greatest near the wire and tube wall, but are again on the order of 5-10 microns. 68
3.27 ADC dependence of residuals (a) before and (b) after applying time slew correction, as well as with the amplitude of the time slew correction varied 25% (c) down and (d) up. The residuals are divided by the average drift velocity to put the y axis in units of nanoseconds. The red dots show the mean residual/drift velocity in coarse bins and the black line is a straight line fit to this coarsely binned data. The text boxes show the equations found for the straight line fits. 70
3.28 Efficiency within the gas volume measured as a function of the number of layers penetrated. 0 layers penetrated corresponds to the top layer, 1 to the second from the top layer, and so on. The constant fit has a chi-square per degree of freedom 0.45, and is not in tension with the data. No systematic uncertainty is adopted. 73
3.29 (a) Noise rate as a function of threshold for a single chamber with the HV on and off. When the HV is on, cosmic-ray hits are seen at the expected rate, about 10 Hz. (b) Chamber noise for 30 chambers at 39 mV threshold, HV on and off. 74

3.30 (a) Observed resolution as a function of radius, after performing the deconvolution technique to remove the effects of multiple scattering. (b) Efficiency as a function of drift radius shown with statistical error bars. The region near the tube wall is responsible for most of the inefficiency of the chamber within the gas volume. The 0.9 mm gap between gas volumes, consisting of the two 0.4 mm tube walls and the 0.1 mm space between tubes, is the reason for the rest of the inefficiency. 76
3.31 Measured resolution for the first 30 sMDT chambers constructed at UM. . 77
3.32 Measured (a) tube and (b) tracking efficiency for the first 30 sMDT chambers constructed at Michigan. The tube efficiency is the efficiency when the particle is inside the instrumented volume, and the tracking efficiency is the efficiency of a layer of tubes, with the geometric acceptance included. The expected tracking efficiency is 94.2%. 77
4.1 Electron efficiency as a function of the number of primary vertices [13]. Three identification WPs are shown, and the ratio plot shows that the scale factor will be different for the choice of WP. 82
4.2 Prompt muon efficiency at the various working points compared to muons from light hadrons as a function of (a) η and (b) p_T [14]. The MS has less coverage in the region $|\eta| < 0.1$, causing a drop in efficiency. 84
4.3 Distribution of muon isolation discriminant for prompt and non-prompt muons [14]. In this search, the discriminant $p_{\mathrm{T}}^{\mathrm{varcone30}} + 0.4E_{\mathrm{T}}^{\mathrm{neflow20}}$ was selected because it gives the best discrimination between heavy flavor backgrounds and prompt muons. 85
4.4 Efficiency for (a) b-jets, (b) c-jets, and (c) light jets to be accepted by DL1r in a $t\bar{t}$ MC sample. The different DL1r WPs are shown. 87
4.5 Performance of the RNN tagger for tau identification [15]. The performance is shown for 1- (red) and 3-prong (blue) τ_{had} candidates. The BDT (dashed line) and RNN (solid line) algorithms are compared. The markers show the WPs for the RNN. 89
5.1 Example of the Σm_T distribution for the $\mu\tau_{\mathrm{had}}$ channel. The $t\bar{t}$ CR is defined by $\Sigma m_T > 120$, and the SR and high mass CR both have the cut $\Sigma m_T < 120$ applied. 104
5.2 Control and validation regions used for DY and $t\bar{t}$ in all channels. There is no upper limit to Σm_T for the $t\bar{t}$ CR, nor is there an upper limit on the number of b-jets in the $t\bar{t}$ CR or signal region. In this context, DeXTer and DL1r jets are counted as a "b-tag." 105
5.3 Orthogonalization scheme for the three trigger regions in the $e\mu$ channel. . 106
5.4 Category definitions for the three heavy flavour signal categories (1 b-jet, >1 b-jet, and DeXTer category) and the 0 b-jet CR for all channels. 109
5.5 Trigger efficiency in the ggF production mode for the $e\mu$ channel 111
5.6 Multiplicity of DL1r tagged (anti-k_t $R = 0.4$) jets for different simulated mass samples. 111
5.7 Visible mass of the $e\mu$ system in simulated signal samples in the $e\mu$ channel. 112
5.8 MMC most likely mass of the $e\mu$ system in simulated signal samples in the $e\mu$ channel. 113

5.9 Reconstructed mass of the $\nu_1 + \nu_\tau$ system using the MMC for simulated signal samples in the $e\mu$ channel. 113
5.10 D_ζ for $m_a = 30$ GeV simulated signal sample and $Z+$ jets background, showing the discriminating power of this variable. 114
5.11 Validation plots in the $t\bar{t}$ CR for the $\mu\tau_{\text{had}}$ channel. The N b-jets, N jets inclusive, and H_T distributions are shown before reweighting (first row), with only N b-jet reweighting applied (second row), and with the full sequential reweighting applied (third row). 118
5.12 The $t\bar{t}$ weights derived during the sequential reweighting in the $\mu\tau_{\text{had}}$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template. 119
5.13 Validation plots in the $t\bar{t}$ CR for the $e\tau_{\text{had}}$ channel. N b-jets, N jets inclusive, and H_T are shown before reweighting (first row), with only N b-jet reweighting (second row), and with the full sequential reweighting (third row). 120
5.14 The $t\bar{t}$ weights derived during the sequential reweighting in the $e\tau_{\text{had}}$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template. 121
5.15 (a) H_T reweighting is necessary in the $e\mu$ channel, as observed in the $\ln(H_T)$ distribution. (b) Reweighting fit as a function of $\ln(H_T)$. 123
5.16 Validation plots in the $t\bar{t}$ CR for the $e\mu$ channel. N b-jets, N jets inclusive, and H_T are shown before reweighting (first row), with only N b-jet reweighting (second row), with N b-jet and N jet reweighting (third row), and with the full sequential reweighting, including H_T (fourth row). 124
5.17 The $t\bar{t}$ weights derived during the sequential reweighting in the $e\mu$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template. 125
5.18 Validation plots in the high mass CR for the $\mu\tau_{\text{had}}$ channel (a) before and (b) after DY reweighting. 128
5.19 Validation plots in the high mass CR for the $e\tau_{\text{had}}$ channel (a) before and (b) after DY reweighting. 128
5.20 Validation plots in the high mass CR for the $e\mu$ channel (a) before and (b) after DY reweighting. 129
5.21 Validation plots in the high mass CR with the DeXTer selection applied for the (a) $\mu\tau_{\text{had}}$ (b) $e\tau_{\text{had}}$ and (c) $e\mu$ channels. 129
5.22 Observed real (a) electron and (b) muon efficiencies which are obtained using $Z \to \tau\tau$ MC. These efficiencies are used as inputs to the matrix method to calculate the lepton fake rates in all three analysis channels. 132
5.23 The observed fake electron efficiencies in the region containing at least 1 b-jet. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3. 133
5.24 The observed fake muon efficiencies in the region containing at least 1 b-jet. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3. 134

5.25 The invariant mass of the $e\mu$ system (a) before and (b) after applying the matrix method correction in the SS CR. The SS CR with 0 b-tags is also shown (c) before and (d) after applying the correction. The correction (see Eq. 5.11) adjusts the weights for events with overlapping isolation cones and two "Loose" leptons proportional to the fraction of the isolation cones which are overlapping. The error bars on the background prediction include both statistical errors and the propagation of errors on the matrix method weights. 137
5.26 The best fit values of r_{QCD} in the (a) $e\tau_{\text{had}}$ and (b) $\mu\tau_{\text{had}}$ channel, binned by selection and decay mode. The bins with r_{QCD} =0 are shown in white. . 140
5.27 Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the inclusive SR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 142
5.28 The best fit values of f_{LJVT} in the (a) $e\tau_{\text{had}}$ and (b) $\mu\tau_{\text{had}}$ channels, binned by selection and decay mode. Bins with f_{LJVT} =0 are shown in white. . . . 142
5.29 Visible mass of the di-τ system in the $e\tau_{\text{had}}$ channel for the SS CR selections. The dominant contributions are from fake τ_{had} (gold) and fake leptons (green). 143
5.30 Visible mass of the di-τ system in the $\mu\tau_{\text{had}}$ channel for the SS CR selections. The dominant contributions are from fake τ_{had} (gold) and fake leptons (green). 143
5.31 Visible mass of the di-τ system in the $e\mu$ channel for the SS CR selections. The dominant contributions is from fake leptons (gold). 144
5.32 Combined FFs for the $\mu\tau_{\text{had}}$ channel under the SR selection ($>$ 0 b-tag). Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 145
5.33 Fake muon efficiencies for (a) 0 b-tag and (b) $>$0 b-tags in the $\mu\tau_{\text{had}}$ channel. 146
5.34 Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category. 150
5.35 (a) Cross-entropy loss function in the $\mu\tau_{\text{had}}$ 1 b-jet category. (b) Network output distribution for the $\mu\tau_{\text{had}}$ 1 b-jet category. 152
5.36 Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. 153
5.37 (a) Prefit and (b) postfit plots for the $m_a = 30$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 171
5.38 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 30$ GeV mass point. 172
6.1 A comparison of the expected and observed mass spectra in each of the nine channels. The $\mu\tau_{\text{had}}$ channel is shown for (a) DeXTer, (b) 1 b-tag, (c) 2 b-tag categories. Next, the $e\tau_{\text{had}}$ channel is shown for (d) DeXTer, (e) 1 b-tag, (f) 2 b-tag categories. Finally, the $e\mu$ channel is shown for (g) DeXTer, (h) 1 b-tag and (i) 2 b-tag categories. 174

6.2 Expected limits in the $\mu\tau_{\text{had}}$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d). 176
6.3 Expected limits in the $e\tau_{\text{had}}$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d). 177
6.4 Expected limits in the $e\mu$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d). 178
6.5 Combined expected limits across all channels and categories. 179
6.6 The observed limits are interpreted under each of the types of 2HDMs. Each of Types II, III, and IV reduce to a Type-I case when $\tan\beta$=1. The plots saturate at a branching fraction of 20% because at a branching fraction this large, analyses of the Higgs Yukawa couplings constrain the 2HDM more than this search. 180
A.1 The observed fake electron efficiencies in the region containing 0 b-jets. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3. 185
A.2 The observed fake muon efficiencies in the region containing 0 b-jets. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3. 186
A.3 Seed jet width fits in the $e\tau_{\text{had}}$ channel for the inclusive SR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 187
A.4 Seed jet width fits in the $e\tau_{\text{had}}$ channel for the inclusive 0 b-tag VR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 187
A.5 Seed jet width fits in the $e\tau_{\text{had}}$ channel for the SS CR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 188
A.6 Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the inclusive 0 b-tag VR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 188
A.7 Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the SS CR selections. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 189
A.8 Combined FFs for the $\mu\tau_{\text{had}}$ channel for the 0 b-tag VR selections. Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 189
A.9 Combined FFs for the $e\tau_{\text{had}}$ channel for the SR selection (> 0 b-tag). Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 190
A.10 Combined FFs for the $e\tau_{\text{had}}$ channel for the 0 b-tag VR selections. Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn. 190
A.11 Validation plots in the 0 b-tag VR for the (a) $\mu\tau_{\text{had}}$ (b) $e\tau_{\text{had}}$ and (c) $e\mu$ channels. 191

A.12 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 60 GeV) CR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. . 193
A.13 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 60 GeV) CR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. . 194
A.14 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 45 GeV) CR for the $e\mu$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. . 195
A.15 Additional plots validating the kinematic modeling in the 0 b-tag VR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 196
A.16 Additional plots validating the kinematic modeling in the 0 b-tag VR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 197
A.17 Additional plots validating the kinematic modeling in the SS CR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 198
A.18 Additional plots validating the kinematic modeling in the SS CR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 199
A.19 Additional plots validating the kinematic modeling in the SS CR for the $e\mu$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 200
A.20 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 60GeV) CR for the $\mu\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 201
A.21 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 60GeV) CR for the $e\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 202

A.22 Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 45GeV) CR for the $e\mu$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 203
A.23 Additional plots validating the kinematic modeling in the SS CR for the $\mu\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 204
A.24 Additional plots validating the kinematic modeling in the SS CR for the $e\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 205
A.25 Additional plots validating the kinematic modeling in the SS CR for the $e\mu$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds. 206
A.26 Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. Continued in Figure A.27. 208
A.27 Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. Continued from Figure A.26. 209
A.28 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. Continued in Figure A.29. 210
A.29 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. Continued from Figure A.28. 211
A.30 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, 1 b-jet category. 212
A.31 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. Continued in Figure A.32. 213
A.32 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. Continued from Figure A.31. 214
A.33 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. Continued in Figure A.34. 215
A.34 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. Continued from Figure A.33. 216
A.35 Pre-Processed network input distributions for the $e\mu$ channel, 1 b-jet category. 217
A.36 Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. Continued in Figure A.37. 218
A.37 Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. Continued from Figure A.36. 219
A.38 Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. Continued in Figure A.39. 220

A.39 Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. Continued from Figure A.38. . . . 221
A.40 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.41 . . . 222
A.41 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.40. . . . 223
A.42 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.43. . . . 224
A.43 Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.42. . . . 225
A.44 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. . . . 226
A.45 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued in Figure A.46. . . . 227
A.46 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued from Figure A.45. . . . 228
A.47 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued in Figure A.48. . . . 229
A.48 Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued from Figure A.47. . . . 230
A.49 Pre-Processed network input distributions for the $e\mu$ channel, 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. . . . 231
A.50 Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.51. . . . 232
A.51 Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.50. . . . 233
A.52 Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.53. . . . 234
A.53 Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.52. . . . 235
A.54 Cross-entropy loss function in the $\mu\tau_{\text{had}}$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories. . . . 236

A.55 Cross-entropy loss function in the $e\tau_{\text{had}}$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories. 237
A.56 Cross-entropy loss function in the $e\mu$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories. 238
A.57 Network output distributions for the $\mu\tau_{\text{had}}$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded. 239
A.58 Network output distributions for the $e\tau_{\text{had}}$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded. 240
A.59 Network output distributions for the $e\mu$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded. 241
A.60 (a) Prefit and (b) postfit plots for the $m_a = 12$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 243
A.61 (a) Prefit and (b) postfit plots for the $m_a = 15.5$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 244
A.62 (a) Prefit and (b) postfit plots for the $m_a = 20$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 245
A.63 (a) Prefit and (b) postfit plots for the $m_a = 25$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 246

A.64 (a) Prefit and (b) postfit plots for the $m_a = 40$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 247
A.65 (a) Prefit and (b) postfit plots for the $m_a = 50$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 248
A.66 (a) Prefit and (b) postfit plots for the $m_a = 60$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix. 249
A.67 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 12$ GeV mass point. 250
A.68 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 15.5$ GeV mass point. 251
A.69 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 20$ GeV mass point. 252
A.70 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 25$ GeV mass point. 253
A.71 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 40$ GeV mass point. 254
A.72 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 50$ GeV mass point. 255
A.73 (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 60$ GeV mass point. 256

LIST OF TABLES

Table

1.1 Properties of the bosons in the SM [18]. There are 8 gluons, which each carry 2 color charges. . . . 3

1.2 Properties of the fermions in the SM [18]. Each type of fermion ("up-type quark", etc.) has three nearly identical "generations" of particles with different masses. The generations are listed in order from top to bottom and are typically denoted I, II, and III. Only the top quark t mass is directly observed; for more information on the calculation of quark masses, see [18]. Each particle also has an anti-particle, which is identical except has opposite electric charge. It is an open question in beyond the SM physics whether the neutral leptons (neutrinos) are their own anti-particle or not. In the SM anti-neutrinos are distinct from neutrinos. . . . 5

2.1 Design resolution for various ATLAS detector subsystems [5, 19, 20]. All units of energy are in GeV. For the forward calorimeter, only a two-parameter resolution model is used [19]. . . . 25

2.2 Intrinsic resolutions of subdetectors in the ID [6]. . . . 28

3.1 sMDT tube materials and operating parameters [16]. . . . 40

3.2 Fitted parameters to Equation (3.7) to describe the muon component of the cosmic-ray spectrum . . . 65

3.3 Mean size of systematic shift for track parameters and impact on resolution. 68

3.4 Summary of error bars for single-hit resolution measurement. Total error is obtained by adding the individual errors in quadrature. . . . 75

3.5 Fitted polynomial coefficients for the residual vs. radius curve, a function of drift distance. . . . 75

4.1 Object reconstruction Working Points (WPs) for all channels ($e\mu$, $\mu\tau_{\text{had}}$, $e\tau_{\text{had}}$) in the analysis. For more information, including object removal procedure, see text. . . . 80

4.2 Lifetimes for tau leptons and select hadrons [18]. . . . 89

4.3 Overlap removal procedure defined in the STANDARD working selection. The steps are performed sequentially and only surviving objects participate in subsequent steps. . . . 92

5.1 Integrated luminosity by data-taking year. . . . 95

5.2 Overview of the MC generators used for the main signal and background samples. . . . 99
5.3 High level triggers used for the single muon, single electron, and di-lepton triggers as defined for this analysis. The single muon trigger is used in the $e\mu$ and $\mu\tau_{\text{had}}$ final states. The single electron trigger is used in the $e\mu$ and $e\tau_{\text{had}}$ final states. The di-lepton trigger is only used for the $e\mu$ final state. . 101
5.4 Best fit parameters to Eq. 5.3 in the $e\mu$ channel. No H_T reweighting is performed in the $\ell\tau_{\text{had}}$ channels. . . . 122
5.5 Summary of nominal unconstrained weights derived for DY in all channels and selections. The nominal weights in this table are applied to the DY MC, including the component of the DY MC which leaks into the data-driven CRs for construction fake lepton and hadronic tau templates. The scale factors in the > 0 b-tag selection are unconstrained normalizations in the PLR fit. The quoted error is statistical and from the propagation of errors on the data-driven fake template. . . . 130
5.6 Lepton efficiency bins. Bins in p_{T} and $|\eta|$ are used for real and fake leptons, while the trigger region and N b-jet bins are only used for fake leptons. . . 132
5.7 Bin edges used for jet$\rightarrow \tau_{\text{had}}$ fakes in the $\ell\tau_{\text{had}}$ channels. . . . 139
5.8 The number of nuisance parameters in the fit, broken down by category. . 155
5.9 Table of fake lepton origin in the $e\mu$ channel . . . 160
5.10 Size of the systematic variation from PDF uncertainties for the DY background. Uncertainties with a $< 1\%$ affect on the background are omitted. . 164
5.11 Summary of binning scheme for statistical interpretation. Note that the SR and low NN output bin definitions change depending on the mass hypothesis being tested. For more information, see text. . . . 170
5.12 The cut on $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ which produces the highest sensitivity in each category is adopted. These cuts leave some region of $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ between the CR and SR for each category. . . . 170
B.1 Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $e\mu$ final state. . . . 257
B.2 Weighted cutflow table for t full Run 2 MC for the $e\mu$ final state. . . . 258
B.3 Weighted cutflow table for diboson full Run 2 MC for the $e\mu$ final state. . . 258
B.4 Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $e\mu$ final state. . . . 258
B.5 Weighted cutflow table for $Z \rightarrow \tau\tau$ full Run 2 MC for the $e\mu$ final state. . . 258
B.6 Weighted cutflow table for SM $H \rightarrow \tau\tau$ full Run 2 MC for the $e\mu$ final state. 259
B.7 Weighted cutflow table for $H \rightarrow aa \rightarrow bb\tau\tau$ MC at $m_a = 12$ GeV for the $e\mu$ final state. . . . 259
B.8 Weighted cutflow table for $H \rightarrow aa \rightarrow bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $e\mu$ final state. . . . 259
B.9 Weighted cutflow table for $H \rightarrow aa \rightarrow bb\tau\tau$ MC at $m_a = 20$ GeV for the $e\mu$ final state. . . . 260
B.10 Weighted cutflow table for $H \rightarrow aa \rightarrow bb\tau\tau$ MC at $m_a = 25$ GeV for the $e\mu$ final state. . . . 260
B.11 Weighted cutflow table for $H \rightarrow aa \rightarrow bb\tau\tau$ MC at $m_a = 30$ GeV for the $e\mu$ final state. . . . 260

B.12 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $e\mu$ final state. . . . 261
B.13 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $e\mu$ final state. . . . 261
B.14 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $e\mu$ final state. . . . 261
B.15 Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. . . . 261
B.16 Weighted cutflow table for $Z \to \tau\tau$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. 262
B.17 Weighted cutflow table for $Z \to \ell\ell$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. 262
B.18 Weighted cutflow table for SM $H \to \tau\tau$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. . . . 262
B.19 Weighted cutflow table for t full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. . . . 262
B.20 Weighted cutflow table for diboson full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. 263
B.21 Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state. . . . 263
B.22 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 12$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 263
B.23 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 263
B.24 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 20$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 264
B.25 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 25$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 264
B.26 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 30$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 264
B.27 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 264
B.28 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 265
B.29 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $\mu\tau_{\text{had}}$ final state. . . . 265
B.30 Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $e\tau_{\text{had}}$ final state. . . . 265
B.31 Weighted cutflow table for $Z \to \tau\tau$ full Run 2 MC for the $e\tau_{\text{had}}$ final state. 265
B.32 Weighted cutflow table for $Z \to \ell\ell$. . . . 266
B.33 Weighted cutflow table for SM $H \to \tau\tau$ full Run 2 MC for the $e\tau_{\text{had}}$ final state. . . . 266
B.34 Weighted cutflow table for t full Run 2 MC for the $e\tau_{\text{had}}$ final state. . . . 266
B.35 Weighted cutflow table for diboson full Run 2 MC for the $e\tau_{\text{had}}$ final state. 266
B.36 Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $e\tau_{\text{had}}$ final state. . . . 267
B.37 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 12$ GeV for the $e\tau_{\text{had}}$ final state. . . . 267
B.38 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $e\tau_{\text{had}}$ final state. . . . 267
B.39 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 20$ GeV for the $e\tau_{\text{had}}$ final state. . . . 267

B.40 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 25$ GeV for the $e\tau_{\text{had}}$ final state. 268
B.41 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 30$ GeV for the $e\tau_{\text{had}}$ final state. 268
B.42 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $e\tau_{\text{had}}$ final state. 268
B.43 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $e\tau_{\text{had}}$ final state. 268
B.44 Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $e\tau_{\text{had}}$ final state. 269

LIST OF APPENDICES

APPENDIX

A. Additional Validation Plots 184

B. Cutflow tables 257

LIST OF ABBREVIATIONS

2HDM Two Higgs Double Model

2HDM+S Two Higgs Double Model with additional (psuedo)scalar

ALICE A Large Ion Collider Experiment

ATLAS A Toroidal LHC ApparatuS

BIS Barrel Inner Small

BMG Barrel Middel Gap

BSM Beyond the Standard Model

CERN the European Center for Nuclear Research

CMS Compact Muon Solenoid

CR Control Region

CSC Cathode Strip Chamber

CSM Chamber Service Module

DLT di-lepton trigger

DY Drell-Yan

EWSB Electroweak Symmetry Breaking

FCNC Flavor Changing Neutral Current

FF Fake Factor

HLT High Level Trigger

IBL Insertable B-Layer

ID Inner Detector

JVT Jet VerTex

LAr Liquid Argon

LEP Large Electron-Positron Collider

LHC Large Hadron Collider

LHCb Large Hadron Collider beauty

LO leading order

MC Monte Carlo

MDT Monitored Drift Tube

ME matrix element

MM Matrix Method

MMC Missing Mass Calculator

MS muon spectrometer

NLO next-to-leading order

NNLO next-to-next-to-leading order

NP nuisance parameter

OF Opposite Flavor

OLR overlap removal

OS Opposite Sign

PDF Parton Distribution Function

PLR Profile Likelihood Ratio

PNN Parameterized Neural Network

p-*p* Proton-Proton

PS Proton Synchrotron

PSB Proton Synchrotron Booster

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

RF Radio Frequency

RPC Resistive Place Chamber

SCT SemiConductor Tracker

SLT single lepton trigger

SM Standard Model

sMDT Small Monitored Drift Tube

SPS Super-Proton Synchrotron

SR Signal Region

SS Same Sign

TDAQ trigger and data acquisition

TGC Thin Gap Chamber

TRT Transition Radiation Tracker

VR Validation Region

WP Working Point

ABSTRACT

This thesis presents a search for decays of the 125 GeV Higgs boson (h) to a pair of new (pseudo)scalar particles (a), $h \to aa$, where the a-bosons decay to a b-quark pair and tau lepton (τ) pair. Adding a new (pseudo)scalar to the Standard Model is a simple extension which produces a wide range of phenomenology. The addition of a new (pseudo)scalar is also motivated by dark matter, which could couple to ordinary matter via the a-boson, and by recent experimental excesses in the muon anomalous magnetic moment and W boson mass, which both are modulated by a new scalar.

The search is performed in the mass range $12 < m_a < 60$ GeV in different analysis channels defined by the decay products of the tau pair. The search uses 139 fb^{-1} of proton-proton collision data with $\sqrt{s} = 13$ TeV recorded by the ATLAS experiment at the LHC between 2015 and 2018. The analysis searches in three decay channels of the $\tau\tau$ system and makes use of both leptonic (e or μ) and hadronic (τ_{had}) tau lepton decays: $e\mu$, $\mu\tau_{\text{had}}$, and $e\tau_{\text{had}}$. A further categorization is performed by the experimental signature of the heavy-flavor jets, including a boosted double b-jet tagger. Background processes are modeled with Monte Carlo for event topologies with two prompt leptons or hadronic taus and a data-driven technique is used to estimate the contribution from non-prompt leptons and hadronic tau decays. Control regions are employed to constrain different background processes: high transverse mass for $t\bar{t}$, high $m_{\tau\tau}^{\text{vis}}$ for Drell-Yan, and same-sign visible tau decay products for non-prompt leptons and taus. A Parameterized Neural Network is employed to discriminate between signal and background in the low-mass ($m_{\tau\tau}^{\text{vis}} < 60$ GeV) regime in which $h \to aa$ decays are kinematically allowed.

Limits on the branching ratio for the $h \to aa \to bb\tau\tau$ process are set and no evidence of

any beyond the Standard Model phenomena is observed. This thesis sets the most sensitive limits worldwide on exotic Higgs decays in the $bb\tau\tau$ final state. The branching ratio is limited to less than 1.5% over the mass range from 12 to 60 GeV for the mass of the new psuedoscalar m_a.

A measurement of the resolution, efficiency, and noise rate of drift tube chambers for the ATLAS Phase-II upgrade is also presented. Notably, a novel treatment of the multiple Coulomb scattering correction is included in the drift tube resolution measurement.

CHAPTER I

Introduction and Theoretical Motivation

1.1 Introduction

Particle physics is the study of nature at the smallest distance scales. Fundamental particles are described by excitations in quantum fields which permeate spacetime. Particle *accelerators* and *colliders* are used in laboratories to produce high-energy interactions, which can in turn produce previously undiscovered heavy particles. In laboratories particle interactions are studied with *detectors*, which create a record of the energetic particles which pass through them. The detector itself is made up of an ensemble of fundamental particles in a low energy "ground" state. When an energetic particle passes through the detector it deposits some energy and the ensemble of detector particles are excited. A digital record of this excitation is created using photodetectors or by measuring a voltage or charge deposition. Modern experiments use up to 10^8 individual detector channels, each creating their own digital record every time a particle passes through them. Using these many individual detectors and sophisticated *reconstruction algorithms* high level *observables* are constructed. A change in the distribution of some observable gives sometimes surprising inference about the nature of physical phenomena; for example, the distribution of the scattering angle of particles incident on gold foil led Rutherford to conclude that atoms have a dense nucleus. We use these tools: accelerators, colliders, detectors, and algorithms, to study the decays of the Higgs boson.

This work presents a search for a peak in the mass distribution of a pair of tau leptons, produced in association with b-quarks. The presence of a resonant peak in the mass distribution could be indicative of a new particle (a) resulting from a decay of another particle, the Higgs boson (h). The search is interesting in part because of a recent excess in a similar final state in which muons were used instead of tau leptons [21]. Decays to tau leptons are an orthogonal search to the muonic decays used in the prior search, and would have a much larger decay rate due to the higher mass of the tau leptons, provided that the decay rate is proporitional to the mass, as is the case for the Higgs boson. However, the experimental signature is much more challenging than the search with muons because tau leptons are reconstructed with poor resolution compared to muons. The goal of this thesis is to overcome the experimental challenges of a search in the $bb\tau\tau$ final state and provide a limit which could confirm the excess in the $bb\mu\mu$ final state, or exclude the hypothesis that the excess is due to a decay of $h \to aa$. The search for the decay $h \to aa \to bb\tau\tau$ is revisited in Section 1.4, but first the Standard Model (SM) of particle physics and possible extensions are discussed in more detail.

1.2 The Standard Model of Particle Physics

The SM [22, 23, 24] of particle physics enumerates fundamental particles and describes their interactions. In quantum field theory, particles are objects which transform under irreducible unitary representations of the Poincaré group [25]. The bosons, or "force carriers" transform as tensor representations of the Poincaré group, while the fermions, or "matter particles" transform as spinor representations. The last SM particle experimentally confirmed is the Higgs boson (h) which was discovered in 2012 [26, 27]. In the intervening decade no significant deviation from the SM Higgs boson has been observed for this 125 GeV[1] particle [28, 29]. However, it is possible that Beyond the Standard Model (BSM) physics could be detected at the Large Hadron Collider (LHC) through exotic decay modes of the 125 GeV

[1]Throughout this work, natural units ($c = \hbar = 1$) are used

Boson	Mass [GeV]	Electric Charge [e]	Spin	Color Charge
γ	0	0	1	0
g	0	0	1	carries 2
$W^{\pm}$	80.377±0.012	±1	1	0
Z	91.1876± 0.0021	0	1	0
h	125.25±0.17	0	0	0

Table 1.1: Properties of the bosons in the SM [18]. There are 8 gluons, which each carry 2 color charges.

Higgs boson not predicted by the SM. This chapter describes the SM and motivate both the Two Higgs Double Model with additional (psuedo)scalar (2HDM+S) and the $bb\tau\tau$ decay mode as of particular interest.

The properties of the bosons in the SM are summarized in Table 1.1. Photons carry the electromagnetic force including radio waves, visible light, and x-rays. Particles with *electric charge* interact with photons; for example: the electron has a charge of $e = 1.6 \times 10^{-19}$ Coulombs which is proportional to the strength of its interaction with the photon. Gluons carry the strong nuclear force, which binds together hadrons (including the neutrons and protons in the nuclei of atoms). Particles with *color charge* interact with gluons. There are three "types" of color charge (red, blue and green), in contrast with the single electric charge for the electromagnetic force. Color charges for antiparticles are "antired", "antiblue", or "antigreen"; and the gluon itself carries two color charges (so a gluon could be a red-antiblue gluon, for example). The $W^{\pm}$ and Z bosons are responsible for the weak nuclear force, which causes atomic nuclei to decay in radioactive isotopes. However, there is some mixing between the weak and electromagnetic force between the Z boson and photon. The Z boson is a mass eigenstate which is not only mediating the weak force, but is mixing with the photon and is better described as one of the bosons in *electroweak* theory: the unification of the electromagnetic and weak nuclear interactions. The Higgs boson, via the Higgs mechanism, is responsible for giving mass to the fermions and weak bosons ($W^{\pm}$ and Z bosons) in the

SM. The Higgs boson interacts with fermions proportional to their masses rather than their charges, and the reason for this will become clear later in this chapter.

Properties of SM fermions are summarized in Table 1.2. Fermions are categorized into leptons and quarks. The leptons carry no color charge and have integer electric charge, while the quarks have color charge and fractional electric charge. Quarks are observed only in confined states of two or more quarks, bound together by gluons. Both quarks and leptons are organized into three generations of matter. The quantum numbers of each generation are identical, but the masses are different.

Interactions between particles are commonly understood using the Lagrangian framework. Terms in the Lagrangian ($\mathcal{L}$) provide an understanding of which particles couple to one another. For example, the Lagrangian term $\mathcal{L} \supset \phi^4$ would admit a coupling of four ϕ fields to one another. There are two main types of interactions between particles in the SM Lagrangian: Yukawa couplings [30] and Yang-Mills interactions [31].

Interactions mediated by spin-1 particles (the photon, gluon, $W^{\pm}$ and Z bosons) are described by Yang-Mills theory. Yang-Mills is a general description of SU(N) symmetry groups with massless spin-1 particles transforming under the adjoint representation. The SM is a $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ theory which is composed of multiple types of Yang-Mills interactions. A common way to illustrate which interactions are allowable is to introduce a *gauge invariance* by embedding the spin-1 particles in tensor fields with additional (non-physical) degrees of freedom. The choice of gauge cannot change the physical phenomenology. An interaction between a set of fields occurs if the set of fields can be composed together in a Lagrangian term to preserve all quantum numbers and respect gauge-invariance. For example, because the vector bosons and left handed fermions live in a SU(2) doublet, and right handed fermions are SU(2) singlets, a W^- boson can decay to an electron and an electron neutrino.

Yukawa couplings describe the interaction between the spin-0 Higgs boson and the fermions of the SM. However, many physical properties of atomic and sub-atomic mat-

Fermion		Mass [GeV]	Electric Charge	Spin	Color Charge
up-type quarks	u	$2.16^{+0.49}_{-0.29} \times 10^{-3}$	$\frac{2}{3}$	$\frac{1}{2}$	carries 1
	c	1.27 ± 0.02			
	t	172.69 ± 0.30			
down-type quarks	d	$4.67^{+0.48}_{-0.17} \times 10^{-3}$	$-\frac{1}{3}$	$\frac{1}{2}$	carries 1
	s	$9.34^{+0.86}_{-0.34} \times 10^{-2}$			
	b	$4.18^{+0.03}_{-0.02}$			
Charged leptons	e	5.11×10^{-4}	-1	$\frac{1}{2}$	0
	μ	0.105			
	τ	1.776			
Neutral leptons	ν_e	$< 1.1 \times 10^{-9}$	0	$\frac{1}{2}$	0
	ν_μ	$< 1.1 \times 10^{-9}$			
	ν_τ	$< 1.1 \times 10^{-9}$			

Table 1.2: Properties of the fermions in the SM [18]. Each type of fermion ("up-type quark", etc.) has three nearly identical "generations" of particles with different masses. The generations are listed in order from top to bottom and are typically denoted I, II, and III. Only the top quark t mass is directly observed; for more information on the calculation of quark masses, see [18]. Each particle also has an anti-particle, which is identical except has opposite electric charge. It is an open question in beyond the SM physics whether the neutral leptons (neutrinos) are their own anti-particle or not. In the SM anti-neutrinos are distinct from neutrinos.

ter are described well by the Yang-Mills theory and its strong, weak and electromagnetic forces, including quark confinement, nuclear decay, and the properties of atomic orbitals. What role does the Higgs boson play? Why should there be a Higgs boson at all?

The Higgs boson is responsible for the generation of mass[2] in the SM via electroweak symmetry breaking [32, 33, 34, 35]. The Higgs boson is responsible for the mass of both the weak bosons and the fermions. A Higgs doublet is needed in the SM because the weak force is observed to be mediated by massive vector bosons, but the vector bosons before spontaneous symmetry breaking are massless. The extra degrees of freedom in the Higgs doublet are "eaten" by the W and Z bosons, giving them longitudinal degrees of freedom, and therefore mass.

Another reason that the Higgs doublet is needed is that the weak force only interacts with *left handed* fermions. For example, consider the leptons (and allow me to ignore conjugate terms and signs). Right handed leptons (ℓ_R) are in a SU(2) singlet, meaning they do not interact with the $W^{\pm}$ or Z bosons, while the neutrinos (ν_ℓ) and left handed leptons (ℓ_L) form SU(2) doublets (the two fields in the doublet are conceptualized similarly to the components of a 2D vector):

$$f_R = \ell_R \qquad\qquad f_L = \begin{pmatrix} \nu_\ell \\ \ell_L \end{pmatrix} \tag{1.1}$$

A fermion has mass m if we can write down a Lagrangian term proportional to $m\bar{f}_L f_R$ which is a singlet under all SM gauge transformations. A SU(2) singlet times a doublet transform like a doublet under SU(2). We need another SU(2) doublet in the Lagrangian term for lepton masses: the Higgs doublet. A complex SU(2) doublet has 4 real degrees

[2]It is worth noting that the Higgs mechanism as described is not the only way to give mass to chiral fermions; in fact, it is not even the only way in the SM that chiral symmetry is spontaneously broken. Chiral symmetry is also broken in bound states of strongly interacting matter, generating far more mass than the Higgs mechanism in protons, for example.

of freedom, three of which are Nambu-Goldstone bosons and are eaten by the $W^{\pm}$ and Z bosons (giving them mass). What remains is a single real degree of freedom $h(x)$, the Higgs field:

$$\Phi = \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix} \tag{1.2}$$

where $v = 246$ GeV is the vacuum expectation value of the Higgs field. If v is nonzero, the chiral symmetry is broken and the singlet Lagrangian term involving the right and left handed fermions and Higgs doublet can be expanded:

$$y_\ell \bar{f}_L f_R \Phi = y_\ell \frac{v}{\sqrt{2}} \ell_R \bar{\ell}_L + y_\ell \ell_R \bar{\ell}_L \frac{h(x)}{\sqrt{2}} \tag{1.3}$$

where y_ℓ is the Yukawa coupling strength of the fermion to the Higgs boson. The first term is a mass for the lepton and the second term is a coupling to the Higgs boson, both set by the same interaction strength y_ℓ. For quarks, the mass and flavor eigenstates are not the same, so the argument is slightly more complex, but similar nonetheless. It is now obvious that the nonzero vacuum expectation value in the Higgs field gives rise to masses in chiral fermions, and that the masses are proportional to the fermion coupling to the Higgs boson. As discussed previously, the Higgs boson also completes the description of the weak force, which is short-ranged due to the mass of the vector bosons. In Yang-Mills theory without an additional symmetry-breaking doublet the vector bosons are massless.

1.3 Beyond the Standard Model

Why look beyond the SM? What is left to explain? There is strong astrophysical evidence [36] that the SM does not describe all of the particle matter in the universe and that "dark"

matter, which apparently interacts at most very, very weakly with SM particles, makes up five times as much matter as SM particles[3]. Additionally, precision measurement of the muon anomalous magnetic moment [37] and the W boson mass [38] suggest modifications to electroweak theory are required. The Higgs boson is an especially good candidate particle to study in the search for BSM physics. The SM singlet operator $|\Phi^2|$ can couple to new scalars (or fermions, with an effective operator). However, observations of the 125 GeV Higgs boson are so far consistent with the SM Higgs. Extensions of the SM which modify the Higgs sector must only slightly modify the 125 GeV particle. One such family of modifications to the SM is the Two Higgs Double Model (2HDM) [39]. A complete discussion of the many extensions to the SM which produce exotic Higgs decays is found in [1, 40].

In the 2HDM there are now two SU(2) doublets that participate in electroweak symmetry breaking, Φ_1 and Φ_2. The ratio of the vacuum expectation values of the doublets is parameterized by $\tan(\beta) = v_2/v_1$. The addition of a new SU(2) doublet adds four degrees of freedom: the particles A, H^0 and $H^{\pm}$. These particles are assumed to be heavy enough to evade detection.

In general, the 2HDM allows large Flavor Changing Neutral Currents (FCNCs), which are not observed experimentally. There are four ways to choose which fermions couple to which Higgs doublet, solving the problem of large FCNCs (imposing a $\mathbb{Z}_2$ symmetry). By convention, let the up-type quarks u_R always couple to Φ_2. In Type I, all fermions couple to Φ_2, in Type II only u_R couples to Φ_2, in Type III the quarks couple to Φ_2 and leptons to Φ_1, and in Type IV leptons and u_R couple to Φ_2. Setting $\tan(\beta) = 1$ in Types II, III and IV recovers the phenomenology of a Type I 2HDM.

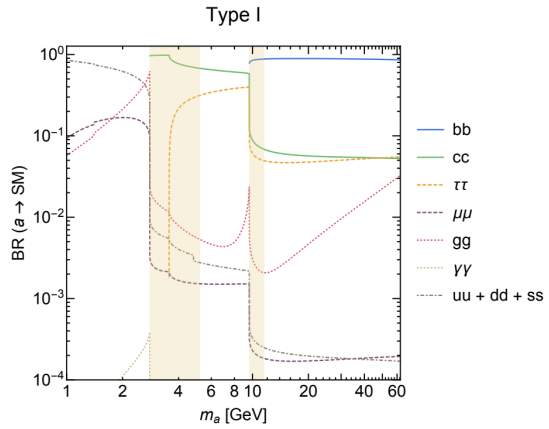
The SM-like h is an admixture of two neutral scalar modes from two Higgs doublets parameterized by the mixing angle α. As such, the couplings of h are modified by the 2HDM. However, in the decoupling limit $\alpha \to \beta - \pi/2$ the coupling strengths are unmodified.

It appears almost as if we have arrived where we started by enforcing the decoupling limit

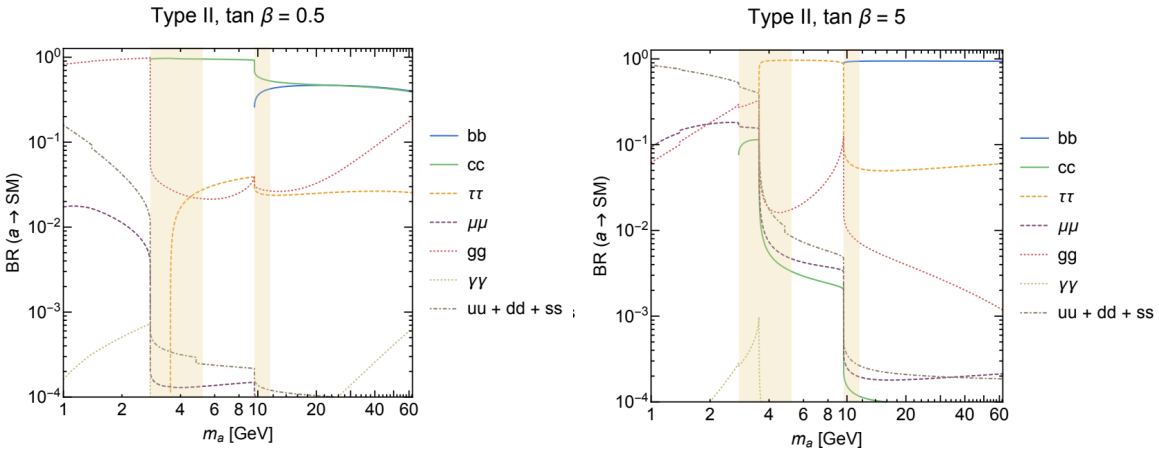
[3]The SM also notably lacks a description of gravity, and other BSM searches and theories are motivated by the particle description of gravity.

and requiring the additional 4 modes from the second doublet are too heavy to be detected (yet). However, a rich phenomenology appears when an additional light psuedoscalar a is added to the 2HDM. Now there is a small mixing between the new psuedoscalar S_I and A parameterized by the angle θ_a, where a is the light state which is mostly composed of S_I. The decay $h \to aa$ is allowable through the scalar potential. The mixing between a and the rest of the scalar sector allows a to decay to SM fermions via the Higgs' Yukawa coupling, despite a having no direct Yukawa couplings. Intuitively, a will now also couple to SM fermions proportional to their masses. The richness of the phenomenology is apparent when the type of 2HDM and $\tan\beta$ are considered. The magnitude of the coupling of a to SM fermions is modified by a factor of $\tan\beta$ when that fermion couples to Φ_1 and $\cot\beta$ when the fermion couples to Φ_2. Even modest values of $\tan\beta$ produce vastly different phenomenology in terms of exotic Higgs decays. Figures 1.1 and 1.2 shows the branching ratio of the a for different values of $\tan\beta$ and different types of 2HDMs. Almost any four-body SM final state is allowable, and a wide variety of them can be dominant in different mass regimes. The wide range of phenomenology motivates a large search program for exotic Higgs decays in many final states. This thesis will study $h \to aa \to bb\tau\tau$.

The $bb\tau\tau$ decay mode has not yet been studied by ATLAS. It is a difficult channel to study due to the poor resolution on the Higgs mass from jets and neutrinos in the final state (for more information on detector resolution, see Section 2.3). The trigger acceptance is low because triggers require leptons and (i) not all $bb\tau\tau$ decays will have a lepton, and (ii) when a lepton is present it will often not be above the ≈ 27 GeV p_{T} trigger threshold. While experimentally challenging, the $bb\tau\tau$ final state is particularly interesting. An excess of events near $m_a = 52$ GeV with significance of 3.3σ was observed by ATLAS in the $bb\mu\mu$ final state [21]. The ATLAS excess is suggestive of an a boson, and searches for analogous decays in other final states are necessary to support or exclude the ATLAS excess. This search in the $bb\tau\tau$ final state could confirm the $bb\mu\mu$ excess and discover a new fundamental particle. However, if the $bb\mu\mu$ excess is not the result of a decay mediated by an extended



(a)



(b) (c)

Figure 1.1: Branching ratios of the light psuedoscalar a in the 2HDM+S model [1]. (a) Type I, (b) Type II $\tan\beta = 0.5$, (c) Type II $\tan\beta = 5$.

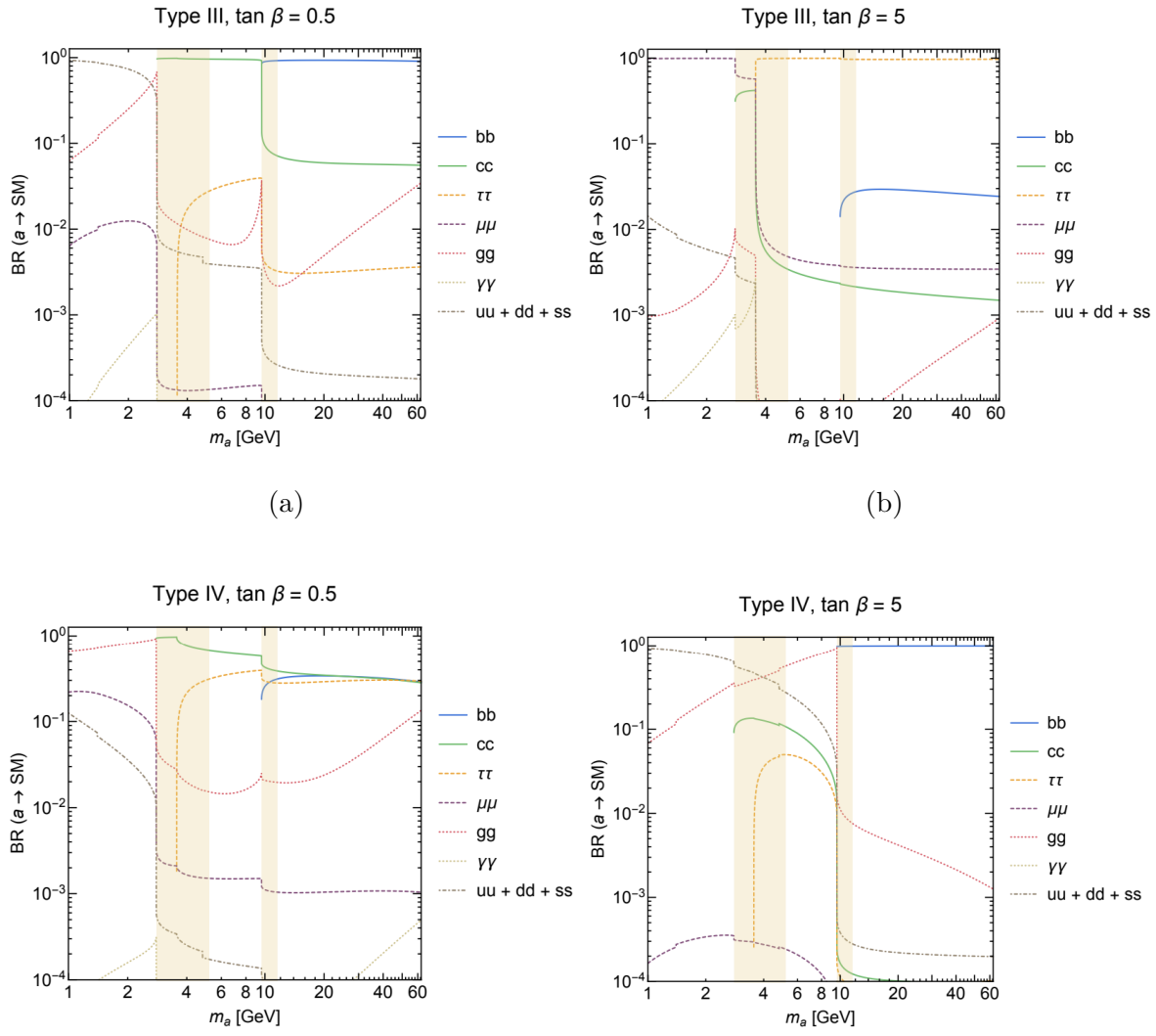


Figure 1.2: Branching ratios of the light psuedoscalar a in the 2HDM+S model [1]. (a) Type III $\tan\beta = 0.5$, (b) Type III $\tan\beta = 5$, (c) Type IV $\tan\beta = 0.5$, (d) Type IV $\tan\beta = 5$.

Higgs sector, and the mediator is therefore not inheriting Yukawa couplings from the SM-like Higgs boson, then the excess could still be the hint of a new particle. For example, a Z' boson would decay democratically to all leptons, and a search in $bb\tau\tau$ would not be as sensitive as a search in $bb\mu\mu$. Therefore, this search in $bb\tau\tau$ will either confirm the excess in $bb\mu\mu$ or be an important step to rule out one possible explanation (2HDM) of the $bb\mu\mu$ excess. Whether there is a significant signal in $bb\tau\tau$ or not, the search will have a significant impact on the interpretation of the $bb\mu\mu$ excess.

Note that additional degrees of freedom in the model are allowable, but not necessary to produce exotic Higgs decays. If a scalar is instead added to the 2HDM an even richer phenomenology is possible. There is one additional degree of freedom, α', which is a rotation angle parameterizing the mixing of the two neutral Higgs bosons in the added scalar. The angle α' does not control the normalization of the mixing, which is necessarily small. The additional scalar case reduces to the added pseudoscalar case when the angle $\alpha' = \beta$. The freedom from α' simply increases the maximum branching fraction to the desired final state. Furthermore, the new scalar can acquire a vacuum expectation value. While this is not required in order to produce exotic Higgs decays, it can modulate the relative frequency of $h \to aa$ or $h \to Za$ decays [1]. Should a $h \to aa$ decay be discovered in the relative absence of $h \to Za$, these additional degrees of freedom could prove important to the full description of nature. While the additional degrees of freedom does provide a richer phenomenology, it is often sufficient to discuss the pseudoscalar case. Moving forward, the psuedoscalar is discussed for simplicity.

1.4 $bb\tau\tau$ Search Strategy

Predictions of the SM or a BSM theory are described in matrix elements: quantum mechanical amplitudes for initial state particles to transition to final state particles. In our specific case, we search for a Higgs boson decaying to two b quarks and two τ leptons[4], shown

[4]Tau leptons will decay to other particles before detection.

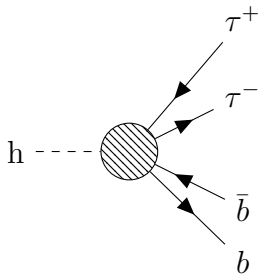


Figure 1.3: Feynman diagram: a Higgs boson decays to two b quarks and two τ leptons. The hatched circle represents a "blob," which stands in for multiple possible mediators

in Figure 1.3. If a signal is found, then by counting the number of times an event with the $h \to bb\tau\tau$ event topology seems to occur in our detector we can measure the branching fraction $h \to aa \to bb\tau\tau$. In the absence of a signal, limits on the model parameters $\tan(\beta)$, α, and the type of 2HDM, which would give rise to $h \to aa \to bb\tau\tau$, are computed. However, Higgs bosons cannot be considered incident particles in this experiment. We cannot procure a sample of Higgs bosons in the way one could mine a radioactive element; rather, they must be produced in a laboratory by high energy particle collisions.

The only machine which produces Higgs bosons in high enough quantity is the LHC. Other machines including the Large Electron-Positron Collider (LEP) and the Tevatron likely produced some Higgs bosons, but not in observable quantities[5]. The LHC uses Proton-Proton (p-p) collisions to produce Higgs bosons. The desired final state particles are still $bb\tau\tau$, but the incident particles are the gluons and quarks that compose the proton, not the Higgs boson. Figure 1.4 shows the dominant production modes of the Higgs boson at the LHC, and Figure 1.5a shows an example of a full diagram of the incident protons to final state $bb\tau\tau$. The LHC discovered the Higgs boson using p-p collisions at a center-of-mass energy ($\sqrt{s}$) of 7 and 8 TeV, and this thesis uses data at $\sqrt{s} = 13$ TeV. The production cross sections of Higgses at the LHC are a function of the center-of-mass energy, so 13 TeV collisions have approximately 3x the Higgs production cross section as 7 TeV collisions.

The Higgs bosons produced at the LHC will decay quickly enough to evade direct detec-

[5]At LEP the production of Higgses was disfavored due to the Higgs Yukawa coupling to electrons, meaning the lower cross section ZH mode dominated, and the Tevatron was lower energy than the LHC

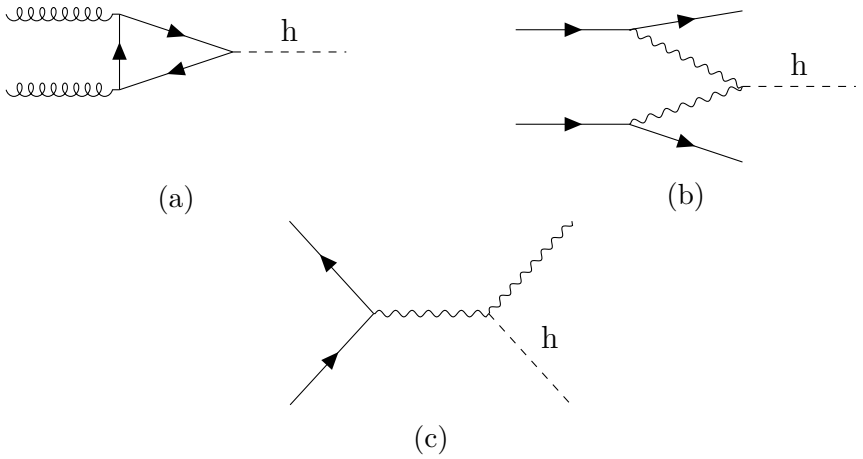


Figure 1.4: Dominant Higgs production modes at the LHC. (a) gluon-gluon fusion (b) vector-boson fusion (c) associated production of vector and Higgs bosons. $t\bar{t}H$ is ommitted because it is not present in the simulated higgs boson production samples used for this work. For more information on simulated samples, see Section 5.1.2.

tion, and are only detected via correlations between the Higgs's decay products. Furthermore, even some of the decay products of the Higgs boson will decay and are also indirectly detected. Tau leptons have a lifetime on the order of 10^{-13} seconds, and will decay to an electron and missing energy (from 2 neutrinos), a muon and missing energy (from 2 neutrinos), or hadrons and missing energy (from 1 neutrino). The b-quarks will hadronize, producing showers of secondary particles, including soft leptons. The signature of b-quark showers are different than showers initiated by light quarks (u, d, s) and charms (c), so flavor-tagging algorithms are used to discriminate b-jets (moving forward, b-jets and b-tags will be used interchangably to refer to the number of jets which are flavor-tagged). Therefore, in order to search in the $bb\tau\tau$ final state, this thesis describes the techniques for detecting electrons, muons, b-jets and missing energy using the ATLAS detector.

The LHC produces Higgs bosons, but that is not all it produces. It is unavoidable that other interactions between partons will occur. For example, consider the production of two top quarks, $t\bar{t}$, at the LHC (see Figure 1.5). The same incident and final state particles as our signal are present[6]. The $t\bar{t}$ diagram is a *background*, while the Higgs decay to $bb\tau\tau$ is a

[6]Except for the neutrinos, which interact so rarely that they are "invisible" to our detector. An extremely large neutrino flux is required to detect their interactions.

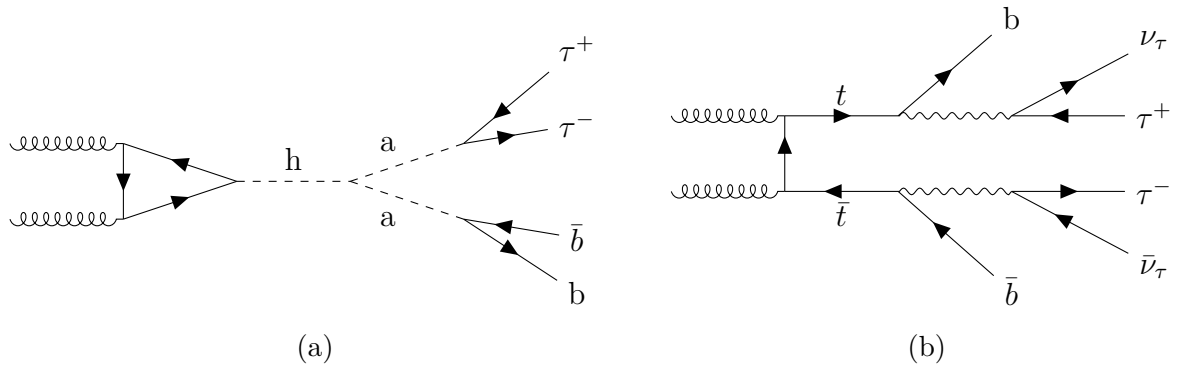


Figure 1.5: Comparison of initial and final state particles for (a) exotic Higgs decay signal and (b) $t\bar{t}$ background. The (visible) incident and outgoing particles are the same, so a detector that can measure the energy and direction (4-vector) of each particle is necessary.

signal. The background processes, including but not limited to $t\bar{t}$, are orders of magnitude larger than the signal, so an "inclusive" search, which would make little to no concerted effort to separate signal and background, would not be very sensitive to the presence or absence of the signal process. The signal and background can be discriminated with a detector that can measure the *energy*, *position*, and *identity* of the final state particles. The matrix elements for these two processes have different kinematics. For instance, the four-body $bb\tau\tau$ system reconstructs the Higgs mass in signal, but not in the background. The kinematics of Ws and top quarks in the background are not present in the signal. The desire to separate these signal and background processes sets the requirements for our detector: it must have the ability to precisely measure the energy (or momentum) and position of the decay products so the kinematics of the Higgs boson, top quark, or other intermediate particles are resolved.

There are two main ways to measure the properties of the final state particles: *tracking* and *calorimetry*. Tracking refers to the non-destructive process of capturing a record of the trajectory of a charged particle as it moves through the detector. Tracking is used to measure the momentum of muons as their path bends in a magnetic field. Tracking is also used to measure the position of all charged particles and associate them with a primary vertex, allowing disambiguation of multiple interactions. Calorimetry refers to the process of estimating the energy of a particle by absorbing the particle's energy (or the energy of

a shower of particles). The ionization electrons or light produced in the calorimeter are collected and are used to determine the energy of the incident particle. Chapter II discusses the production of Higgs bosons at the LHC and the tracking and calorimetry of the ATLAS detector. Chapter III discusses of a future upgrade of the ATLAS detector to increase the barrel muon trigger efficiency. The resolution, noise, and efficiency of new drift chambers are measured, including a novel technique for addressing multiple Coulomb scattering of cosmic-ray muons.

The detector records information about the location and energy/momentum of particles, but on its own does not provide the *identity* of the particles. Algorithms are needed to reconstruct high level objects, including but not limited to b-jets, leptons from τ decays, and hadronic τ decays. For example, here is a high level algorithm for identifying electrons and photons: both produce an electromagnetic shower, but electrons also produce a bent track in the inner tracking detector due to their charge and the magnetic field. The algorithms used to identify and calibrate these so-called "physics objects" are discussed in Chapter IV.

The data analysis strategy for $bb\tau\tau$ decays is discussed in Chapter V. In the simplest possible terms, the experiment is to count the number of events in a signal-enriched region (defined by some cuts on the identity, energy/momentum, and position of particles) and compare it to the background-only hypothesis. Then, based on the observed counts in the signal-enriched region, limits on the signal rate (or, a if a large excess is seen, a discovery significance) are computed. The background prediction is constrained by the observed counts in an orthogonal, signal-depleted region. Results are presented in Chapter VI and Chapter VII draws conclusions.

CHAPTER II

Experimental Apparatus

This chapter describes the experimental apparatus necessary to produce Higgs bosons, detect their decay products, and discriminate them from SM backgrounds. Section 2.1 describes accelerators and Section 2.2 describes the Large Hadron Collider. The ATLAS detector, which is used to measure the energy or momentum of decay products from LHC collisions, is discussed in Section 2.3. Specific identification algorithms using the raw data collected by these detectors are discussed in more detail in Chapter IV.

2.1 Particle Accelerators

The production of Higgs bosons requires the collision of energetic and stable beams of protons at high luminosity (high particle interaction rate). A common way to impart energy into a beam is via a Radio Frequency (RF) cavity. In a RF cavity a rapidly (up to 10s of GHz) oscillating electric field gives energy to charged particles. Each oscillation of the field gives a kick to a small packet, or "bunch", of particles. The design of the RF cavities is such that when the electric field oscillation is opposing the acceleration of the particles, the bunch is in a dead zone between cavities. Thus, the bunch feels the cumulative effect of many small kicks, all in the same direction. High energy particles can be stored in a storage ring, collided, or delivered to fixed targets. A storage ring is economical because it collides the same bunches over and over again as they repeatedly circle the ring. Dipole magnets

are used to bend the trajectory of the particles into a ring. The energy of a beam stored in a ring is proportional to the RF accelerating gradient, the magnetic field strength of the dipole magnets, and the radius of the ring [18].

A stable beam in a storage ring requires stable orbit. If the transverse dynamics are unstable, the particles will fly out of the storage ring before they are collided. Alternating focusing and defocusing quadrupole magnets with focal length f and spaced by a distance L, called a FODO lattice, are used to maintain the required stable dynamics. In the thin-lens approximation, the translation matrix acting on a vector $X = \{x, x'\}$ (where x is the transverse displacement and x' is its derivative) is [2]:

$$M_{FODO} = \begin{pmatrix} 1 - 2\frac{L^2}{f^2} & 2L(1 + \frac{L}{f}) \\ 2(\frac{L}{f} - 1)\frac{L}{f^2} & 1 - 2\frac{L^2}{f^2} \end{pmatrix} \tag{2.1}$$

Stability requires that as a particle repeatedly traverses alike cells stay contained in the beam. Therefore, the absolute value of the eigenvalues of the transfer matrix must be ≤ 1, or $|\mathrm{Tr}(M)| \leq 2$ [2]. For the FODO matrix above, this means that stable beam dynamics require $L/2 < f$. For a general transfer matrix parameterized by the Courant-Snyder parameters α, β, and γ, the stable motion is bounded by an ellipse in phase space x, x', the area of which is proportional to the beam emittance. Figure 2.1 shows the phase space ellipse.

RF cavities, dipole magnets and quadrupole FODO lattices can accelerate and store an energetic beam. The beam collisions produce Higgs bosons in a rare process: only one out of every 2 billion proton-proton collisions at the 13 TeV LHC will produce a Higgs boson. The number of $h \to \mathrm{bb}\tau\tau$ events detected is given by:

$$N = \int \mathcal{L}dt \times \sigma \times A \times \epsilon \tag{2.2}$$

where $\mathcal{L}$ is the instantaneous luminosity, σ is the cross section, A is the acceptance and ϵ is the efficiency. The quantity $\int \mathcal{L}dt$ is the integrated luminosity, and has units of $(\mathrm{length})^{-2}$. The integrated luminosity is typically discussed in terms of *barns* (b), where $\mathrm{b} = 1 \times 10^{-28}\mathrm{m}^2$.

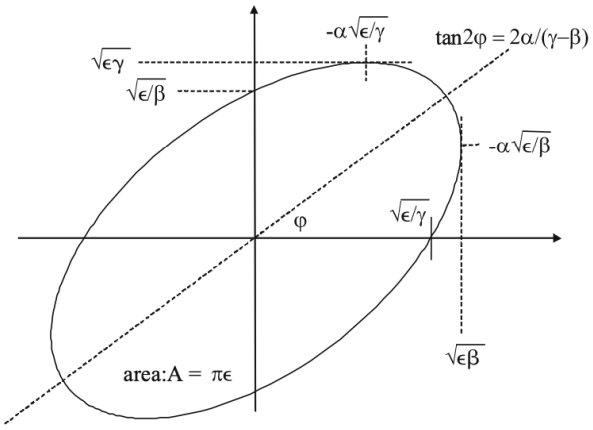


Figure 2.1: Phase space diagram for transverse beam dynamics in one dimension [2]. The x axis is translation and the y axis is momentum. The area of the ellipse is the beam emittance. α, β, and γ are the usual Courant-Snyder parameters.

Acceptance and efficiency are a function of the detector geometry and performance, as well as the cuts applied in data analysis. The number of events created by the collider $\int \mathcal{L}dt \times \sigma$ is increased by (i) running the experiment for a longer time (ii) increasing the cross section σ, and (iii) increasing the instantaneous luminosity $\mathcal{L}$. The cross section is determined by the kinematics of the desired amplitude; in this case, it is the diagram in Figure 1.5a (and similar diagrams). The Higgs production cross section scales with energy, so better RF cavities increase the cross section via a more energetic beam. Instantaneous luminosity of crossing beams is controlled by the amplitude function at the interaction point β^*. More precisely,

$$\mathcal{L} \propto \frac{n_1 n_2}{\sqrt{\epsilon_x \beta_x^* \epsilon_y \beta_y^*}} \tag{2.3}$$

where n_1 and n_2 are the number of particles in the beams and ϵ_x and ϵ_y are the emittances in the x and y direction. Instantaneous luminosity at a collider can therefore be increased by (i) increasing the number of particles in a bunch (ii) designing the storage ring to have low emittance (in other words, quiet transverse dynamics) and (iii) rotating the phase space ellipse at the interaction point to squeeze the beam transversely.

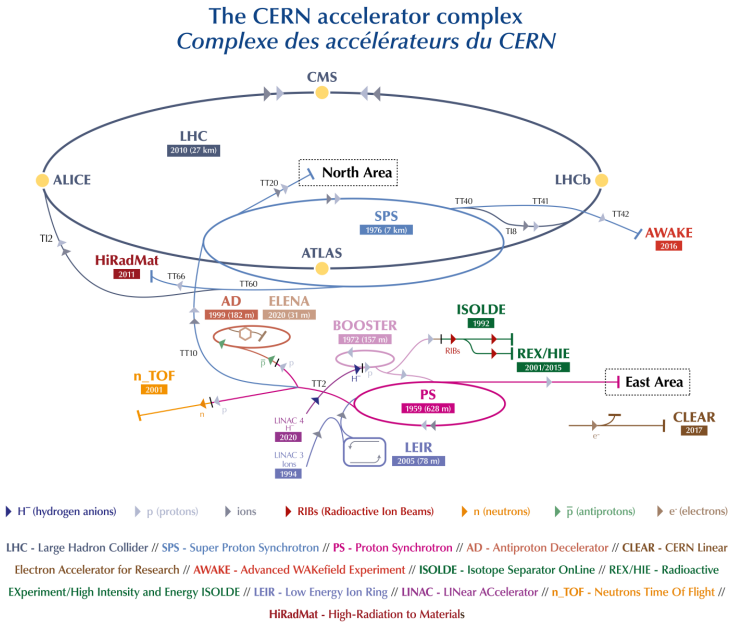


Figure 2.2: The CERN accelerator complex [3]. In this diagram, Linac2 has been replaced by Linac4, but the data collected for this thesis used the Linac2 accelerator. The four beam crossing sites are shown with the experiment name and a yellow circle on the LHC ring.

2.2 The Large Hadron Collider

The LHC [41] is a hadron colliding machine constructed in a 27 km underground tunnel at the European Center for Nuclear Research (CERN). The LHC first collided proton beams in 2009 at a center-of-mass energy of 7 TeV and has since raised the p-p collision energy to 13 TeV for the second main data taking period. The LHC is also an ion colliding machine, with lead-lead (Pb-Pb) ion collisions recorded at a maximum energy of 5 TeV per nucleon [42].

The LHC is supplied with protons from the Linac2[1]—Proton Synchrotron Booster (PSB)—Proton Synchrotron (PS)— Super-Proton Synchrotron (SPS) injector chain (see Figure 2.2). The LHC uses 1232 superconducting dipole magnets to bend proton beams around the 27 km ring. The magnets are cooled by super-fluid helium to 2 K and the working magnetic field is 8 T [41]. The lower operating temperature with respect to similar machines (Tevatron, RHIC

[1]Beginning in 2020, Linac2 was retired and replaced with Linac4 [43].

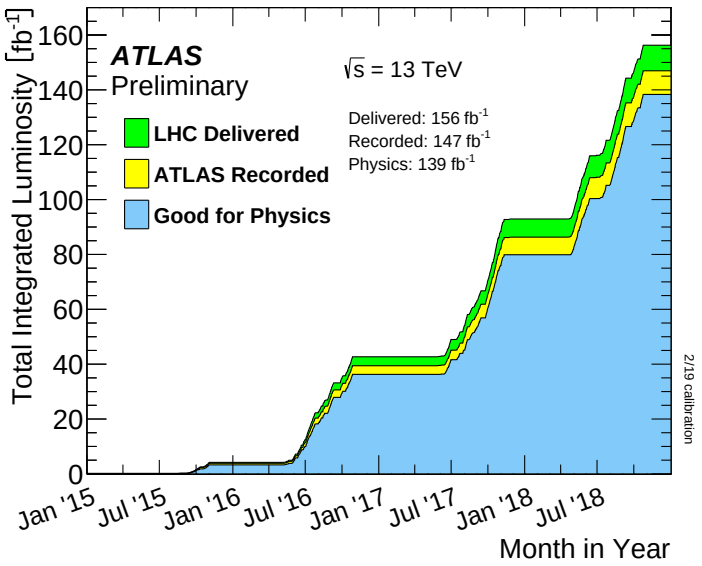


Figure 2.3: Integrated luminosity for the Run 2 data taking period delivered by the LHC (green), recorded by ATLAS (yellow) and good for ATLAS physics (blue) [4].

operate at 4.5 K) means that a smaller energy deposition can trigger a quench. Quadrupole magnets are also employed for focusing in a FODO cell arrangement, with dipoles in between the focusing and defocusing magnets. The LHC beams are bunched with a 25 ns bunch spacing, with each beam storing up to 2,808 bunches of 1.1×10^{11} protons [41].

Experiments and accelerator subsystems are in 8 caverns (Points) spaced around the 27 km tunnel. The LHC beams collide at Points 1 (ATLAS) [6], 2 (ALICE) [5], 5 (CMS) [44] and 8 (LHCb) [45]. Points 3 and 7 are for beam cleaning. Point 4 houses a 400 MHz RF system, and a beam dumping system is at Point 6. Both ATLAS and CMS are general-purpose detectors with nearly 4π solid angle coverage, which makes them well suited to search for new particles. LHCb is a specialized forward (high-η) detector to study CP violation and B mesons. ALICE is a detector designed to study the quark-gluon plasma produced in heavy ion collisions at the LHC, and its design is optimized for the large number of particles produced in heavy ion collisions.

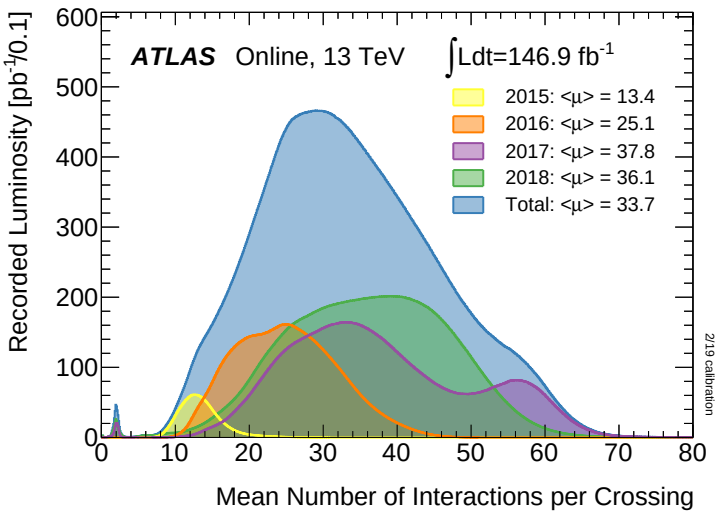


Figure 2.4: Mean number of interactions per bunch crossing for the Run 2 data taking period [4].

2.2.1 LHC Run 2

The LHC Run 2 lasted from 2015 to 2018 and data was collected in p-p collisions at an energy of 13 TeV. Figure 2.3 shows the delivered luminosity and the luminosity recorded by the ATLAS experiment. In total 139 fb^{-1} of integrated luminosity were recorded with an uncertainty of 0.83% [46]. Not all delivered luminosity is "good for physics" if there is a operational problem with the detector at the time of delivery.

When two bunches of 10^{11} protons collide, multiple interactions can occur. Figure 2.4 shows the average number of interactions per bunch crossing (also called pileup or $< \mu >$) over the Run 2 data taking period. If pileup is too high and many interactions appear nearly simultaneously in the detector, it is more difficult to reconstruct the event and associate each particle produced with the one of dozens of interactions. Pileup is calculated as

$$< \mu > = \frac{\mathcal{L} \times \sigma_{\text{inel}}}{f} \tag{2.4}$$

where $\mathcal{L}$ is the instantaneous luminosity of the LHC, σ_{inel} is the inelastic scattering cross section (for 13 TeV p-p collisions, 10^{11} pb) and f is the LHC bunch crossing frequency (40

MHz). The bunch luminosity is modulated by the beam parameters as discussed in Section 2.1. Because the inelastic scattering cross section and revolution frequency of the LHC are fixed, pileup is simply proportional to the instantaneous luminosity of the bunch. Therefore, the while the planned High-Luminosity LHC (HL LHC) will produce a higher number of "interesting" particle collisions (meaning they are useful for searches for new physics, like the one presented in this work), an unavoidable side effect is that the number of simultaneous "uninteresting" collisions will also increase. Pileup collisions must be discriminated against and provide a source of background and systematic uncertainty. The current amount of pileup, between 20 and 60 simultaneous interactions, is tolerable with the existing reconstruction algorithms.

2.3 The ATLAS Detector

2.3.1 Overview

ATLAS is a general purpose detector suited for new particle searches because of its high solid angle coverage and precise measurement of particle kinematics. The tracking detectors in ATLAS are the inner detector and muon spectrometer. A hybrid magnet system consisting of a solenoid and toroid bends the trajectories of charged particles in the ATLAS detector to aid in momentum measurement via tracking. The calorimeter systems in ATLAS are sampling calorimeters. A sampling calorimeter has alternating uninstrumented absorber material (usually with heavy nuclei such as lead or uranium) and instrumented material where charge or light is collected. By contrast a homogeneous calorimeter is composed of a single medium in which light or charge is collected and has the correct material properties to absorb the particle energy. The detector is depicted in Figure 2.5. The detector is cylindrical, with a central barrel region and two endcaps (the precise definition of the barrel vs endcap depends on the geometry of the particular detector subsystem).

The design resolutions of each subsystem, which characterize how precisely the energy

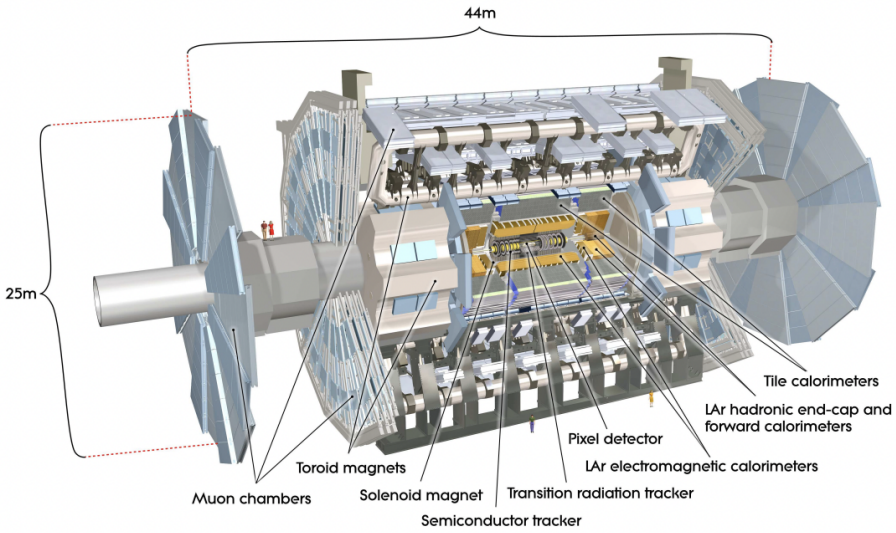


Figure 2.5: The ATLAS detector at the CERN LHC [5]. The detector is 25 m tall and 44 m long and weighs 7000 tonnes.

(or momentum) of a particle is measured by the detector, are summarized in Table 2.1. Resolution is important because measuring the 4-vector of the particles is one of the ways that signal and background are discriminated, as discussed in Section 1.4. Resolution is limited by the particle measured or the choice of detector technology, as will become clear in Sections 2.3.3-2.3.6. In general calorimetric energy resolutions will have some combination of a stochastic ($\sigma_E/E \propto 1/\sqrt{E}$), noise ($\sigma_E/E \propto 1/E$), and constant ($\sigma_E/E \propto 1$) terms. Calorimetric energy resolution therefore improves with higher energy particles because there are more samples recorded by the detector, which improves the contribution from stochastic and noise terms. In contrast, tracking resolutions degrade with increased momentum ($\sigma_{p_T}/p_T \propto p_T$) because the particles will bend less in the given magnetic field and the momentum measurement is via the track curvature.

The nominal interaction point is the origin of the ATLAS coordinate system. The beam pipe is along the z axis and the x and y axes are in the plan transverse to the beam. The positive x direction points to the center of the LHC and the positive y direction points up. The coordinate system is right-handed, so the positive z direction is tangent to the beam in the direction of Point 8 (anti-clockwise around the LHC ring in Figure 2.2). The usual

Detector subsystem	Design Resolution
Tracking	$\sigma_{p_T}/p_T = 0.05\%\ p_T \oplus 1\%$
EM Calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.33\ \text{GeV}/E \oplus 0.7\%$
Hadronic Calorimetry	
barrel and end-cap ($\lvert\eta\rvert < 3.2$)	$\sigma_E/E = 50\%/\sqrt{E} \oplus 1.0\ \text{GeV}/E \oplus 3.4\%$
forward ($3.1 < \lvert\eta\rvert < 4.9$)	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$
Muon Spectrometer	$\sigma_{p_T}/p_T = 10\%$ at 1 TeV

Table 2.1: Design resolution for various ATLAS detector subsystems [5, 19, 20]. All units of energy are in GeV. For the forward calorimeter, only a two-parameter resolution model is used [19].

azimuthal angle ϕ is in the xy plane and the polar angle θ is the angle with respect to the z axis. A useful variable, pseudorapidity, is defined as $\eta = -\ln\tan(\theta/2)$ for massless objects. Transverse momentum p_T is the projection of a momentum onto the transverse plane. The angular distance ΔR is defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. A brief summary of the main components of the ATLAS detector follows. In the course of my graduate study, I also worked on an upgrade to the muon spectrometer. Detector construction and testing for the muon spectrometer upgrade is summarized in Chapter III.

2.3.2 Magnet System

As discussed previously, identification of particle species and estimation of its energy and/or momentum is required in order to perform signal-background discrimination with the detector. Particle identification is performed in tracking detectors by measuring the bend of particles in a magnetic field. The ATLAS detector uses four large superconducting magnets to bend the paths of charged particles in the two tracking detectors: the inner detector and the muon spectrometer. The magnet system is 22 m in diameter and 26 m in length, and stores 1.6 GJ of energy [6].

A diagram of the magnet coils is shown in Figure 2.6. A thin solenoid magnet provides

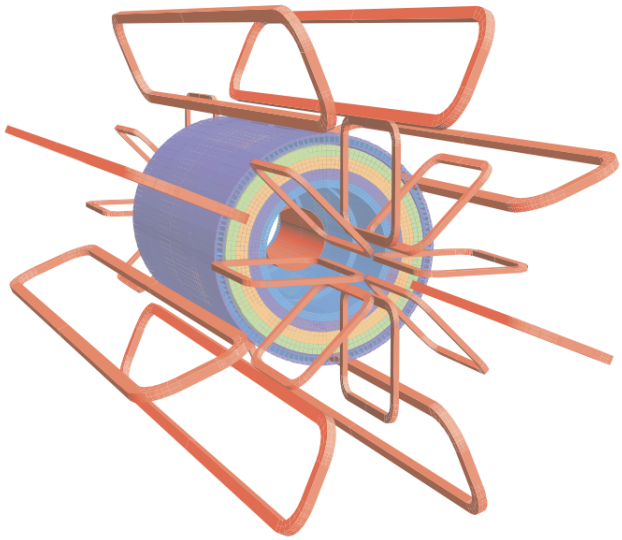


Figure 2.6: The geometry of the ATLAS magnet system. Magnet windings are shown in red, including the inner solenoid (central red cylinder) and three toroids (one barrel and two endcaps), which each have eight coils. The various layers of the tile calorimeter, including the return yoke, are shown in blue, yellow and green between the solenoid and toroid magnets [6].

a 2 T magnetic field in the $\hat{z}$ direction for the inner detector, and a field of 0.5 and 1 T in the $\hat{\phi}$ direction for the muon spectrometer barrel and endcaps, respectively.

The solenoid magnet must have as little material as possible to avoid particles depositing an appreciable fraction of their energy before reaching the calorimeters. The solenoid magnet has ≈ 0.66 radiation lengths of material in front of the calorimeters [6].

2.3.3 Inner Detector

The Inner Detector (ID) is the primary tracking detector of ATLAS. The ID is inside of the solenoid magnet's 2 T magnetic field. High detector granularity is required to measure the momentum of approximately 1000 particles per bunch crossing (25 ns). In addition to momentum measurements, the ID must resolve the primary vertex by associating each particle with one of the many interactions per bunch crossing. Three detector technologies are employed: silicon pixels, silicon strips in the SemiConductor Tracker (SCT), and straw

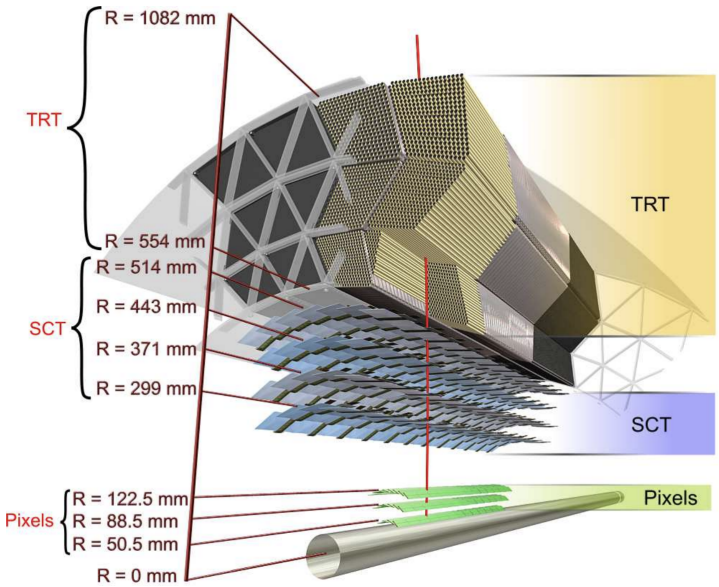


Figure 2.7: The Inner Detector [6]. Not shown is the Insertable B-Layer: an additional pixel layer added in May 2014 at R=33 mm from the beam axis after a smaller beam pipe was installed [7, 8, 9].

tubes in the Transition Radiation Tracker (TRT). There are 80.4 million pixels, 6.3 million strips, and 351,000 straw tubes [6]. A small slice in ϕ is shown in Figure 2.7. Similarly to the solenoid magnet, the ID must have as little material in front of the calorimeters as possible to avoid degrading calorimetric energy resolution. The ID has between 0.3 and 1.3 radiation lengths, varying with respect to $|\eta|$ [47].

The pixel sensors are 250 µm thick wafers of n-type silicon. When charged particles traverse the silicon electron-hole pairs are created in the bulk, which drift in an electric field and induce a voltage on readout electronics. The nominal pixel size is $50 \times 400\mu m^2$, but there are some edge cases were pixels are larger. The SCT uses strips of silicon $\approx$ 100 mm in size and $\geq 6cm$ in length with a pitch of 80 µm in the $\hat{\phi}$ direction. The SCT provides a low radius "space point" in the tracking, which allows for vertex reconstruction and pileup rejection.

The TRT is composed of straw tubes, the walls of which are thin films of Kapton and Polyurethane, which have different dielectric properties. The straw tubes are filled with a

Item	Intrinsic accuracy [µm]
Pixel	
Layer-0	10 $(R-\phi)$ 115 (z)
Layer-1 and -2	10 $(R-\phi)$ 115 (z)
Disks	10 $(R-\phi)$ 115 (R)
SCT	
Barrel	17 $(R-\phi)$ 580 (z)
Disks	17 $(R-\phi)$ 580 (R)
TRT	130

Table 2.2: Intrinsic resolutions of subdetectors in the ID [6].

$Xe/CO_2/O_2$ gas mixture. Transition radiation occurs when a particle crosses an interface due to a time-varying dipole between the charge and its image [48]. At some scales, the energy radiated is proportional to the incident energy, but a saturation occurs at high enough energies.

The ID can perform tracking for particles with $p_T > 0.5$ GeV with an energy resolution of ≈2%, which degrades to ≈17% at $p_T = 500$ GeV. The intrinsic position accuracy varies for the different subdetectors and in different dimensions. The accuracies are summarized in Table 2.2. Electron identification is performed up to 150 GeV of p_T [47]. In May 2014, the Insertable B-Layer (IBL), a new layer of pixels, was installed after the beam pipe was upgraded to be smaller. The IBL helps measure secondary vertices of particles with finite lifetimes, including b-mesons and τ leptons, both of high relevance for this search.

2.3.4 Electromagnetic Calorimeter

The energy loss of photons and electrons above 10 MeV in matter is due to Bremsstrahlung and the creation of electron-positron pairs [18, 48]. The cascade of Bremsstrahlung photons and electron-positron pairs creates an electromagnetic shower (EM shower). The characteristic length scale of this process is the radiation length, X_0, set by the atomic properties of the material. Detectors must be constructed with the radiation length in mind so that they capture the shower (nearly) completely. EM showers are initiated by both photons and

electrons/positrons. Electrons will have a matched track in the ID, and photons will not because they have no electric charge. The EM calorimeter detects electrons and photons in ATLAS.

The electromagnetic calorimeter is a lead-Liquid Argon (LAr) sampling calorimeter with an accordion absorber geometry, shown in Figure 2.8. In the barrel the lead plates have a thickness of 1.53 mm and the distance between plates is 2.1 mm. Readout electrodes are mounted on the lead absorber and the nominal HV for the barrel EM calorimeter is 2 kV [6]. The accordion geometry reduces charge collection time because the electrodes have low capacitance. Signal rise time can be decreased by a factor of 100 as compared to calorimeters with absorbers perpendicular to the incident charged particle, leading to better timing resolution [48]. The accordion geometry also lends itself naturally to full 2π azimuthal coverage, necessary for a general-purpose detector like ATLAS.

Any matter between the interaction point and EM calorimeter degrades the resolution of the calorimeter. As particles traverse the upstream material they stochastically deposit some of their energy in it, and some of the information of the particle energy is lost (if the process was not stochastic, it could be calibrated for; for example, the energy being simply the measured energy plus some number of GeV to account for upstream material). The amount of material in front of the EM calorimeter is shown in Figure 2.9 as a function of $|\eta|$. In the range $1.37 < |\eta| < 1.52$ a support structure means a significant amount of material is in front of the calorimeter. Reconstructed electrons in this eta region, known as the "crack" are vetoed. There is a minimum of 22 X_0 of material in the EM calorimeter.

The energy resolution of the EM calorimeter is

$$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.33 \text{ GeV}/E \oplus 0.7\%. \tag{2.5}$$

The choice of detector technology can have a significant affect on the energy resolution. For comparison, the CMS EM calorimeter has a resolution of $\sigma_E/E = 2.8\%/\sqrt{E} \oplus 0.128 \text{ GeV}/E \oplus 0.3\%$ [49]. CMS achieves this resolution by using a homogeneous crystal

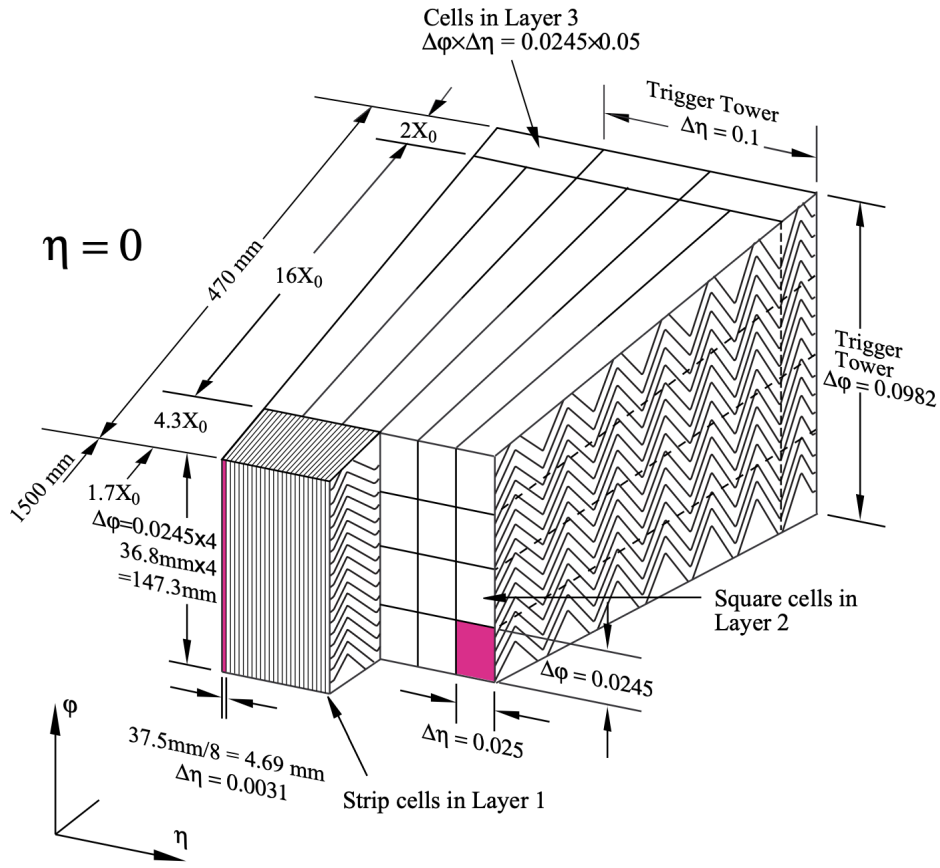


Figure 2.8: Diagram of cells in the ATLAS EM calorimeter [6]. There are 3 layers of cells with varying sizes. Charge is collected on electrodes lining the lead accordion.

calorimeter. Unlike the ATLAS calorimeter, which has an alternating serious of absorber and instrumented material, the CMS calorimeter has no dead volume from absorbers.

2.3.5 Hadronic Calorimeter

When a strongly interacting particle interacts with matter, a shower of secondary, tertiary, and further hadrons are formed until each individual hadron is stopped by ionization losses or absorbed in a nuclear process [48]. This process is called a hadronic shower. It is difficult to precisely reconstruct the energy deposited in a hadronic shower for several reasons. The energy resolved by a hadronic calorimeter will be incomplete because (i) some shower particles, like muons and neutrinos from pion decay, will escape the calorimeter and (ii)

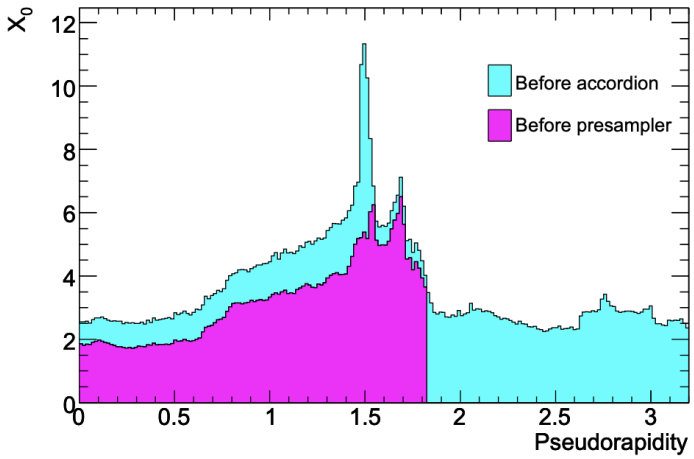


Figure 2.9: Amount of material in front of the EM calorimeter [6]. The crack between 1.37 and 1.52 is vetoed due to a significant amount of material from a support structure for other detector components.

depending on the detector geometry and composition, low energy γ rays, protons, neutrons, and other nuclear fragments produced in the bulk of the absorber may not be detected [48]. Pions will on average only produce about 80% as much visible energy in a sampling hadronic calorimeter than electrons with the same energy, and the visible fraction will change depending on the ratio of π^0 to $\pi^{\pm}$ in a given hadronic shower [48]. Therefore, a large (irreducible) uncertainty will arise from the measurement of hadronic shower ("jet") energy in this thesis.

Three independent calorimeters make up the ATLAS HCAL system. The tile calorimeter is a sampling calorimeter with plastic scintillator tiles as the instrumented volume and steel absorbers. The LAr hadronic end-cap calorimeter is located behind the EM calorimeter and shares a LAr cryostat. The LAr forward calorimeter covers the range $3.1 < |\eta| < 4.9$ and has a higher number of nuclear interaction lengths than the barrel calorimeters.

The tile calorimeter has a nearly periodic steel-scintillator structure with a ratio 4.7:1 by volume [5]. The approximate cell size is 2 cm of steel in the z direction. The geometry of the tile calorimeter is shown in Figure 2.10. The scintillator is oriented radially to allow for seamless azimuthal coverage [5]. The steel girder at the top of each segment provides a return yoke for the magnetic flux of the solenoid magnet.

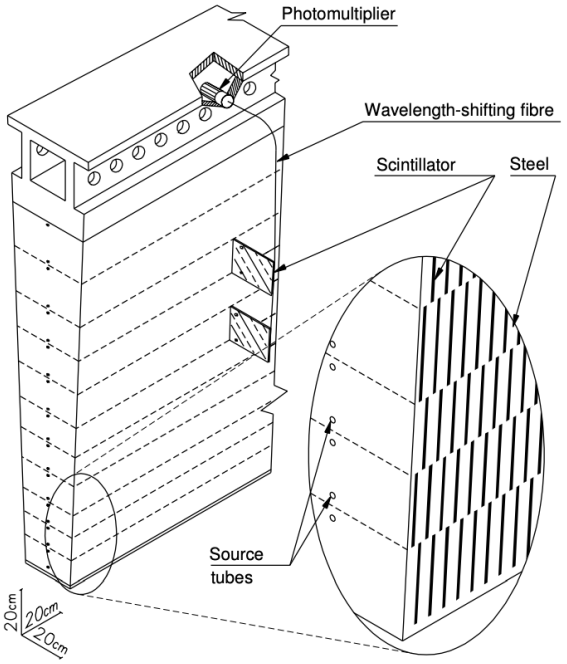


Figure 2.10: A single segment of the ATLAS tile calorimeter [6]. There are 64 azimuthal segments for full coverage.

The hadronic end-cap calorimeter is a sampling calorimeter with copper absorber and LAr for the active volume. The hadronic forward calorimeter uses a tungsten absorber due to the high rate of particles at high η. Both hadronic calorimeters use a smaller LAr gap than the EM calorimeter (2 mm for EMCal and 0.2-0.5 mm for HCal [6]).

The resolution of the HCAL is $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ in the barrel and $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ in the endcap. The resolution of the HCAL is 5-10x worse than the EM calorimeter. The EM calorimeter resolution was driven by the amount of absorber, and an example was shown that a homogeneous calorimeter (with zero absorber) would improve the resolution drastically. The HCAL resolution is driven by both (i) the intrinsic difficulties in measuring hadronic showers, discussed above; and (ii) the extra absorber needed to contain the energy deposition. It is uncommon to have a homogeneous hadronic calorimeter because on an experiment with jets as energetic as ATLAS the size would be much larger than permissible.

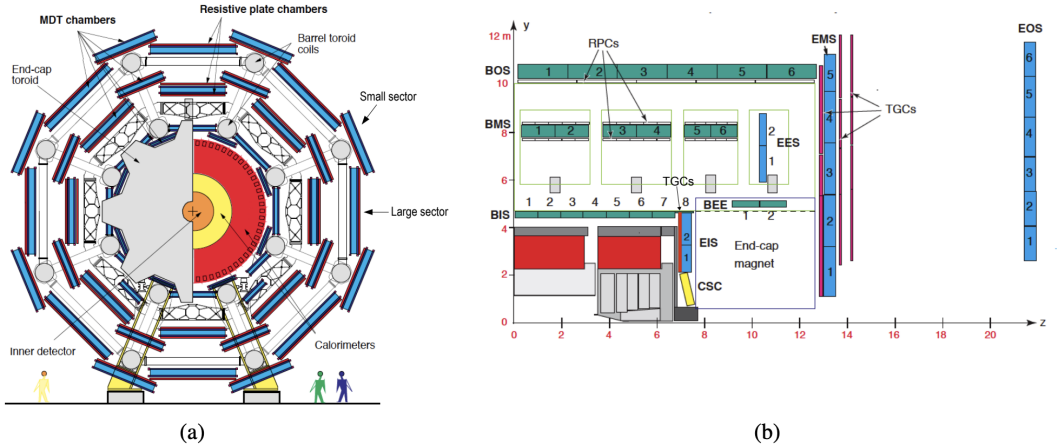


Figure 2.11: The ATLAS muon spectrometer [10]. (a) Barrel RPCs and MDT chambers depicted with the toroid magnets and calorimeters. Alternating large and small MDT chambers are shown. (b) Muon spectrometer layout in the (z, y) plane for a small azimuthal sector.

2.3.6 Muon Spectrometer

Muons are nearly minimum ionizing particles over a large energy range [18]. As their energy loss is so minimal, they will pass through the lead and steel absorbers of the EM and hadronic calorimeters. The momentum of a muon must be determined by its bend under the Lorentz force in the magnetic field. This is achieved in an array of position measuring detectors at large radius and with its own set of toroid magnets.

The ATLAS muon spectrometer consists of Thin Gap Chambers (TGCs), RPCs, Cathode Strip Chambers (CSCs) and MDT chambers. RPCs and TGCs are used for muon triggering and determining the second coordinate[2] for the precision trackers: CSCs and MDTs. Chambers are organized in three concentric stations and into barrel and end-cap regions. Small and large sectors alternate azimuthally and overlap for full coverage. A diagram of the muon spectrometer is shown in Figure 2.11.

In order to achieve high spatial resolution, the relative alignment of neighboring muon chambers (and between concentric stations) is measured to a precision of 30 μm [6]. An op-

[2]Second coordinate determination refers the fact that MDTs and CSCs do not record 3D tracking information. Hits on RPCs and TGCs (for the barrel and end-cap, respectively) measure the coordinate orthogonal to precision trackers.

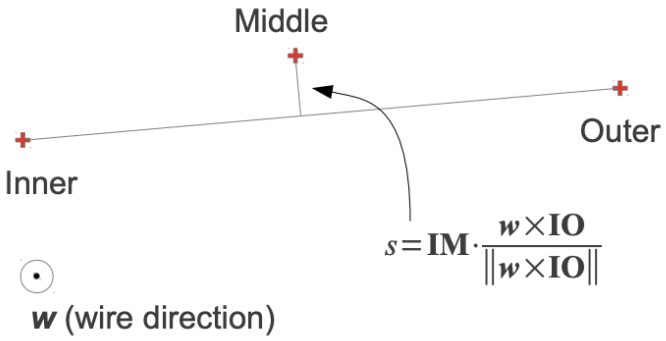


Figure 2.12: Definition of the sagitta [11]. For a negatively charged muon traversing the inner, middle, and outer stations (in that order), a magnetic field into the page will produce the observed deflection.

tical alignment system is used to measure the deformation of chambers and relative position of nearby chambers. In the full Run 2 data set, a MDT single-hit resolution of 81.7 ± 2.2 μm was observed [50].

Muon momentum is calculated from the sagitta (see Figure 2.12) of a bending muon track as it intersects the three precision stations. When the toroid magnet is off (and muons travel in a straight line), the performance of the optical alignment system is calibrated. When the magnet is on, the sagitta measures the energy of the muon. The energy resolution of reconstructed muons is dominated by the ID resolution at low p_T ($\approx$2% at $p_T = 0.5$ GeV) and dominated by the spatial resolution of MDT chambers at high p_T ($\approx$10% at $p_T = 1$ TeV).

2.3.6.1 Phase-II Upgrade of the ATLAS Muon Spectrometer

An upgrade of the LHC to increase the instantaneous luminosity by a factor of 5 is planned to take place from 2026-2028 [51]. In accordance, many detector subsystems must be upgraded. One such system is the ATLAS muon spectrometer. A new RPC layer will be added for muon triggering. The MDT chambers closest to the magnet toroids must be reduced in size to make room for the additional RPC layer. In total 96 MDT chambers will be replaced with Small Monitored Drift Tube (sMDT) chambers [52]. sMDT chambers

are designed such that they have nearly identical operating characteristics to MDTs. This upgrade will increase the trigger efficiency for muons. A significant component of this thesis is the work conducted for the muon spectrometer upgrade, including novel methods for characterizing drift chamber performance. More information on sMDT chambers produced at the University of Michigan is presented in Chapter III.

2.3.7 Trigger and Data Acquisition

The LHC collides beams every 25ns, so there are on the order of 10^7 bunch crossings per second. The ATLAS detector has approximately 100 million detector channels. It is not feasible to record and analyze the full results of every collision. Thus, a trigger and data acquisition (TDAQ) system is used on ATLAS to sort through data in real time and keep only those events which are most interesting, reducing the data rate from 10^7 to 10^3 per second. The data set for this thesis uses triggers designed to record events with high p_T leptons. Information from the EM calorimeter and muon spectrometer is used to decide whether the data from all detector systems is recorded.

Electron triggers apply dedicated fast reconstruction algorithms which include various isolation and E_T cuts. Figure 2.13 shows the single electron trigger efficiency as a function of transverse energy and pseudorapidity, which is generally above 95%. In this search single electron triggers are used to detect electrons which are the products of tau lepton decays.

In the muon spectrometer, triggering is performed with information from RPCs and TGCs, which are prompt detectors. Figure 2.14 shows the barrel trigger efficiency with respect to muon p_T and number of interactions per bunch crossing. Muon trigger efficiency is roughly flat above the appropriate offline p_T threshold (trigger threshold +1 GeV), but the high p_{T} triggers used in this work saturate at about 70% muon trigger efficiency[3]. The aforementioned work on the muon spectrometer upgrade will help improve the muon trigger efficiency for future searches and measurements with ATLAS. With increasing an increasing

[3]The barrel covers the region $|\eta| < 1.05$. In the endcap ($1.05 < |\eta| < 2.4$) the trigger efficiency saturates at about 90 GeV[53]

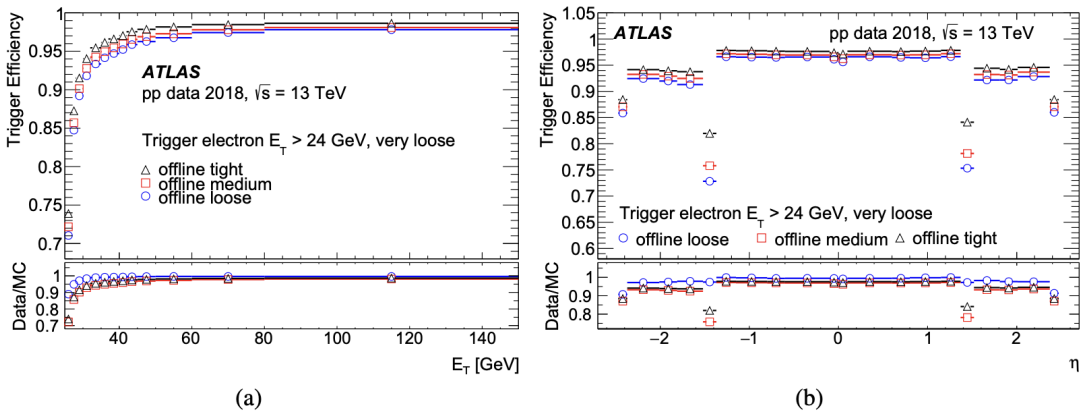


Figure 2.13: Trigger efficiency for single electron triggers used in this search [12].

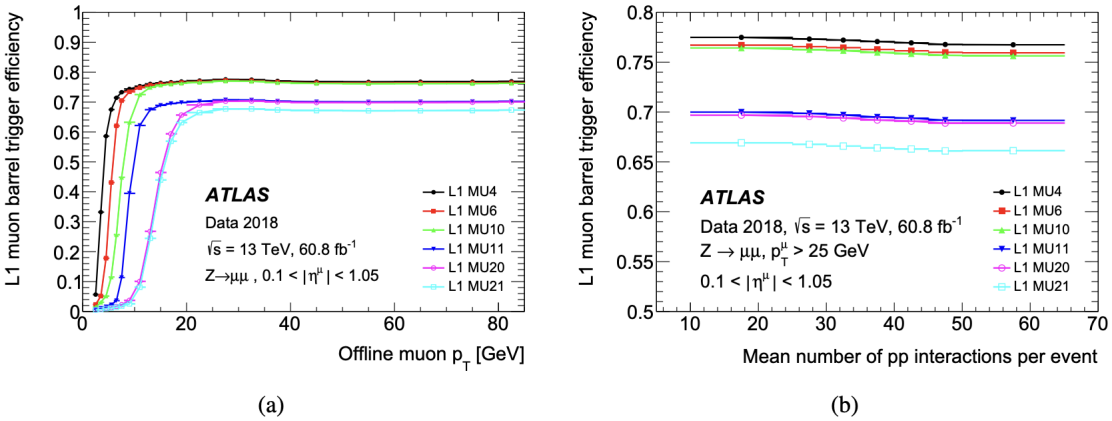


Figure 2.14: Muon trigger efficiency measured in the barrel with RPCs [10] with respect to (a) offline muon p_T and (b) mean number of interactions per bunch crossing (pileup). In the endcap ($1.05 < |\eta| < 2.4$) the trigger efficiency plateau is approximately 90%.

number of interactions per bunch crossing the trigger efficiency degrades slightly because of the effective dead time produced by latency. In this search single muon triggers are used to detect muons which are the products of tau lepton decays. Mixed triggers requiring both an electron and muon are also used to cover more lepton p_T phase space in the $e\mu$ channel.

CHAPTER III

Testing of sMDT Chambers for the ATLAS Phase-II Upgrade

The ATLAS detector described in Chapter 2 is constantly being upgraded and improved. This chapter describes a major upgrade to the central muon system which will be very important for maintaining and enhancing muon trigger efficiency in future data taking and any future version of this search. Discussion of the search itself resumes in Chapter IV.

3.1 Introduction

The sensitivity of the search presented in this thesis is limited primarily by the statistics of the data collected. Indeed, many similar searches and other precision probes of the Higgs sector are also limited by statistics. As the ATLAS collaboration and the LHC look into the future, upgrades to instrumentation will maximize the potential for discovery. One limiting factor for ATLAS specifically is the barrel muon trigger efficiency, shown in Figure 2.14. If the ATLAS L1 muon trigger efficiency was improved to the performance observed by CMS [158], statistics-limited studies dependent on single-muon triggers would increase in sensitivity by a factor of ≈ 1.15. Therefore, improving the L1 muon trigger efficiency is a well-motivated detector upgrade.

In the ATLAS Phase II upgrade, scheduled for 2026-28, an extra layer of small-gap RPC

chambers will be added to the barrel inner layer to improve the efficiency of the barrel muon trigger [159]. The Run 2 trigger efficiency is shown in Figure 2.14. In order to accommodate the new RPCs in the upgrade, the "small" sector MDTs will be replaced with small-diameter MDTs (sMDTs), which are are reduced in size compared to MDTs. The "large" sectors already have space for the new RPCs, so the current MDTs will remain.

The sMDT chamber design was developed at the Munich Max Planck Institute (MPI) [160] and use drift tubes of 7.5 mm radius, compared to 15 mm for the original MDTs. Aside from the difference in tube radius the ATLAS MDT and sMDT drift tubes are very similar. Both use aluminum tubes of 400 µm wall thickness, 50 µm diameter gold-plated W-Rh wires, and a drift gas admixture ($Ar:CO_2$, 93:7 %) at 3 bar absolute pressure. To keep the gas gain (2×10^4) identical, the operating high voltage is 2730 V for sMDTs compared to 3080 V for MDTs. These operating specifications were chosen in order for the nominal time-to-space transfer functions (r(t) functions) for the sMDT and MDT chambers to agree as closely as possible. A summary of the design sMDT chambers with a comparison MDTs can be found at [160]. Table 3.1 details the operating parameters of sMDT tubes. The University of Michigan and MPI Munich share the production of sMDT chambers for the ATLAS Phase-II muon detector upgrade.

Stringent quality controls of all systems under realistic operating conditions must be met to construct precision detectors. sMDT chambers must have high spatial resolution and efficiency to allow an accurate momentum measurement using the sagitta of muon tracks in the ATLAS muon spectrometer. The sMDT design single-hit resolution is 106 µm [160]. The resolution for MDTs is 81 µm [50], so it might seem at face value like the momentum resolution of the muon spectrometer would be degraded. However, a new low-noise ADC chip is also in development. The low-noise chip will allow a lower operating threshold, reducing jitter from signal rise time and increasing the resolution of sMDT chambers up to ≈ 83 µm. The efficiency for muons passing through the active gas volume must be nearly 100%. The only expected drop in efficiency is due to geometric acceptance from finite tube wall thickness.

The chambers must also have sufficiently low noise. Noise can cause the detector to "miss" a coincident hit from a real muon due to dead time or can create ambiguity in reconstructing tracks if the reconstruction algorithm must choose between nearby fake and real hits. This chapter summarizes tube and chamber production and discusses the methodology for quality control testing, including efficiency, noise rate, and resolution measurements.

Section 3.2 summarizes the sMDT tube construction, chamber construction and electronics installation. Section 3.3 describes the experimental setup and the procedure for quality control tests, including deriving the r(t) function for and fitting muon tracks. Section 3.4 describes the multiple scattering correction to the resolution, and Section 3.5 describes the systematic uncertainties considered for resolution and efficiency measurements. Measured resolution, efficiency, and noise rate results are reported in Section 3.6. Figures may refer to both the Barrel Middle Gap (BMG) prototype detector and the Barrel Inner Small (BIS) chambers. Results between the two chambers are consistent, and important deviations are noted.

3.2 sMDT Construction

3.2.1 Tube construction

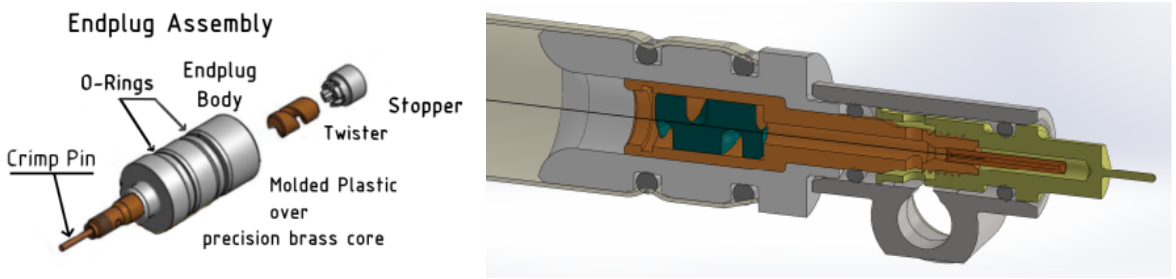
Drift tubes are constructed at the University of Michigan [16] and Michigan State University. Table 3.1 shows the design and operating parameters of the drift tubes. Quality control tests for tubes include tube straightness, length, tension, gas leakage, and dark current. Straightness tests are required so that the assembled chamber meets the mechanical precision requirement of tube position being accurate to within 20 µm RMS. If tubes have widely different lengths, then front-end electronics will not properly mount onto the chamber. If wire tension is too high or too low wires could break or slip, causing the tube to become inoperable. Similarly, a gas leak or high dark current could render a tube effectively inoperable and it would need to be disconnected from the HV system if it is installed on a

Parameter	Value (sMDT)
Tube material	Aluminium AW6060-T6/AlMgSi
Tube surface	Surtec 650 chromatization
Tube outer diameter	15.000 mm
Tube wall thickness	0.4 mm
Tube length	1615 mm
Tube straightness	0.8 mm / tube
Wire material	W-Re (97:3)
Wire diameter	50 µm
Wire resistance	44 Ω/m
Wire tension	350±15 g
Gas mixture	Ar:CO_2 (93:7)
Gas pressure	3 bar (abs.)
Gas leak rate limit	$< 1 \times 10^8 \frac{\text{mbar} \times \text{cm}^3}{\text{sec}}$ per tube
Gas gain	2×10^4
Wire potential	2730 V

Table 3.1: sMDT tube materials and operating parameters [16].

chamber. All of the quality control tests discussed are designed to prevent inoperable tubes from being installed on chambers.

Before assembly each tube is measured for straightness. If tubes are bent too much, then they will not fit together in the compact lattice and there will be some misalignment in the chamber. Tubes are measured for straightness using a microscope mounted on a beam known to have a maximum deviation < 0.05 mm. Approximately 6% of raw Aluminium tubes in early shipments had a maximum bend > 0.8 mm. Feedback was provided to the



(a) (b)

Figure 3.1: (a) Exploded diagram of endplug assembly. (b) Cross-section of assembled tube with signal-cap (yellow), gas manifold (gray) and twister (blue) [16].

manufacturer and newer shipments of tubes have a straightness failure rate of 2%. Tubes passing the straightness check are cleaned with isopropyl alcohol to prepare for assembly.

Figure 3.1 shows the anatomy of a completed drift tube. The wire is strung through the tube and the two endplugs. The endplug assembly seals the gas volume with O-rings and provides precision surfaces on both ends of the tube. The twister is a precision brass locator that keeps the wire centered in the tube. The brass core is a precision surface to locate the tube when it is placed within a chamber. The wire is crimped on one end first, then brought under tension to 350 grams, and then the other end is crimped. The tube must not be over-tensioned or else the wire is liable to break, but also must not be under-tensioned so that the wire does not sag and degrade the position resolution of the tube. After the wire is tensioned the endplug is swaged (see the indents in the tube wall in Figure 3.1b). The swaging must also be calibrated: too deep a swage can crack the tube body, but too shallow a swage can cause the tension to drop.

Assembled tubes must undergo quality control checks for tension, gas leakage and dark current before chamber assembly. The tension is tested to be within 350±15 grams at assembly and two weeks after to ensure that the wire is not slipping over time. The tension is tested by running a current through the wire while the tube is placed in a permanent magnet. The Lorentz force induces transverse oscillations of the wire; effectively: the string is plucked. The induced current is measured and the frequency harmonics determine the tension of the wire [161]. The observed wire tension is 357±6 g, and the tension on average drops by 2 grams after 2 weeks [16].

Tubes are checked for gas leaks when pressurized with 3 bars absolute pressure. Pressurized tubes are inserted into a vacuum vessel and any significant leakage of Helium into the vacuum vessel will be detected by the sniffer. No tubes assembled by the University of Michigan fail the gas leak check.

Tubes are checked for dark current to be < 2 nA when pressurized with 3 bar of the 93:7 Ar:CO_2 mixture and at 2800 V (slightly higher than the operating voltage of 2730 V). A

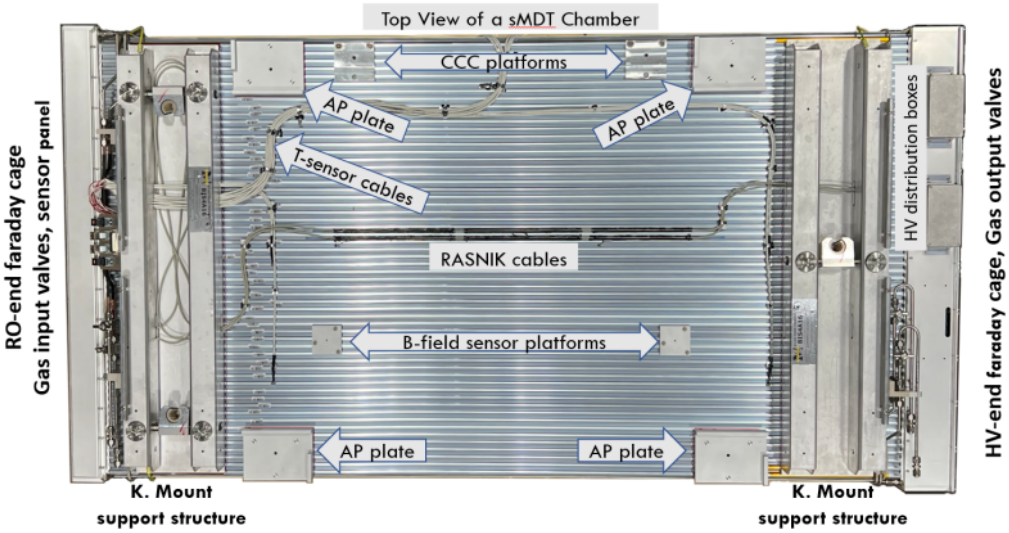
small real current is drawn from the tube due to ionizing radiation from cosmic-rays, but some tubes draw large (100s of nA) currents. The large current is due to dirt or debris inside the tube which the isopropyl alcohol cleaning did not remove. Dark current is remedied with negative HV treatment: the tubes are treated with a reverse polarity HV (-3 kV) for 60 minutes to burn off any debris. All tubes assembled at Michigan reached a dark current of < 2 nA with negative high voltage treatment.

In general, tubes passing straightness, length, tension, gas leak, and dark current tests are ready for chamber construction. However, a small number of various defects were found after tubes were affixed in chambers. When a defect is found, the only solution is often to remove the wire from the tube and render the tube completely inoperable. This is discussed in the following sections as appropriate.

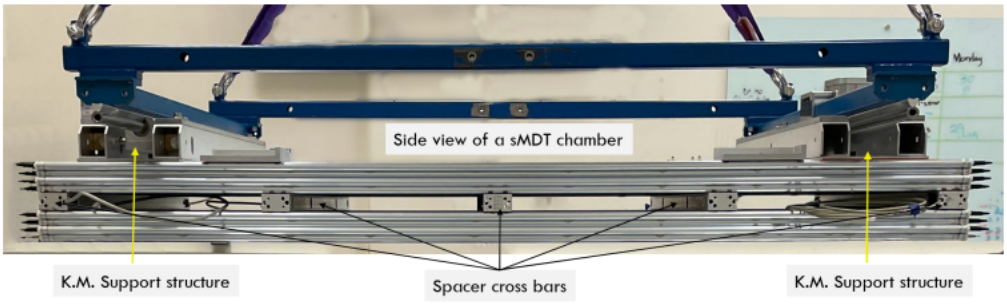
3.2.2 Chamber construction

Drift tubes passing all quality control tests are assembled into chambers. Each chamber has 8 layers organized into 2 multilayers (ML) with 4 layers per ML. Each layer has either 70 (BIS1) or 58 (BIS2-6) tubes per layer. Assembled chambers have various services mounted on them, including HV distribution, platforms for optical alignment system, temperature sensors and kinematic mounts. An in-plane alignment system is in the space between the two multilayers. The alignment system is used to measure and correct for chamber deformation when fitting muon tracks. A fully assembled chamber is shown in Figure 3.2.

Chambers are assembled in a clean room on a 3×7 m^2 granite table flat to within 25 microns. Tubes are laid in precision jigging and fastened to one another with an automatic epoxy dispensing machine mounted on the granite table. The granite table, glue-machine, and a glued sMDT chamber can be seen in Figure 3.3. The precision jigging combs constrain the location of the tube endplugs on the HV and RO ends of the chamber. Figure 3.4 shows the construction of the jigging on the granite table and the comb locating the tube position using the precision reference surface. The horizontal pitch is 15.1 mm (tube outer diameter



(a)



(b)

Figure 3.2: An assembled sMDT chamber (a) top view with platforms, kinematic mounts, sensor cables, HV distribution boxes and Faraday cages. (b) side with kinematic mounts attached to crane (blue) and spacer cross bars visible between the two MLs.

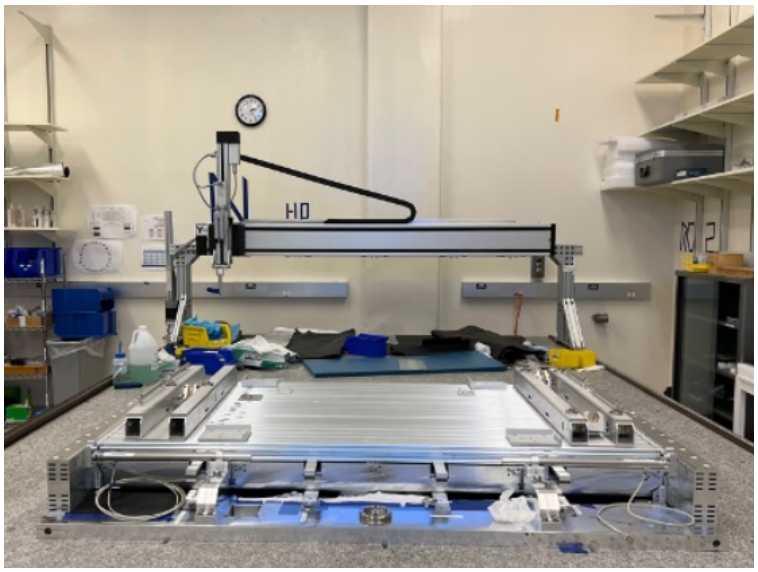
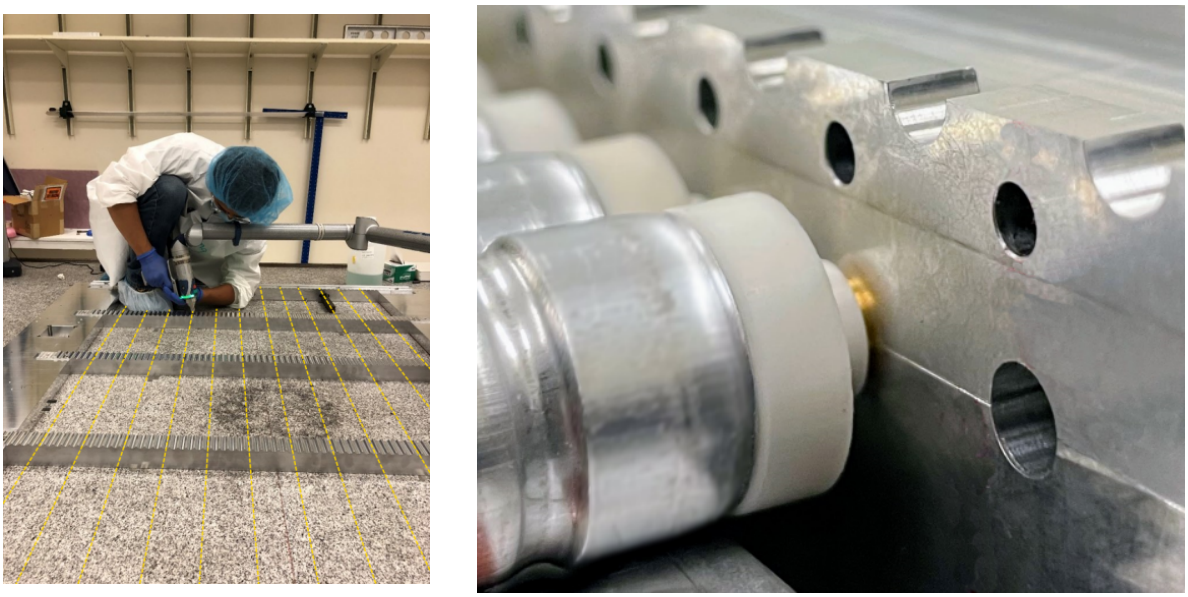


Figure 3.3: Automatic glue-machine mounted on the granite table. A fully-glued sMDT chamber sits in the precision jigging.

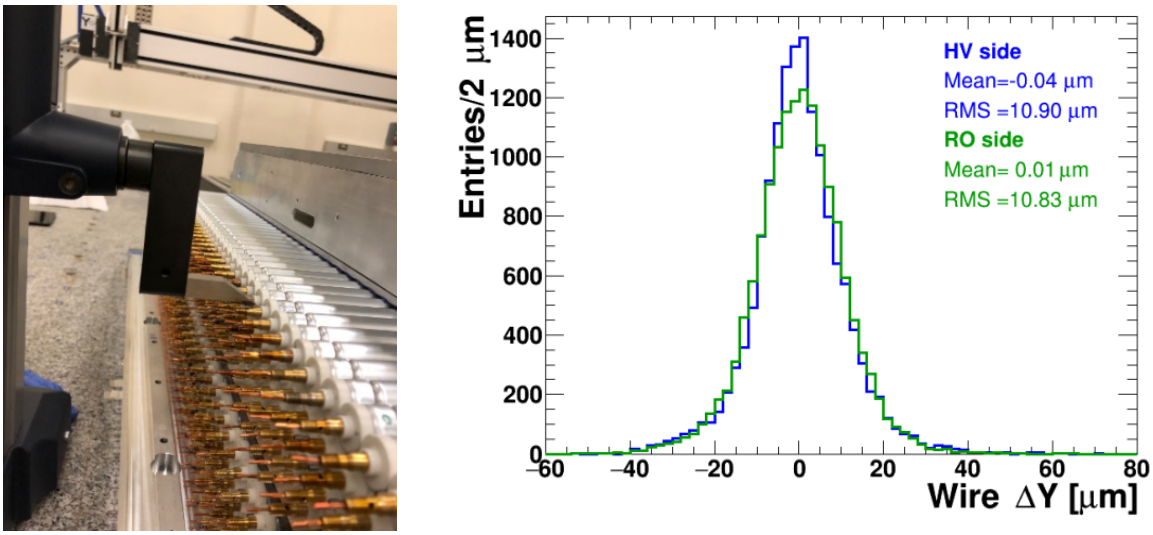
is 15.0 mm). The vertical pitch is 13.077 mm. A height gauge is used to measure the tube positioning after each layer is glued. The horizontal pitch is much more constrained by the precision jigging (to within ±4 µm), so the best measure of precision construction is the vertical pitch. The specification is for vertical pitch to have an RMS width 20 µm. For the first 30 sMDT chambers, the measured RMS was 10.9 µm on the HV side and 10.8 µm on the RO side, as shown in Figure 3.5.

An in-plane alignment system is installed inside the spacer frame between the two multilayers. The in-plane system consists of 4 RASNIK lines, two diagonal across the spacer frame and two along the length of the tubes at the edge of the spacer frame. Figure 3.6 diagrams one RASNIK line with its backlit coded mask, projection lens, and CCD sensor. The RASNIK system is calibrated for each chamber on the granite table, and then chamber deformation can be measured off the granite by subtracting the CCD image from the reference image. Figure 3.7 shows the measured chamber deformation for the first 30 modules. The X dimension is transverse to the layer of tubes and there is very little deviation. The Y direction is perpendicular to the layer of tubes and the chamber is much less rigid in this direction. It is immediately apparent that in general a correction for chamber deformation



(a) (b)

Figure 3.4: (a) The precision comb is assembled on the granite table. (b) The comb is shown with a partially assembled sMDT chamber. The comb constrains the horizontal position to ±4 µm and the vertical position to ≈10 µm RMS.



(a) (b)

Figure 3.5: (a) The height gauge in use on the granite table. (b) Results from height gauge measurements from 30 sMDT chambers. The RMS width is below the specification of 20 µm.

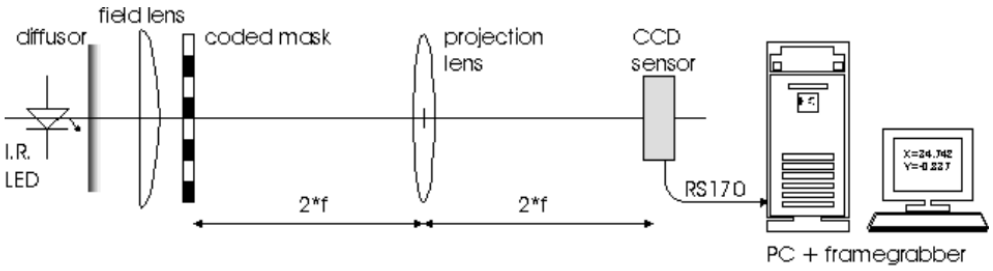


Figure 3.6: Diagram of a single RASNIK line. The RASNIK system is sensitive to movement of the mask, lens, and CCD sensor to measure chamber deformation.

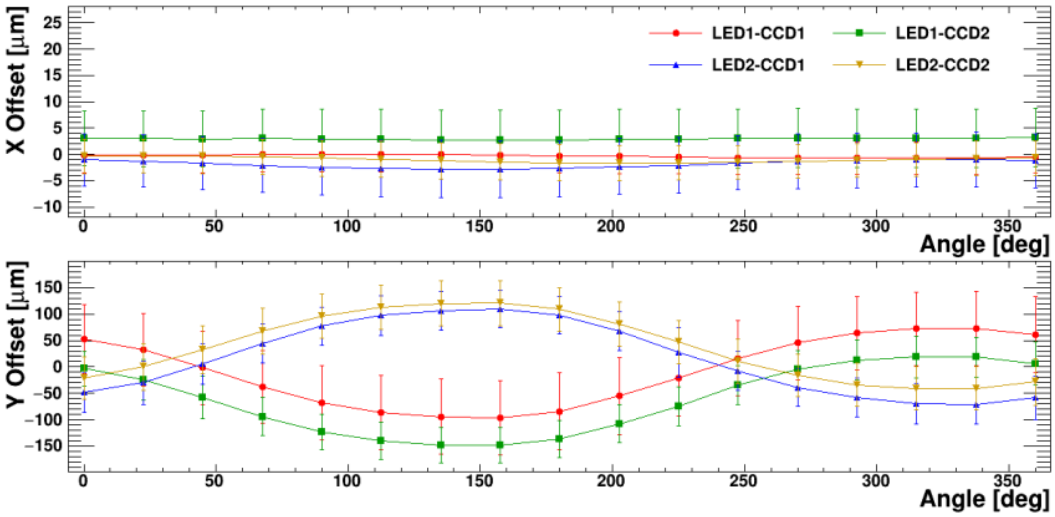
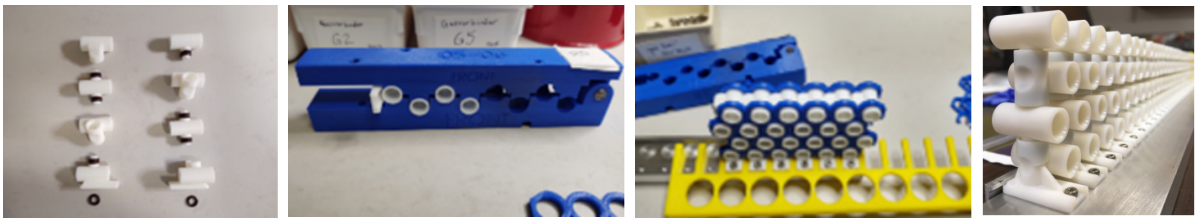


Figure 3.7: Measured chamber deformation as the chamber is rotated through 360°. The error bars are the standard deviation of the measurements from 30 chambers.

must be applied because the Y deviation is on the order of 100 µm and the design resolution is 106 µm. However, for the resolution tests conducted in this thesis the chamber is always laid flat and the effect of the deformation is minimal. When mounted in the muon spectrometer, chambers will be at various angles and deformation must be considered.

Temperature sensors are attached using thermal paste epoxy in twelve positions on the top and bottom of the chamber. The temperature is read-out by checking the voltage drop across a resistor which is calibrated with respect to temperature. The r(t) time-to-radius transfer function is dependent on the gas temperature, and a correction can be derived [50]. Temperature cables are routed to the RO end and a plate is installed on the RO end to make



(a) (b) (c) (d)

Figure 3.8: Assembly of the gas distribution manifold. (a) two gas stacks (with opposite chirality) shown before assembly and with o-rings. (b) Gas stack assembly in 3D printed jigging. (c) gas stacks being mounted on a gas bar with more jigging. (d) fully assembled gas bar. Four gas bars are used per chamber.

connections to the RASNIK system and temperature readout system (see "sensor plate" in Figure 3.2a).

The gas system is assembled on the chamber off of the granite table. Figure 3.8 shows the assembly process for the gas bar. Four gas bars are assembled per chamber for the RO and HV side for each ML. Each plastic part accepts a signal pin from a single tube. The gas bars have 70 (58) holes for BIS1 (BIS2-6), corresponding to the number of tubes per layer. Each stack of 4 plastic parts distributes gas to the four layers in each ML. An important finding in chamber testing was that gas blockages were occurring in early production. This could be seen easily because a set of 4 tubes in a stack had nearly zero cosmic-ray hits. This was due to metal "hanging chads" from the machining of the holes in the gas bars. On the gas inlet side, the gas could push the piece of metal back into place and fully block the flow of gas into a set of 4 stacked tubes. The gas bar cleaning procedure was updated to be more thorough, and the gas flow problems were fixed.

The fully assembled gas distribution system must meet ATLAS specification for leak rate of 1×10^{-5}[mbar liter/s]$\times(2N_{\text{tube}})$, where $2N_{\text{tube}}$ denotes the total number of endplugs present. This specification is equivalent to 0.288 mbar/hour for each ML. Since the pressure reading depends on the gas temperature, the installed temperature sensors are used to correct the pressure measurements. The correction is applied as

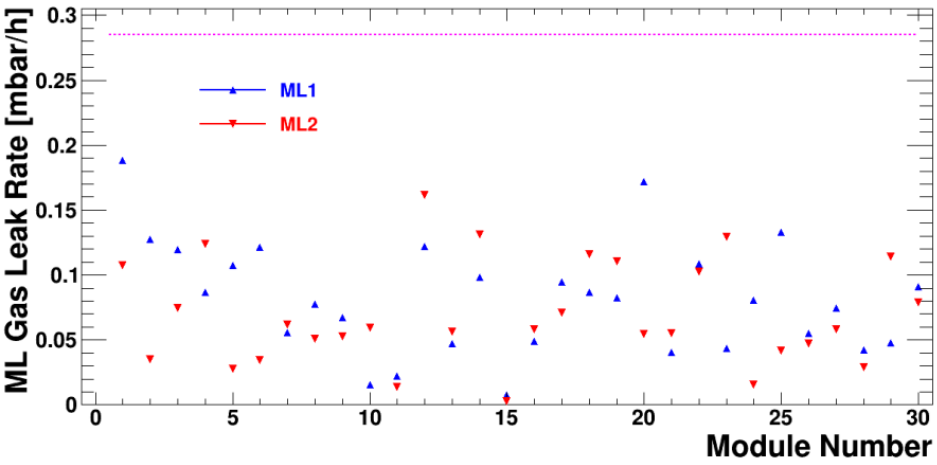


Figure 3.9: ML gas leak rate for 30 sMDT chambers. The dashed line is the maximum acceptable gas leak rate.

$$\left(\frac{\Delta P}{\Delta t}\right)_{\text{norm}} = \frac{\Delta P_{\text{corr}}}{\Delta t} \tag{3.1}$$

$$\Delta P_{\text{corr}} = P_f \frac{T_{\text{ref}}}{T_f} - P_i \frac{T_{\text{ref}}}{T_i} \tag{3.2}$$

where ΔP_{corr} is the corrected change in pressure over a time interval Δt; P_i, P_f, T_i, T_f are the initial and final pressures and temperatures, and $T_{\text{ref}} = 273.15°$ K is used as a reference temperature. Figure 3.9 shows the recorded leak rates for 30 sMDT chambers. All MLs tested pass the leak rate specification.

3.2.3 Electronics installation

Specially designed electronics, including Application-Specific-Integrated-Circuits (ASICs), are needed to maintain high voltage on the wires and read-out the change in potential on a wire (which operates at 2730 V). Grounding is also taken into careful consideration to provide short paths to ground and avoid loops, which cause interference between channels. Figure 3.10 shows a schematic diagram of the electrical connections for a single MDT tube. The Hedgehog (HH) boards are essentially high- and low-pass filters so that the AC component

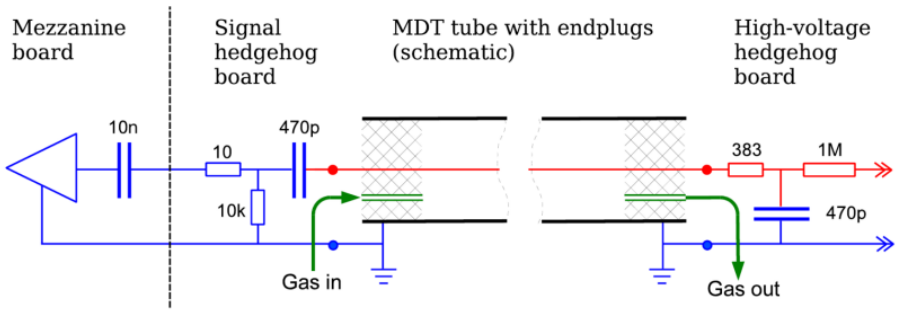


Figure 3.10: Diagram of electrical connections for a single MDT tube [17]. Resistor values are in Ohms. For sMDTs, the 383 Ω resistor is replaced with a 330 Ω resistor.

can be extracted on the RO side and the DC HV can be distributed on the HV side.

Before any electrical components are installed, the grounding and Faraday cage are assembled. Grounding screws are installed in the small gaps between tubes. The threads of the grounding screws make electrical contact with the exterior of the three neighboring tubes. The Faraday cage surrounds the HH cards and prevents outside noise from being picked up by the RO electronics. Small holes in the Faraday cage allow the HV lines in and connection to the FE electronics.

The safety grounding cables connect the gas bars, the support structure and the side paneling together so that they have a common ground without creating a ground loop. The grounding cables are installed before the Faraday cage side plates for access to the gas bars. A total of seven grounding cables must be installed, with three on the HV side and four on the RO side. Figure 3.11 shows the grounding cables installed in each of the four corners of an sMDT chamber. The grounding cables are the green and yellow cables.

Figure 3.12 shows each of the individual components used to distribute HV to and readout the voltage from each drift tube. HH cards are used in the HV and RO side inside the Faraday cage. The Front-End (FE) electronics on the RO side consist of the Amplifier-Shaper-Discriminator (ASD) and High-Performance Time-to-Digital-Converter (HPTDC). The details of the ASD and HPTDC chips is discussed in Section 3.3.1. Inside the Faraday cage on the HV side, various jumpers are used to connect neighboring HH cards and distribute

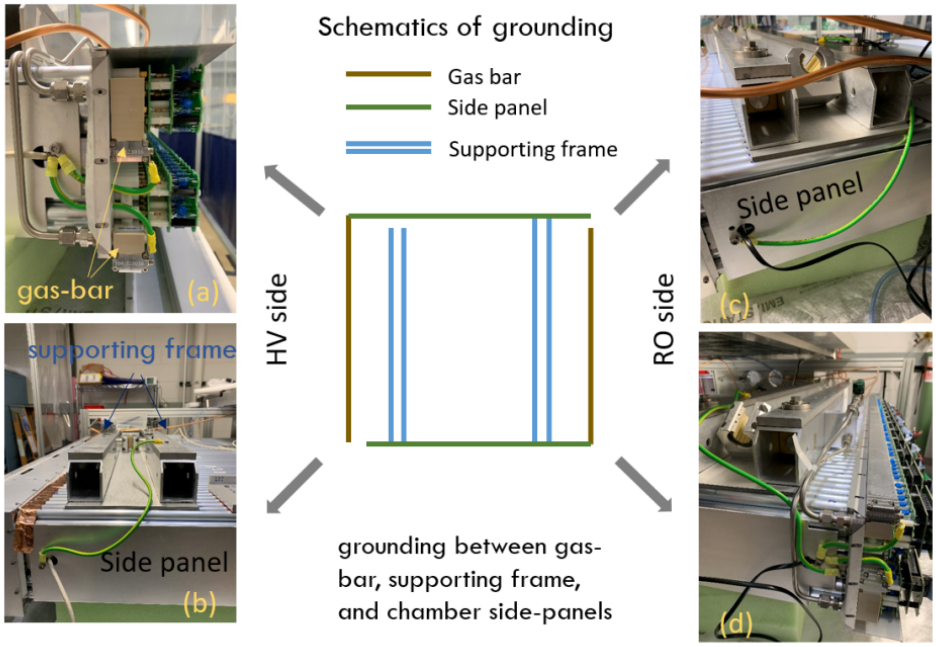
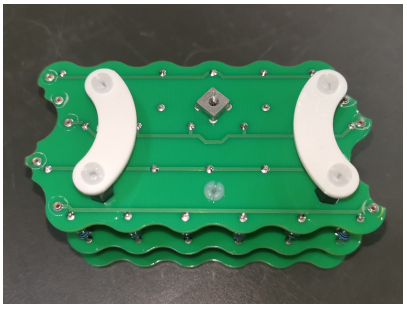


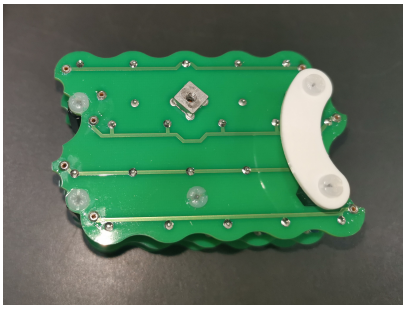
Figure 3.11: Grounding cable connections. (a) Gas bar to the side panel on the HV side. (b) Side panel to the support structure on the HV side. (c) Gas bar and the support structure to the side panel on the RO side. (d) Side panel to the support structure on the RO side.

HV within a layer.

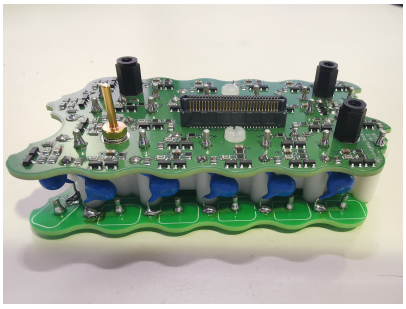
Chambers with fully assembled gas systems, HV distribution, and RO electronics are tested with cosmic-rays and the efficiency, resolution, and noise rate are measured. During testing, defects were discovered with the RO HH cards ("normal" 4x6 variety). Through-hole capacitors are used between the two PCB layers, which can be seen in Figure 3.12, 2(a). For two neighboring channels (numbers 16 and 20 out of 0-23), the capacitor legs can touch. In this failure mode, signals on one channel always show up on its neighbor, and vice versa. These two channels will then have about twice as many hits as expected. The other failure mode is more subtle. Sometimes, the leads of the through-hole capacitors can be swapped entirely, without touching one another. In this mode the channels have the expected number of hits, and will pass efficiency and noise rate tests. However, when reconstructing a track through these tubes, hits will be swapped and it will be almost impossible to construct a precision track through either of these tubes. After its discovery, this problem of swapped



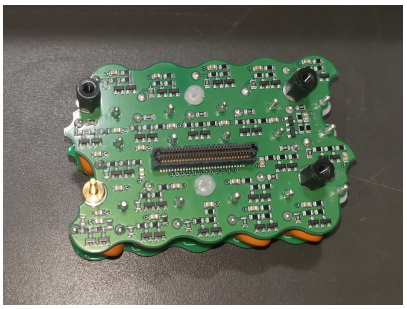
1(a) HV HH normal



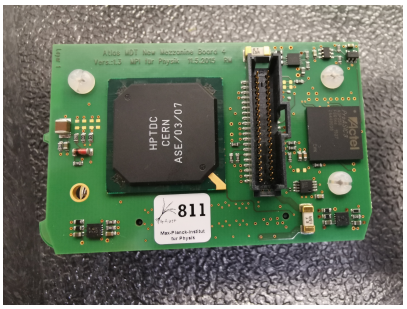
1(b) HV HH corner



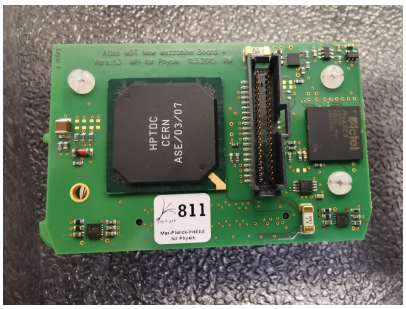
2(a) RO HH normal



2(b) RO HH corner

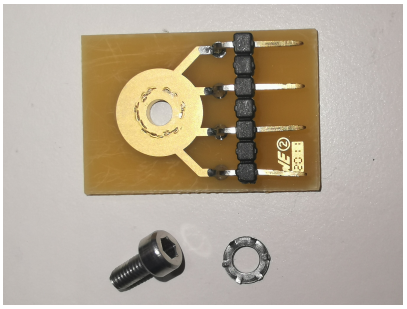


3 ASD mezz

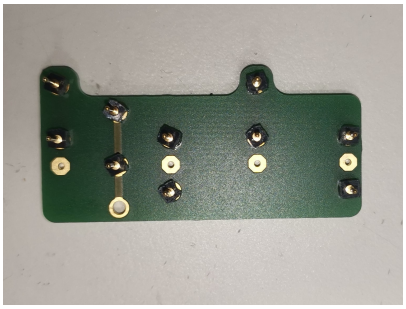


4 HPTDC mezz

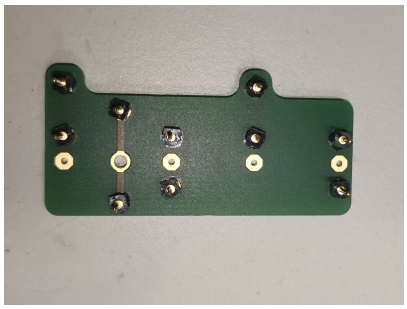
5 HV dis box



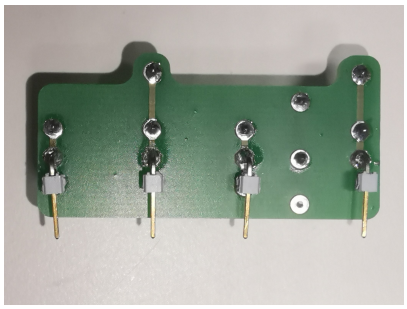
6 HV grounding board



7(a) jumper



7(b) jumper corner



7(c) jumper with connector

Figure 3.12: Front end electronics for read-out and high voltage distribution.

channels was verified in the lab by connecting a function generator to channels one at a time and checking which RO channel had hits on it. This problem can only be found when constructing precision tracks and measuring the resolution, because the tubes operate exactly as expected but the reconstruction code will not know the "right" location of these two swapped tubes. After testing 30 chambers, about 5% of all RO HH cards have a capacitor issue with channels 16 and 20: either touching leads or swapped leads. There is no remedy for this issue but to test chambers and discard HH cards when an issue is found. However, it does demonstrate the efficacy of the testing procedure, which is discussed in the next section.

3.3 Chamber Quality Control Tests

3.3.1 Data Acquisition

A mini-data acquisition (mini-DAQ) system [162] developed at the University of Michigan, as shown in Figure 3.13, is used to collect cosmic ray muon signals from the sMDT chamber. A large (≈ 1.5 m^2) scintillator sits above the test chamber with photo-multiplier tubes (PMTs) mounted on each end and provides the trigger to the mini-DAQ by the coincidence of the two PMTs. Front-end electronics (the ASD and HPTDC) are installed on a three-layer stack of circuit boards mounted on the end of the chamber (see Section 3.2.3). The sMDT cosmic-ray test station is shown in Figure 3.14.

When a muon passes through a tube, the electrons from the primary ionization clusters drift to the central wire which has a potential at 2730 V. The earliest arrival time of the ionized electrons to reach the wire is captured by the ASD and the drift time is obtained by comparing the difference between this arrival time and the scintillator trigger time. In addition, the pulse height of the detector signal for the first $\sim$20 ns is also measured by a Wilkinson ADC on the ASD chip, which allows the pulse-height-dependent slewing corrections to be adapted for the timing measurement. The pulse height of the signal is encoded as the time interval between the leading and trailing edges of the ASD output logic pulse.

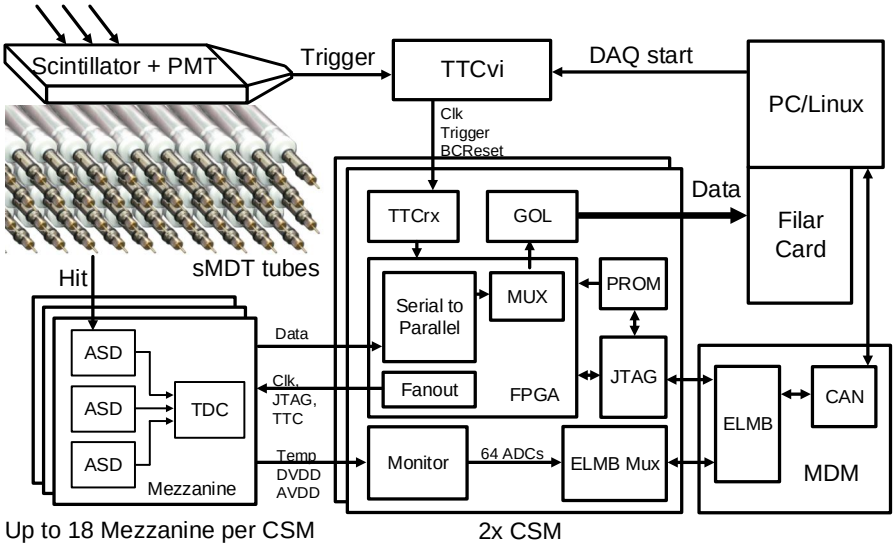


Figure 3.13: Block diagram of the sMDT readout electronics and mini-DAQ system. See text for explanation of components.

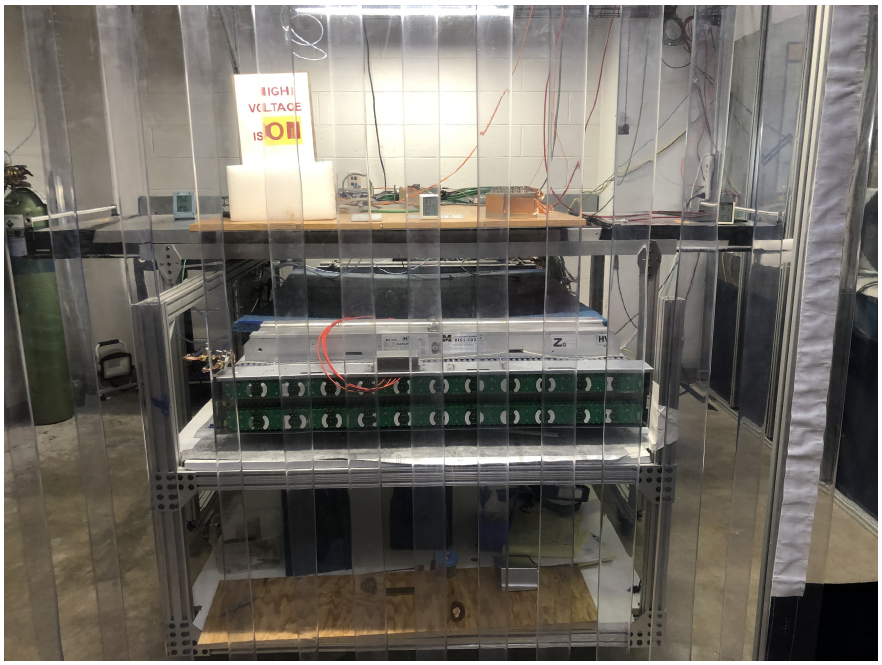


Figure 3.14: sMDT cosmic ray test station in a humidity-controlled room. The sMDT chamber sits on a cart with a 1.5m^2 scintillator mounted above. In this image the Faraday cage is not yet fully assembled and the HV HH cards are seen on the chamber.

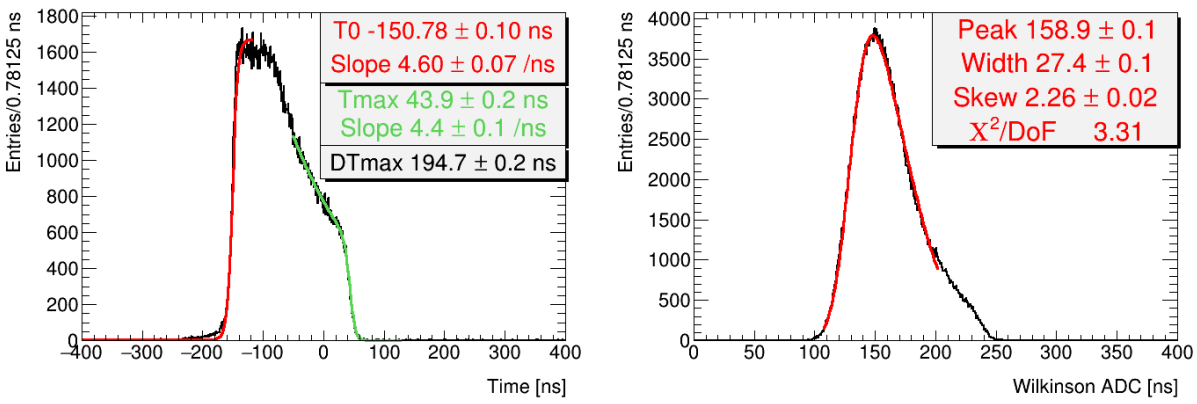
The HPTDC digitizes the both edges of the ASD output, performs the trigger matching, and sends the matched data to the Chamber Service Module (CSM) for readout. One CSM multiplexes data from up to 18 HPTDCs and transmits it to a Linux machine, where the data related to the same trigger is packaged as an event and saved for offline analysis. The scintillator trigger signal timing is also recorded by one channel of one of the HPTDCs.

The TTCvi [163] module is used to provide both the trigger and clock to the front end electronics. It receives the coincident PMT signal and then distributes this signal to the CSMs. An internal 40 MHz clock is also provided by the TTCvi to the CSMs. The CSMs then distribute the clock and trigger signal to all the mezzanine cards.

Triggering on cosmic-rays via scintillator PMTs has a significant trigger time smearing due to the large size of the scintillator, as is addressed in Section 3.3.2. The ASD has a threshold of ≈ 24 primary ionization electrons corresponding to an electronics signal threshold of 39 mV. To filter out noise hits that are not related to a muon track, hits from adjacent tubes on different layers are grouped into clusters. At least one but no more than three clusters of 1-10 hits is required in each multilayer of the chamber. The requirement of at least one hit in each multilayer means some triggered events will not pass the reconstruction cuts. The rejection of events with too many clusters vetos large shower events. The expected geometric acceptance is 73%. The reconstruction cuts are approximately 30% efficient on top of the geometric acceptance, and the majority of the events rejected with hits in both multilayers involve multiple and/or large clusters of hits, in which it appears a cosmic-ray particle has undergone a shower process, or coincident noise hits not adjacent to the primary cluster of hits.

3.3.2 Data Analysis

The data analysis for sMDT chambers encompasses two main areas: (1) calibrating the drift tubes and reconstructing tracks from raw data to determine the tracking resolution; (2) estimating the multiple scattering of cosmic-rays so that this multiple scattering can be



(a) (b)

Figure 3.15: (a) Example drift time spectrum for one channel from a cosmic-ray data run collected using the BMG prototype chamber. The red curve is the Fermi-Dirac fit of the rising edge to determine t_0. The green curve is a fit the falling edge to determine t_{max}. DTmax is $t_{max} - t_0$ (b) Example ADC spectrum for the same channel, where the red curve and text show the results of a skew normal fit.

removed from the resolution measurement to allow comparison with ATLAS data from p-p collisions. The analysis strategy is as follows. First, the drift time is corrected for time slew and the rising and falling edges of the drift time spectrum are fit to derive the minimum and maximum drift time for each tube. Then a calibration of the r(t) function is performed, in which r(t) is parameterized as a 10-degree Chebyshev polynomial. Then straight, 2-D tracks are fit in the plane perpendicular to the wires and distributions of biased and unbiased tracking residuals are created. The residual is the difference between the drift radius and the radius predicted by the straight line fit. Biased residuals are those from fits using all tube hits, whereas unbiased residuals are found by refitting a track multiple times by removing one hit from the from fits and finding the residual of the hit removed from the fit. To account for multiple coulomb scattering, which is much more significant for cosmic-rays than muons in the ATLAS experiment with ≥ 20 GeV of p_T, residual distributions are deconvoluted with a Monte Carlo (MC) truth multiple scattering residual distribution.[1] Finally, the spatial resolution is calculated.

Now the procedure is described in greater detail. A time slew correction is subtracted

[1] $p_T \geq 20$ GeV is the cut used in the ATLAS MDT Run 2 resolution measurement [50].

from the drift time to compensate for the timing jitter caused by variations in the signal amplitude. Large signals have shorter rise time and cross the ASD threshold more quickly than small signals causing a pulse-size dependent time slew. The time slew is a function of the response of the particular Amplifier/Shaper/Discriminator (ASD) chip used in the front-end electronics, as described in the corresponding manual (see figures 34 and 35 in [164]). The time slew correction used in this study is taken from ATLAS MDT calibrations and shown in Equation 3.3:

$$\text{timeslew} = 35.59e^{-ADC/61.33\text{ns}}\ [\text{ns}] \tag{3.3}$$

where ADC is the output pulse width of the ASD chip. Typical values for ADC are on the order of 100 nanoseconds, and typical values for the time correction are $\approx 6 \pm 2.5$ ns. The nonzero mean slew correction is irrelevant and will be accounted for in the t_0 calibration described below.

The signals also contain an timing offset, t_0, due to electronics and other delays, which must also be subtracted from the drift time. t_0 is defined as the half-way point in the leading edge of the drift time spectrum. t_0 is determined by fitting the leading edge of the drift time spectrum with a Fermi-Dirac step function as defined in Equation 3.4 and shown in Figure 3.15a:

$$b + \frac{A}{1 + e^{-(t-t_0)/T}} \tag{3.4}$$

where b is the background noise floor, A is the amplitude, T is the t_0 slope (rise time). In addition, the maximum drift time, $t_{\max}$, is determined via a Fermi-Dirac fit on the falling edge of the drift time spectrum.

The ADC distribution is fit using a skew normal distribution. Note that in order to cut noise, a threshold cut of 40 ns was imposed on ADC values. Example drift time and ADC spectra are shown in Figure 3.15b.

Auto-calibration [165] and track fitting are both performed using a linearized least-squares procedure. Auto-calibration is the process by which the parameterization describing the r(t) function is updated iteratively. The auto-calibration algorithm iterates over two steps: (i) fitting tracks for set of events using a fixed r(t) function; (ii) updating the r(t) function using the least-squares method. The process continues until the parameterization of the r(t) function converges. Auto-calibration requires knowledge of the resolution as a function of drift radius, which is the quantity under study. For the r(t) auto-calibration the resolution function is initialized to that observed using ATLAS MDTs [50].

The r(t) function is parameterized as a Chebyshev polynomial with 10 terms. Via the least-squares method the changes to apply to each Chebyshev coefficient on each iteration of auto-calibration are calculated. Chebyshev functions are defined on the domain [-1,1], so the domain of the Chebyshev functions are scaled to $[t_0, t_{\max}]$ for the individual t_0, $t_{\max}$ for each tube. The r(t) function is also constrained so that the predicted radius $r(t_0)=0$. Without this constraint the auto-calibration will tend to fit an un-physical nonzero radius for the minimum drift time.

A robust means of validating this procedure is to compare the chi-square distribution obtained to the expected one with the same number of degrees of freedom. The chi-square distribution obtained from the track fits using the auto-calibrated r(t) function is shown in Figure 3.16, compared to the predicted chi-square distribution. The non-closure in the tail is likely the result of multiple coulomb scattering, which is not yet accounted for at this stage, but is addressed in Section 3.4.3. The angular distribution for all fitted tracks is shown in Figure 3.17. The r(t) function for the prototype chamber is shown in Figure 3.18. The error band is the resolution as a function of radius, shown in Figure 3.30a.

Track fitting is performed as follows. The first step is to make an initial guess of track parameters via a simple pattern recognition algorithm to seed the chi-square minimization algorithm. The pattern recognition algorithm used is a brute force method to determine if the track passes to the right or left of the wire. $2^{(\text{n hits})}$ different combinations of hit

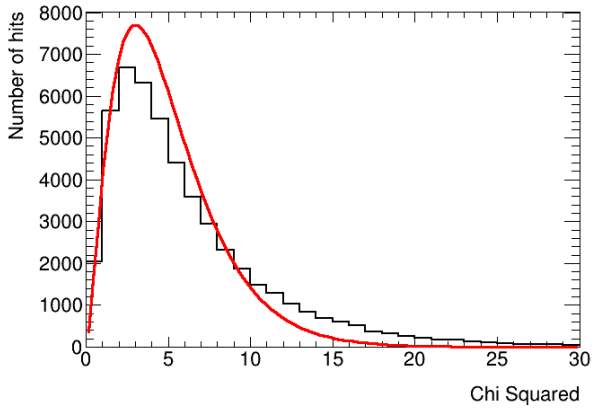


Figure 3.16: Chi-square distribution for fitted straight line tracks. The tracks are fit in a 2-D plane perpendicular to the wires, as shown in Figure 3.20. The red curve shows the expected chi-square distribution, given the number of degrees of freedom, normalized to the same area as the histogram.

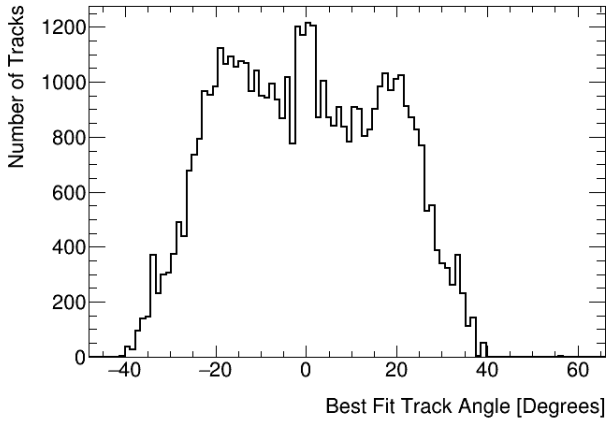


Figure 3.17: The angular distribution for all fitted tracks for 24 hour cosmic-ray run taken with the BMG prototype.

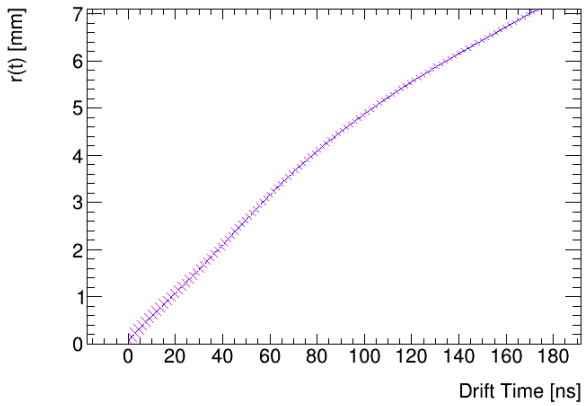


Figure 3.18: The auto-calibrated r(t) function. Note that the Chebyshev polynomials are defined on the range [-1,1] so this r(t) function is linearly scaled to [t_0,$t_{\max}$] for each tube. Therefore, different tubes may have slightly different r(t) functions if they have a differing maximum drift time $t_{\max} - t_0$. The hatched region represents the resolution as a function of radius presented in Figure 3.30a.

location are tried, displaced either to the right or left of the wire by the radial distance. The combination of right/left decisions that has the best chi-square is used to provide initial values of the slope and intercept of the track. For the first pass, the error used is the expected MDT resolution function [50]. After the radial resolution function is measured, the entire resolution procedure is iterated until the coefficients in Table 3.5 converge.

Track fitting again uses the linearized least-squares technique. Tracks are modeled in a 2-dimensional plane and can be parameterized as a straight line with an angle with respect to the vertical and x-intercept (see Figure 3.20 for a definition of the x axis). The choice of angle with respect to the vertical and x-intercept to describe a straight line avoids possible singularities in slope and y intercept for a purely vertical track. Additionally, tracks have a third parameter: a global time shift Δt_0. The Δt_0 is introduced to account for the random variations in trigger time from cosmic rays which are not synchronized with the clock used by the DAQ. The Δt_0 distribution shown in Figure 3.19.

Tracks are fit using the current r(t) function, and residuals from the best fit track are used in the auto-calibration routine to update the Chebyshev coefficients. An example event display with a fitted two-dimensional track is shown in Figures 3.20.

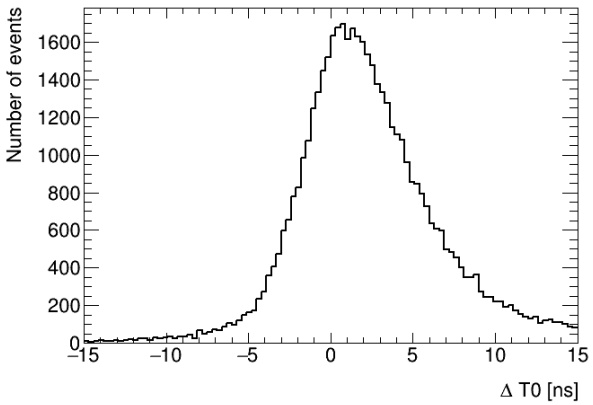


Figure 3.19: Δt_0 distribution from fit residual calculation for cosmic-ray data taken with the BMG prototype.

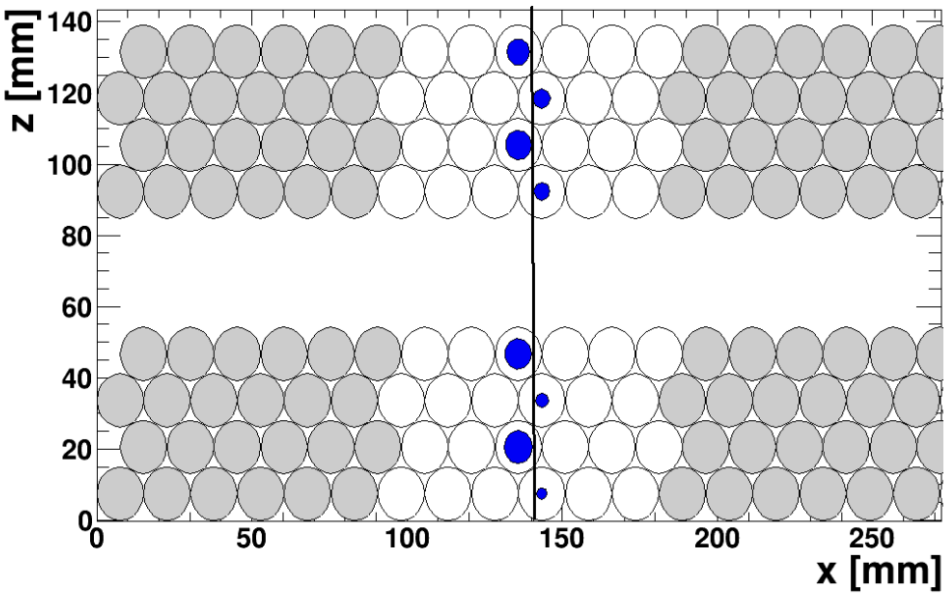


Figure 3.20: A single cosmic ray muon event in a BIS chamber with a straight line track fitted. The blue circles show the drift radius of each hit, which is determined using the auto-calibrated r(t) function. The grey and white regions show the demarcations between different readout electronics cards.

3.3.3 Resolution

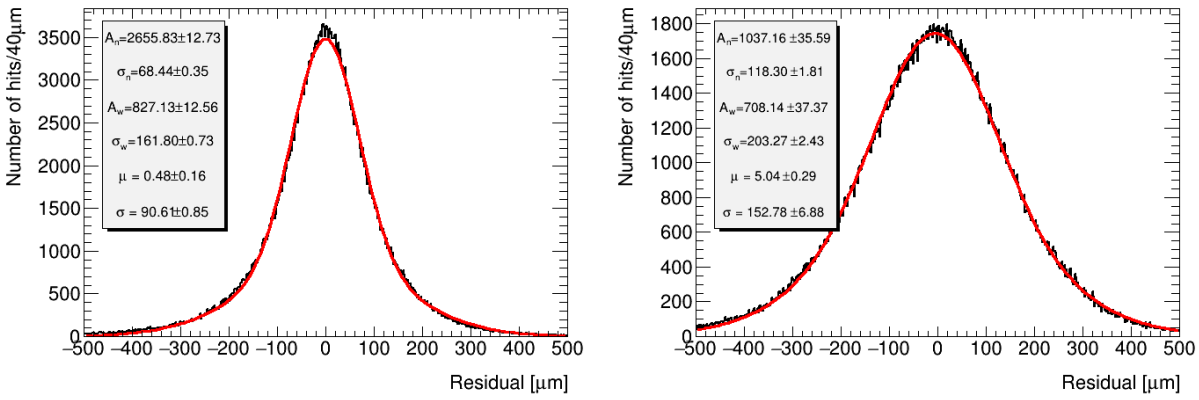
After calibrating the r(t) function a resolution measurement is performed. Using the calibrated r(t) a straight line track is fit in the two dimensional plane perpendicular to the wires, with the additional event-by-event Δt_0 timing shift and the biased and unbiased residual distributions are calculated. The biased residual distribution uses all hits in tracking, while the unbiased residual drops out a tube, fits a track, and then records the residual on the tube which was dropped from the fit. The quantities σ_b (biased width) and σ_u (unbiased width) are defined as the average width of a double Gaussian fit to the corresponding residual distribution, weighted by the amplitude of each Gaussian [50]:

$$\sigma_{[b,u]} = \frac{A_w \times \sigma_w + A_n \times \sigma_n}{A_w + A_n} \tag{3.5}$$

Where A is the amplitude of the Gaussian, σ is its standard deviation, and the subscripts n and w are the narrow and wide Gaussians of the double Gaussian distribution. The double Gaussian fit is constrained to have a common mean for the two Gaussians. Examples of biased and unbiased residual distributions with double Gaussian fits are shown in Figure 3.21a and Figure 3.21b, respectively. The resolution σ is defined to be the geometric mean of the two residual distributions [166]:

$$\sigma = \sqrt{\sigma_b \times \sigma_u} \tag{3.6}$$

Applying Equation (3.6) to the biased and unbiased residuals from the cosmic-ray data obtained using the BMG prototype chamber, a single-hit resolution of 117.7 $\pm$2.1 µm is observed. The quoted error is the statistical error from the fit parameters, propagated through Equation (3.6). The primary reason for the tension with the expected resolution of 106 µm is multiple scattering of low energy cosmic-rays, as is discussed in Section 3.4.



(a) (b)

Figure 3.21: (a) Biased residuals from a cosmic-ray run taken with the BMG prototype. The red curve shows the double Gaussian fit. The fit parameters shown in the box are: A_n, σ_n = Amplitude and width of the narrow Gaussian fit, respectively; A_w, σ_w = Amplitude and width of the wide Gaussian fit, respectively; μ = common mean of wide and narrow fits; σ = the biased residual distribution width from Equation 3.5. (b) Unbiased residuals from cosmic-ray data taken with the BMG chamber. The red curve shows the double Gaussian fit. The fit parameters are the same as shown in (a) except σ = the unbiased residual distribution width using Equation 3.5.

3.3.4 Efficiency

Two types of efficiency measurements are defined for the drift tubes: layer efficiency and tube efficiency. The layer efficiency is defined as the number of times a track passes through a layer with a hit recorded divided by the number of times a track passes through a layer regardless of whether a hit was recorded. The layer efficiency accounts for both the efficiency of the tubes in reconstructing hits and the dead regions due to tube walls and the gaps between tubes.

Tube efficiency is the number of tracks passing through the gas volume of a tube with a hit recorded divided by the number of tracks passing through the gas volume of that tube whether or not a hit was recorded. Tube efficiency is expected to be near 100%. When evaluating whether a track passes through a tube volume that tube is left out of the track fit so as not bias the fit to be closer to that tube, as is done to calculate the unbiased residual distribution. Hits greater than 5σ away from a track are not counted towards the efficiency.

3.4 Multiple Coulomb Scattering

3.4.1 Geant4 Simulation

In order to estimate the impact of multiple scattering, a Geant4 [137] simulation of an sMDT chamber is used. The simulated geometry includes the tubes (walls, gas, and wires) and spacer frame for BMG-type chambers. Figure 3.22 shows a simulated BIS1 chamber in Geant4.

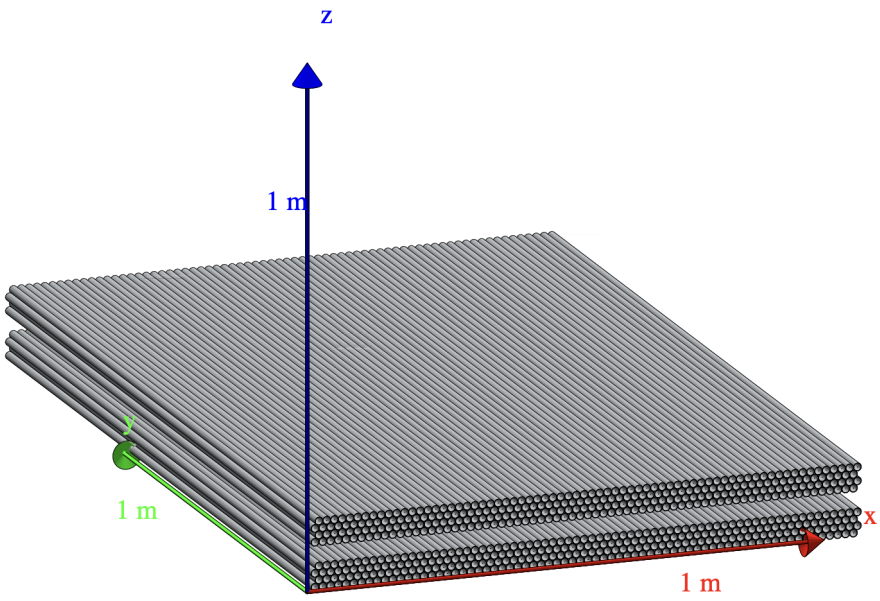


Figure 3.22: 3D event display of a BIS1 sMDT chamber in Geant4.

Using MC truth information for charged particles passing through the simulated gas volume, events are reconstructed. Thus, a data set is created with perfectly reconstructed "MC truth" hit radii. The hit radii are reconstructed as the closest approach of any charged particle within the gas volume, so the effects of delta rays or any other energetic secondaries are included. This data is fed through exactly the same data analysis steps discussed in Section 3.3.2 and Section 3.3.3. The width of the residual distributions are solely the result of multiple scattering, since tube resolution has not been added to the MC. For an example

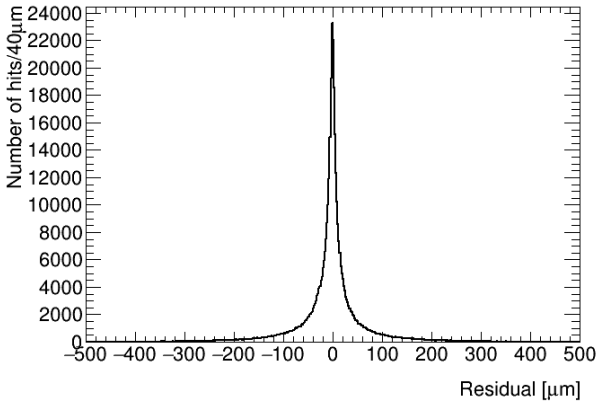


Figure 3.23: Residual distribution obtained from simulating multiple scattering of cosmic-rays passing through the test chamber. There is a sharp central peak and non-Gaussian tails.

residual distribution see Figure 3.23. Unlike the residual distribution observed in data, it is not well described by a double Gaussian fit, but has a sharp central peak and non-Gaussian tails.

The r(t) calibration procedure is validated on a MC truth data set with 20 GeV muons. Hit times are then created with the inverse r(t) function, which is linearly interpolated in 1.9 ns bins. In such a test over 90% of the fit residuals are less than 10 microns, validating the ability of the auto-calibration to capture a realistic r(t) function.

3.4.2 Simulation of the Cosmic-Ray Spectrum

The effect of the multiple scattering is dependent on the assumptions on the particle and energy content of the cosmic-ray spectrum. Many particle species are present in cosmic-rays due to energetic electromagnetic and hadronic showers in the upper atmosphere. Only the muon and soft-electron components of the cosmic-ray spectrum are simulated, which are the most significant components at sea level.

The Particle Data Group (PDG) article on cosmic-rays [18] provides data describing muon flux as a function of momentum at sea level. This data is fit as the sum of two exponentials as seen in Figure 3.24. The functional form of the probability distribution is

Parameter	Value
A_1	0.3924
λ_1	5.5264
λ_2	1.4180

Table 3.2: Fitted parameters to Equation (3.7) to describe the muon component of the cosmic-ray spectrum

described in Equation (3.7), and the coefficients are listed in Table 3.2. The PDG article also notes that the muon energy spectrum is roughly flat below 1 GeV, so the fit terminates at 0.6 GeV, where the data runs out, and the energy spectrum is simulated as flat below this cutoff.

$$\frac{A_1}{\lambda_1} e^{p/\lambda_1} + \frac{(1-A_1)}{\lambda_2} e^{p/\lambda_2} \tag{3.7}$$

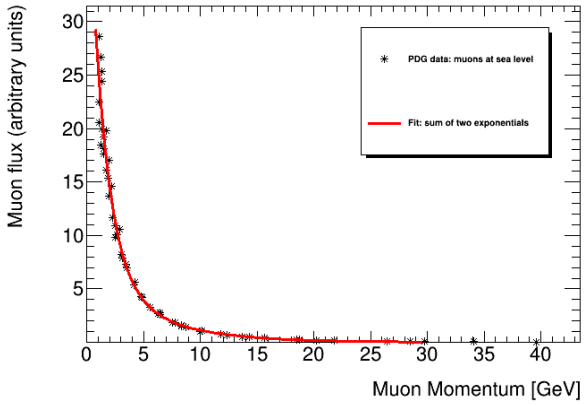


Figure 3.24: Probability distribution for cosmic-ray muon flux at sea level. The sum of two exponential curves are fit to cosmic-ray data. For functional form, see Equation (3.7). For fit parameters, see Table 3.2.

The muon component described thus far is estimated to be 70% of the particle content. The remaining 30% of cosmic-rays are simulated as soft-electrons. This ratio is taken from the PDG article on cosmic-rays [18], which lists the intensity of cosmic-ray muons to be $\approx 166\text{m}^{-2}\text{s}^{-1}$ and $\approx 55\text{m}^{-2}\text{s}^{-1}$ for cosmic-ray electrons above 10 MeV, for a ≈25% electron component (integrated within our geometric acceptance). Therefore 30% is a conservative estimate of the electron component of the cosmic-ray spectrum. Ultimately, Section 3.5.3

shows that the resolution is rather insensitive to the soft-electron fraction. The electron component is modeled as having a minimum energy of 10 MeV and as an exponential with a length scale of 50 MeV.

3.4.3 Deconvolution of Multiple Scattering Effects

The observed residuals presented in Section 3.3.3 are a convolution of the true detector resolution and multiple scattering. Using the MC truth multiple scattering residual distribution, the effect of multiple scattering is removed via a deconvolution and the resolution of this detector for high p_T muons in the ATLAS experiment is estimated.

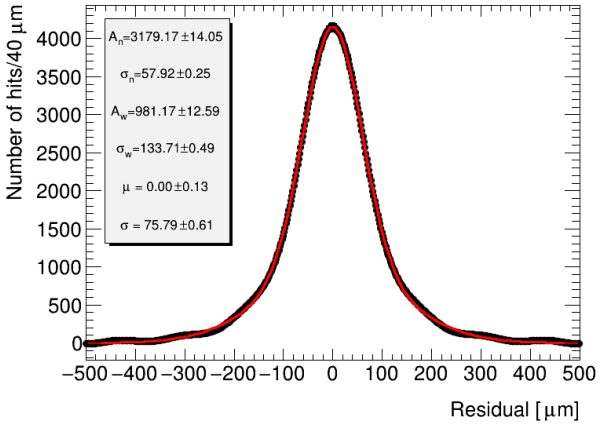


Figure 3.25: Biased residual distribution of the BMG chamber from Figure 3.21a deconvoluted with the simulated multiple scatter residual distribution in Figure 3.23. The red curve is a double Gaussian fit with fit parameters as described in Figure 3.21a shown in box.

The deconvolution is performed using Fourier transforms of the two input distributions. In frequency space, a deconvolution is simply the division of the two histograms. The observed residuals are Fourier transformed and divided by the Fourier transform of the MC truth multiple scattering distribution, and then the inverse Fourier transform of the result is computed. The result of the deconvolution procedure is shown in Figure 3.25.

Then a similar procedure to the deconvolution is performed using a convolution to smear the residuals with the multiple scattering distribution of a sample of pure 20 GeV muons from the MC simulation. 20 GeV muons are chosen to be in accordance with the Run 2 MDT

resolution result [50], which placed a 20 GeV p_T cut on muons. Muons with 20 GeV of p_T have at least 20 GeV of energy, so the 20 GeV monochromatic sample used was conservative compared to the actual spectrum used in the Run 2 MDT resolution measurement. The multiple scattering is much less for this sample, but still adds 3-4 microns to the overall resolution.

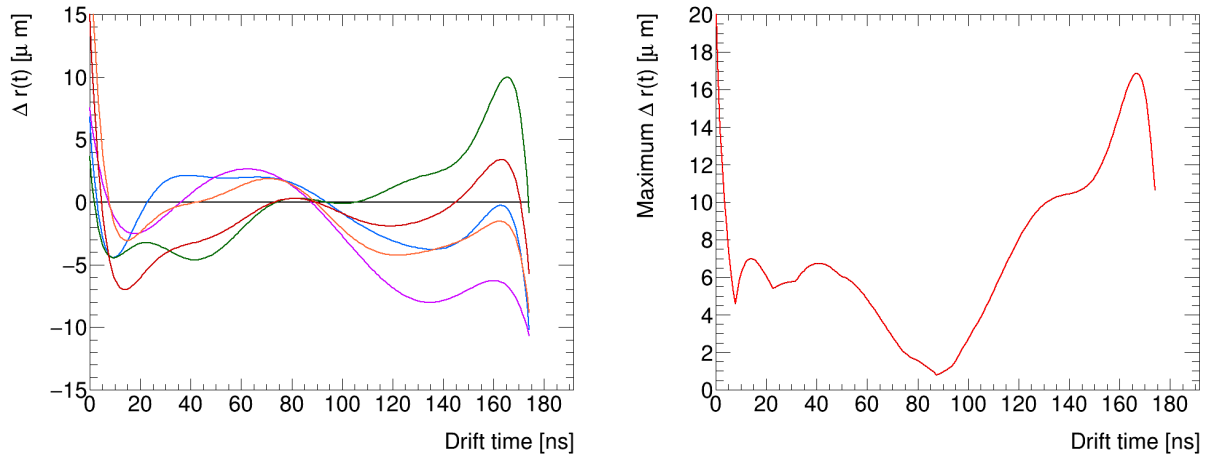
3.5 Systematic Uncertainty Estimation

3.5.1 Auto-calibration and r(t) parameterization

The choice of Chebyshev polynomials, and the exact value of the coefficients, is a potential source of systematic uncertainty. In order to assess the reproducibility of the auto-calibration algorithm, independent data sets are auto-calibrated and the resulting r(t) functions are compared. The results are shown in Figure 3.26. Differences in r(t) functions are on the order of 5-10 microns for the majority of the drift time phase space. In order to assess the impact on the final resolution result the residual distributions (Figures 3.21a and 3.21b) are calculated using an r(t) function from a different data partition. Then the usual deconvolution/convolution methodology is applied and the resolution is calculated. The maximum change in resolution is 0.3 µm. This value is the uncertainty estimate for the Chebyshev parameterization and auto-calibration routine.

3.5.2 Track Fitting

In order to calculate the uncertainty associated with track fitting the likelihood ratio technique is used. The track parameters (Δt_0, angle, and impact parameter) are varied around their central values and the values which cause the likelihood ratio to change by one standard deviation are computed. The size of the perturbation in each parameter required to shift the likelihood by one standard deviation is shown in Table 3.3. After perturbing the track parameters, the biased and unbiased residual distributions are recomputed and



(a) (b)

Figure 3.26: (a) Difference in r(t) function calculated on independent data sets. Differences are generally on the order of 5-10 microns. (b) Maximum difference between any two data partitions. Differences are greatest near the wire and tube wall, but are again on the order of 5-10 microns.

.

the perturbed residual distributions are propagated through the deconvolution procedure to measure the resolution. From this method a maximum systematic error of 5.5 µm is observed due to the track fit. See Table 3.3 for a complete summary of how the individual parameters impact the resolution.

Track Parameter	Mean 1 sigma deviation	Change in resolution
Δt_0	0.81 ns	0.4 µm
b	28 µm	5.2 µm
θ	140 µrad	5.5 µm

Table 3.3: Mean size of systematic shift for track parameters and impact on resolution.

3.5.3 Soft-Electron Component of Cosmic-Ray Spectrum

In this section the systematic effects of the MC procedure are estimated. The primary nuisance parameter is the soft-electron component of the cosmic-ray spectrum, which is nominally chosen to be 30%. The deconvolution procedure also introduces some systematic uncertainties in the Fourier transform, because the result needs to be passed through a low-

pass filter to remove noise. The noise is the result of dividing two doubles near the limit of precision in some bins. The uncertainty associated with the low-pass cutoff and the cosmic-ray electron component are assessed simultaneously by varying the soft-electron fraction to 25 and 35% of the total spectrum and re-computing the deconvolution and final resolution. The maximum variation in resolution observed was 1.7 µm. While this may seem small given that electrons are low energy and larger scattering, a significant fraction of electrons are so low energy that they do not penetrate the entire chamber and pass reconstruction cuts.

3.5.4 Time Slew Correction

The adopted values for the time slew correction function shown in Equation 3.3 have uncertainty. In order to estimate the size of this uncertainty the inverse ADC dependence of the residuals before and after applying the time slew correction is plotted in Figures 3.27a and 3.27b, respectively. The ideal time slew correction would have a slope of zero in the straight line fit in Figure 3.27b. However, a slight correlation remains after applying the time slew correction. Figures 3.27c and 3.27d show the relationship between residuals and ADC with the amplitude of the time slew correction varied down or up by 25%, respectively. The 25% variation is large enough that the straight line fits plotted in Figures 3.27c and 3.27d have slopes of opposite sign, showing that the chosen systematic variation is large enough to account for the degree to which the choice of time slew correction amplitude is non-optimal. Therefore an upper limit of 25% is adopted as the uncertainty on the amplitude of the time slew correction.

The uncertainty in the amplitude of the time slew correction is propagated to the resolution by taking the maximum difference between the resolution measured using the nominal time slew correction and the two systematic variations, with amplitude varied up and down by 25%. The impact on the resolution is 3.7 µm.

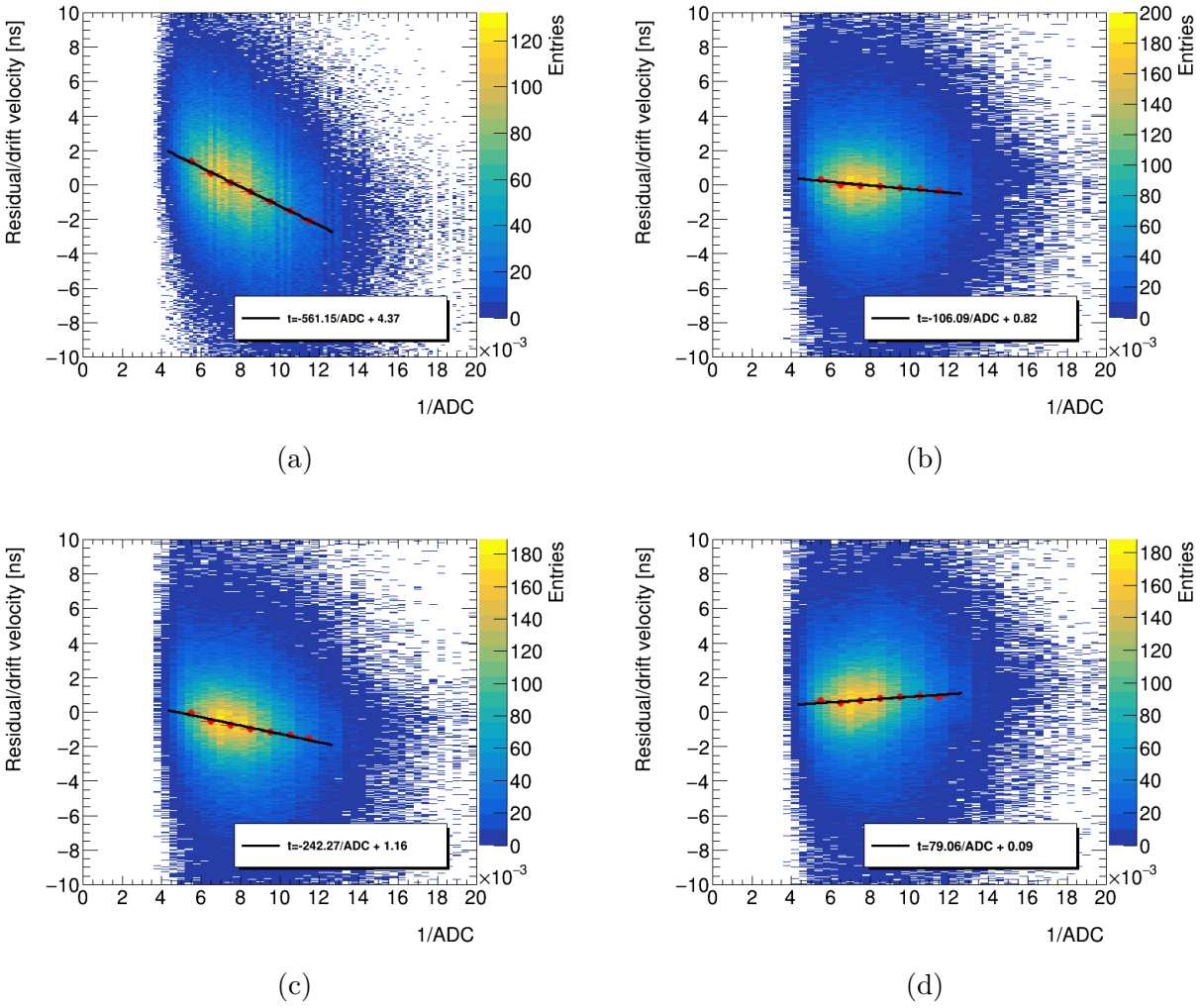


Figure 3.27: ADC dependence of residuals (a) before and (b) after applying time slew correction, as well as with the amplitude of the time slew correction varied 25% (c) down and (d) up. The residuals are divided by the average drift velocity to put the y axis in units of nanoseconds. The red dots show the mean residual/drift velocity in coarse bins and the black line is a straight line fit to this coarsely binned data. The text boxes show the equations found for the straight line fits.

3.5.5 Signal Propagation Time

In the experimental setup the position of the hit along the tube length is not measured, and thus an explicit correction for the propagation time of the signal from hit location to readout electronics cannot be made. The mean signal propagation time is included in the Δt_0 parameter, but inclined tracks have an additional variation of propagation time.

Propagation time is determined by signal propagation speed and distance. First, signal propagation speed is addressed. A drift tube is a coaxial transmission line whose signal propagation speed affected by two factors: (i) the dielectric medium and (ii) the skin depth correction. The dielectric medium used is a 93:7 % admixture of argon and carbon dioxide at 3 bars absolute pressure. The permittivity of the gas mixture is similar to the vacuum and so only the skin depth correction is considered. Using the inductance per unit length of a coaxial cable [167] and assuming a permeability of 1 and a skin depth of 84 microns for 1 MHz signals in aluminum a signal propagation speed of 77% the speed of light is estimated.

The following procedure is used to estimate the systematic error due variations in signal propagation distance. Tracks are generated with random angle drawn from a from a $\cos^2\theta$ distribution [18], up to the maximum geometric angular acceptance (defined by total chamber height and length of wires). The signal propagation distance is calculated for each hit as a function of θ and the mean propagation distance is subtracted (as the mean propagation time is already accounted for by the Δt_0 parameter). Using the signal propagation speed, the distance is converted to a signal propagation time. The standard deviation of the signal propagation time is $\approx$0.4 nanoseconds. Then, assuming hits are uniformly distributed in radius, a random radius is chosen between 0 and the inner tube radius, and the r(t) function is used to calculate the difference in measured radius if the timing was displaced by the signal propagation time. Hits at small radii are more sensitive to changes in timing because the drift velocity is larger near the wire. The observed change in resolution of 2.2 µm. Thus, 2.2 µm is the systematic uncertainty envelope associated with the signal propagation timing.

To check the estimate of the signal propagation uncertainty the original 1.5 m^2 scintillator

is replaced with a smaller one of 0.2 m^2. The difference in residual width between the top layer, closest to the scintillator, and the bottom layer, farthest from the scintillator is observed. The signal propagation time is much more constrained in the top layer than the bottom layer, leading to a wider residual distribution in the bottom layer. Since the scintillator is 20 cm long, there is still a small amount of signal propagation smearing in the top layer. The resolution in the top layer is 1.9 µm narrower than the resolution measured in the bottom layer. This is consistent with the 2.2 µm result obtained via theoretical calculations. Therefore, with empirical evidence and supporting calculations, we adopt a 2.2 µm signal propagation uncertainty.

3.5.6 Systematic uncertainties on efficiency measurement

A possible source of systematic bias in the efficiency measurement is the fact that the efficiency is averaged over all 8 layers. Cosmic-rays are incident from above. If there are soft particles (i.e. electrons) that can penetrate some but not all of the layers, it is possible that the efficiency is a function of the number of layers penetrated. The tube efficiency as a function of the number of layers penetrated is plotted in Figure 3.28 (geometric inefficiencies are not a function of layers penetrated). The constant fit with error bars is also plotted. The constant fit has a chi-square per degree of freedom of 0.45, and is not significantly different than the observed data. Therefore no systematic uncertainty is adopted regarding the efficiency as function of the number of layers penetrated.

3.6 Results

3.6.1 Noise Rate

Noise rate is measured independently for each tube using a 10 kHz random software trigger. The noise rate is measured in both the HV-off and HV-on (2730 V) configurations.

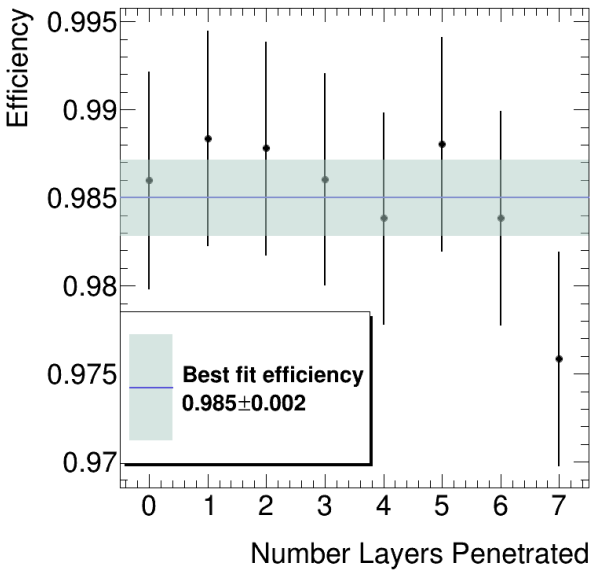


Figure 3.28: Efficiency within the gas volume measured as a function of the number of layers penetrated. 0 layers penetrated corresponds to the top layer, 1 to the second from the top layer, and so on. The constant fit has a chi-square per degree of freedom 0.45, and is not in tension with the data. No systematic uncertainty is adopted.

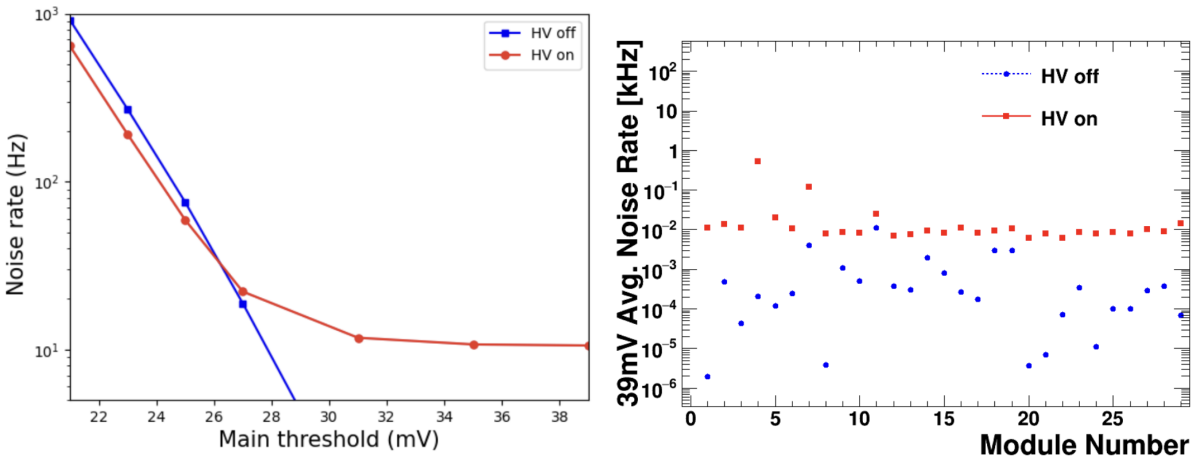
The noise rate is calculated as:

$$\text{Noise rate} = \frac{N_{\text{hit}}}{N_{\text{trigger}} \times \Delta t} \tag{3.8}$$

where $\Delta t = 1.55\mu$s is the HPTDC readout time window, N_{hit} is the number of hits on the channel, and N_{trigger} is the number of total triggers: the trigger rate times the total run-time. The specification is for each tube to have a maximum of 500 Hz noise rate at the 39 mV threshold. Figure 3.29 shows the results for noise rate. Usually, the chambers are well within the limit. One chamber, module 4, has a higher average noise rate. This was due to a one-time error, and module 4 will be re-tested at CERN before installation.

3.6.2 Single-Hit Resolution

As discussed in Section 3.3.3, a resolution 117.7±2.1 µm was observed for the BMG prototype chamber (statistical uncertainty only) without accounting for multiple scattering.



(a) (b)

Figure 3.29: (a) Noise rate as a function of threshold for a single chamber with the HV on and off. When the HV is on, cosmic-ray hits are seen at the expected rate, about 10 Hz. (b) Chamber noise for 30 chambers at 39 mV threshold, HV on and off.

After performing the deconvolution procedure outlined in Section 3.4.3 the resolution improved to 103.7±3.5 µm. Accounting for the systematic uncertainty described in Section 3.5 the observed resolution is 103.7±8.1 µm. For the BIS prototype chamber the resolution with is computed both ASD-1 and ASD-2 chips. Notably, the ASD-2 chip achieves finer resolution due to its higher gain. Using ASD-1 chips on the BIS chamber yields a resolution of 101.8±7.8 µm whereas using ASD-2 chips improves the resolution to 83.4±7.8 µm. The ASD-2 chips will be used on sMDT chambers installed in the ATLAS detector, and will be installed on all existing ATLAS MDT chambers in the ATLAS Phase-II upgrade in 2024. See Figure 3.30a for a comparison of the resolution of the BMG and BIS prototype chambers with the Run 2 MDT result [50].

The systematic uncertainties for both chambers are summarized in Table 3.4. The systematic uncertainties are different due to the differing geometry of the two chambers and the different readout electronics (ASD-1 vs 2). While the two chambers are very similar (same operating voltage, tube radius) there are geometric differences that can subtly impact the performance. Most notably that the two multilayers are approximately 15 cm closer in the BIS prototype as compared to the BMG chamber. Chambers are designed with different

Uncertainty	BMG prototype	BIS prototype
Statistical	3.5 µm	2.4 µm
Track Fit Parameters Error Propagation	5.5 µm	5.2 µm
Soft-Electron Component in MC simulation	1.7 µm	3.0 µm
Signal Propagation Time	2.2 µm	3.3 µm
Time Slew Correction Normalization	3.7 µm	2.8 µm
r(t) parameterization	0.3 µm	0.3 µm
Total	8.1 µm	7.8 µm

Table 3.4: Summary of error bars for single-hit resolution measurement. Total error is obtained by adding the individual errors in quadrature.

geometries in order to fit in various locations within the ATLAS muon spectrometer.

The resolution as a function of radius is also calculated. The biased and unbiased residuals for both the MC truth and observed data are separated into 1 mm radial bins and the deconvolution procedure is performed in each bin. The resulting resolution vs radius curve is shown in Figure 3.30a. The resolution vs radius plots have been fit with a second order polynomial whose coefficients are summarized in Table 3.5. As discussed in Section 3.3, the resolution procedure is performed iteratively until the coefficients in Table 3.5 converge. The coefficients define the error on a hit, depending on its radius.

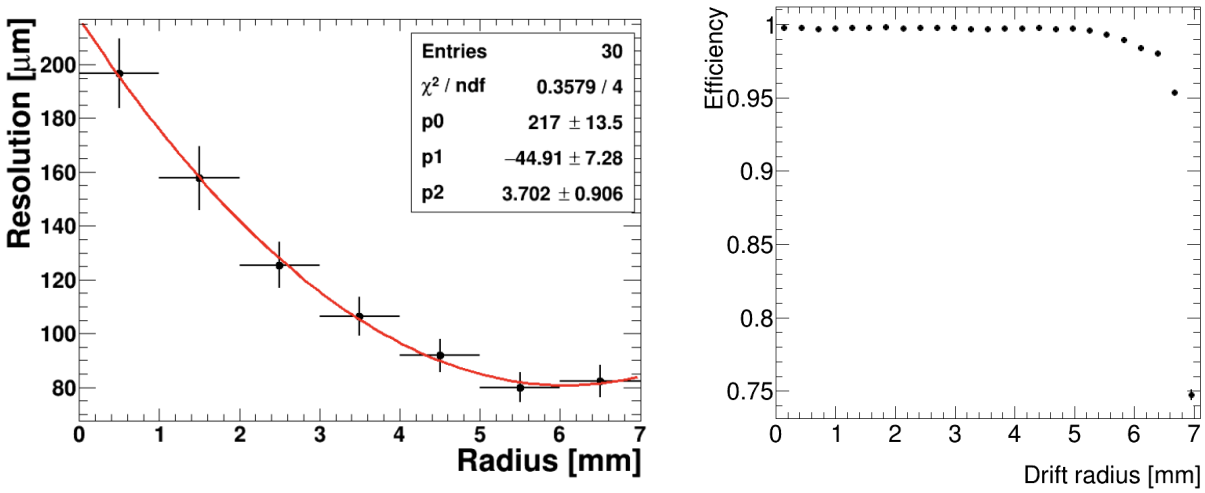
Parameter	BMG Prototype Value [µm]	BIS Prototype Value [µm]
c_0	243.2±10.9	213.4±10.6
c_1	-56.2±6.5	-56.6±6.4
c_2	4.8±0.9	5.3±0.8

Table 3.5: Fitted polynomial coefficients for the residual vs. radius curve, a function of drift distance.

3.6.3 Efficiency

Using data taken on the BMG prototype sMDT chamber the measured (expected) layer efficiency is 0.942±0.002 (0.94) [160]. The measured tube efficiency (when only considering the active gas volume of the tubes) is 0.985±0.002. Figure 3.30b shows the efficiency as a function of drift radius, which demonstrates that the region near the tube wall is responsible

for most of the inefficiency of the chamber, other than the intrinsic geometric inefficiency due to spacing between the gas volumes. Similar results are achieved with the BIS chamber.



(a) (b)

Figure 3.30: (a) Observed resolution as a function of radius, after performing the deconvolution technique to remove the effects of multiple scattering. (b) Efficiency as a function of drift radius shown with statistical error bars. The region near the tube wall is responsible for most of the inefficiency of the chamber within the gas volume. The 0.9 mm gap between gas volumes, consisting of the two 0.4 mm tube walls and the 0.1 mm space between tubes, is the reason for the rest of the inefficiency.

No relationship between efficiency and the number of layers penetrated is observed, so the quoted efficiency is an average of all 8 layers. See Figure 3.32 for a comparison of the efficiencies of the first 30 chambers.

3.7 Conclusions

A resolution of 103.7±8.1 µm and 101.8±7.8 µm was observed for the BMG and BIS prototype sMDT chambers, respectively. These measurements are consistent with the expected resolution of 106 µm which comes from previous studies of sMDT performance [168, 169]. The BIS chamber, when instrumented with the higher gain ASD-2 chips, improved to a resolution of 83.4 ± 7.8 µm. The three crucial components of the analysis strategy were the in-situ r(t) calibration, the addition of a Δt_0 parameter, and measurement of multiple scat-

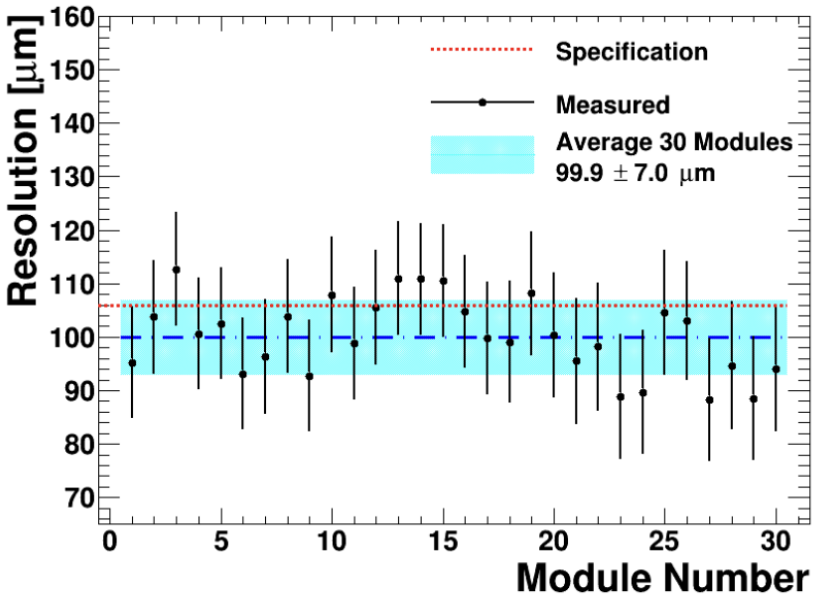
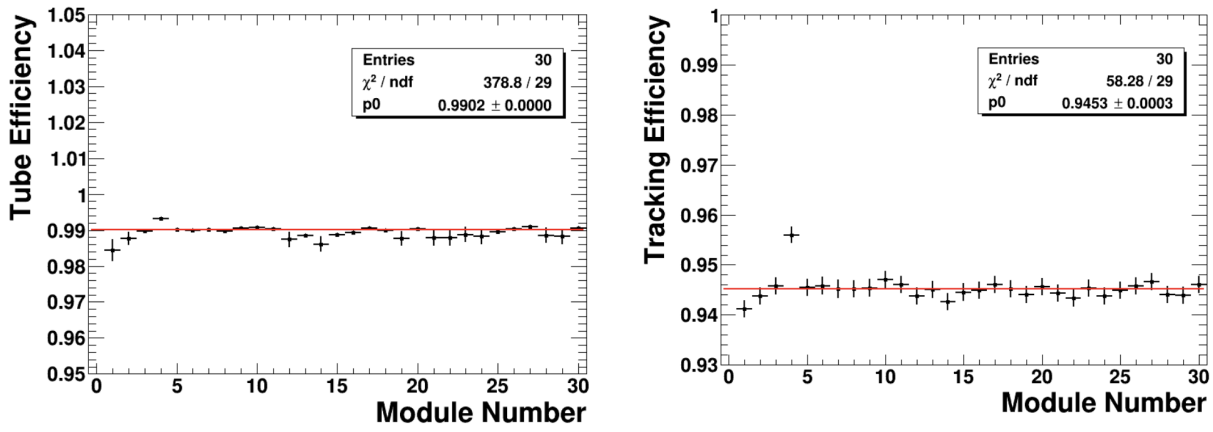


Figure 3.31: Measured resolution for the first 30 sMDT chambers constructed at UM.



(a) (b)

Figure 3.32: Measured (a) tube and (b) tracking efficiency for the first 30 sMDT chambers constructed at Michigan. The tube efficiency is the efficiency when the particle is inside the instrumented volume, and the tracking efficiency is the efficiency of a layer of tubes, with the geometric acceptance included. The expected tracking efficiency is 94.2%.

tering via Geant simulation. The r(t) is determined by an auto-calibration using a linearized least squares procedure.

The addition of the Δt_0 parameter is necessary to account for trigger timing smeared by a large trigger scintillator. The convolution procedure is motivated by the Monte Carlo simulation, which showed a significant difference in multiple scattering for cosmic-rays and 20 GeV muons. By accounting for multiple scattering the prototype chamber resolution as it would perform on muons with 20 GeV of p_T is estimated. The measured efficiency was 0.942±0.002 (0.944±0.002) for the BMG (BIS) chamber, consistent with the expected efficiency 0.94. This efficiency is mainly due to the 0.9 mm gap between neighboring active gas volumes (two 0.4mm tube walls and 0.1mm spacing between drift tubes). The first 30 sMDT chambers assembled at the University of Michigan are consistent with the design specifications for noise rate, resolution, efficiency.

CHAPTER IV

Physics Object Reconstruction

4.1 Overview

Chapter II described the detector and its requirements; namely, that particles of different species are identified and their position and momentum are reconstructed. This chapter describes the algorithms used to translate the raw data collected by the detector into high-level objects relevant for the $h \to bb\tau\tau$ search. All detector subsystems are used in conjunction to reconstruct and identify high level particles; for example, the calorimeter systems add to the spectrometer information for muon identification, and the hadronic calorimeter adds information in electron energy measurement and identification. This chapter is the bridge between discussing "truth-level" particles, be they electrons, muons, photons, etc.; and their reconstructed counterparts. When we reconstruct an electron, for example, we don't *know* that the observed tracks and energy deposits were initiated by an electron. The reconstruction algorithms and detector design are physically motivated, so more often than not that the particle species is correct; but nonetheless the distinction is important. Moving forward, the reconstructed object is almost always simply referred to as an "electron," despite the impossibility of knowing the true particle species in data.

The algorithms discussed both identify physics objects and, in some cases, apply an isolation cut. Isolation is the process of adding up the energy or momentum within a cone surrounding an object (typically a lepton). Leptons from top and weak boson decays tend

to be isolated whereas the leptons from heavy flavor decays tend to have other tracks or calorimeter clusters within their isolation cones from the nearby hadron shower. The lepton from a heavy flavor decay in some sense *should* pass the identification requirement, as it truly is a lepton. However, it is a non-prompt lepton from a decay not involved in the primary proton-proton interaction. Isolation is a tool to discriminate between leptons from heavy flavor decays and prompt leptons from a Higgs boson, for example.

The algorithms discussed have several WPs, and the choice of WP depends on the needs of the particular measurement. The WPs selected are summarized in Table 4.1. Not only is identification important, but so is mis-identification, including which types of particles are mis-identified and at what rates, as "fakes," particles which do not originate from the primary hard-scatter vertex or are identified as the wrong species, will make up a significant component of the SM background. Some energy deposits or tracks may be eligible to be reconstructed in multiple candidate objects. An overlap removal (OLR) procedure is described to avoid double-counting.

Table 4.1: Object reconstruction WPs for all channels ($e\mu$, $\mu\tau_{\text{had}}$, $e\tau_{\text{had}}$) in the analysis. For more information, including object removal procedure, see text.

Object	p_{T} min.	η Range(s)	Identification	Isolation
electron	7 GeV	$0 < \|\eta\| < 1.37$ $1.52 < \|\eta\| < 2.47$	MediumLLH[13]	Baseline: — Signal: Tight_VarRad[13]
muon	7 GeV	$\|\eta\| < 2.7$	medium[54]	Baseline: — Signal: PflowLoose _FixedRad[55]
jet	15 GeV	$\|\eta\| < 2.5$	PFlow jets [56] anti-k_t [57] $R = 0.4$	—
b-jet	15 GeV	$\|\eta\| < 2.5$	85% efficient DL1r [58] WP	—
τ_{had}	20 GeV	$0 < \|\eta\| < 1.37$ $1.52 < \|\eta\| < 2.5$	Baseline: RNN >0.01 Signal: medium WP[59]	—
DeXTer B-jet	20-200 GeV	$\|\eta\| < 2.0$	70% DeXTer WP[60]	—

4.2 Inner Detector Tracking

Separate hits in the ID are associated into tracks using a Kalman filter [61]. A penalty system is used to reward longer tracks with more measurements, and hits from different detector subsystems of the ID have different weights. Additionally, within the penalty system tracks are rewarded by the $\ln(p_{\mathrm{T}})$ to promote energetic tracks [62]. Tracking is initially performed in the silicon detectors and then extrapolated out to the TRT. Candidate tracks are required to have $p_{\mathrm{T}} > 400$ MeV.

4.3 Primary Vertex

Vertices are reconstructed by selecting a set of candidate tracks and iterating a best fit vertex using a so-called "annealing" procedure [62]. The annealing procedure initially suppresses the weight disparity between tracks, but relaxes this as the fit iterates. After the fit, any track more than 7 standard deviations from the fitted vertex is pruned from this vertex and returned to the pool of candidate tracks. After all the possible vertices are fit, the primary vertex is defined as the vertex with the highest sum of the squared transverse momentum of the associated tracks, Σp_{T}^2.

4.4 Electrons

An electron is reconstructed as an ID track matched to a cluster of energy in the calorimeter systems. More specifically a topo-cluster, or a topologically connected set of cells in the EM and HCAL systems, is used. The associated track is re-fit to account for Bremsstrahlung radiation [63]. Topo-clusters are merged into superclusters by selecting the highest energy clusters as supercluster seed candidates and then proceeding to add smaller satellite clusters if they are nearby. For electrons and not photons, the cluster distance is relaxed but the satellite cluster must be matched to the same track.

Pileup clusters and photons are two sources of background to reconstructing real elec-

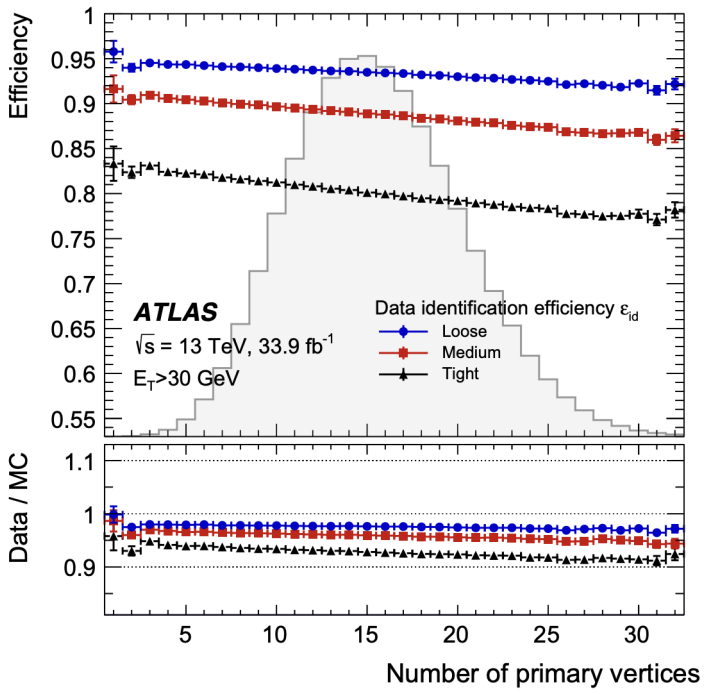


Figure 4.1: Electron efficiency as a function of the number of primary vertices [13]. Three identification WPs are shown, and the ratio plot shows that the scale factor will be different for the choice of WP.

trons. Topo-clusters are required to have an EM fraction (fraction of energy in the EM cal vs total energy in both calorimeters) above 50%, which cuts virtually zero prompt electrons and reduces the pileup background by 60%. Photons will tend not to be matched to an ID track, but rather to a displaced photon conversion vertex, where the photon interacts with the ID and creates an electron-positron pair. The photon conversion rate ranges with $|\eta|$ and is about 20% at low $|\eta|$ and up to 65% at high $|\eta|$.

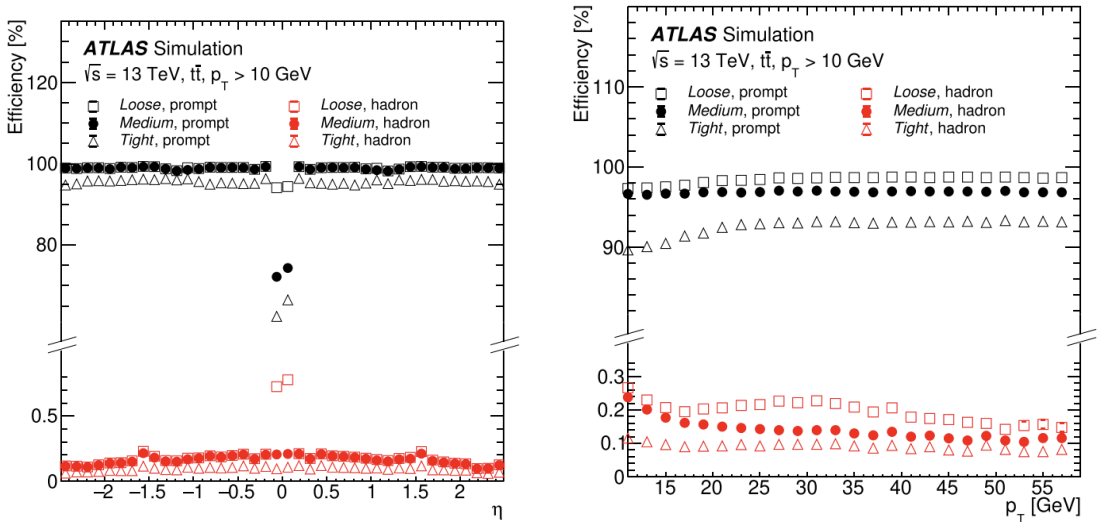
The electron efficiency is shown in Figure 4.1 and is sensitive to the number of pileup interactions (there is nearly a 1-1 relation between pileup and the number of primary vertices). The different WPs "Loose," "Medium," and "Tight" are shown, which each sacrifice progressively more signal acceptance for higher background rejection. In this thesis the MEDIUM WP is used, which has a slight dependence on pileup and is above 85% efficient for real electrons across the pileup spectrum. The ratio plot in Figure 4.1 also demonstrates the need for a scale factor for any Monte Carlo (MC) simulations to correctly weight the

simulation to the data. Calibrations are applied to all reconstructed objects to correctly weight MC and correctly model kinematics. A detailed discussion of calibration is outside the scope of this thesis, but can be found in [64, 65, 66, 67, 68, 69].

Next, the selection cuts applied to candidate electrons for the $bb\tau\tau$ search are described. Electrons are required to have $p_{\mathrm{T}} > 7$ GeV and $|\eta| < 2.47$. A veto is applied in the crack region $1.37 < |\eta| < 1.52$. No cut is applied to the d_0 of the electron candidate track (transverse proximity to the primary vertex) because electrons originating in τ decays are expected to have a displaced vertex. The cut $|z_0 \times \sin(\theta)| < 0.5$ mm (longitudinal proximity to the primary vertex) is still applied. Electrons are required to pass the LooseAndBLayerLLH identification WP. An OLR procedure is applied as described in Section 4.9. Electrons are required to pass the Medium identification WP. Surviving electrons meet the baseline criteria. Signal electrons are required to additionally pass the Tight_VarRad isolation WP. For signal samples with masses $m_a \lesssim 20$ GeV, the leptons will often be within each other's isolation cones. To maintain a high signal acceptance across the mass range, nearby leptons are excluded from isolation calculations.

4.5 Muons

A muon is a minimum-ionizing particle which will tend to pass through the inner detector, calorimeters, and muon spectrometer. Muons are reconstructed using five complementary algorithms to maintain as high efficiency as possible [14]. The five algorithms correspond to various reconstructed signatures. For example, a majority of muons are reconstructed using the "combined" technique, which performs a simultaneous re-fit of tracks in the ID and muon spectrometer (MS). If an ID track is not matched, the muon may be reconstructed as a "muon-spectrometer extrapolated" muon, which helps maintain acceptance up to $|\eta| < 2.7$, as the ID only tracks particles up to $|\eta| < 2.5$. A muon need not even rely on the MS, and a "calorimeter-tagged" muon is reconstructed by matching an ID track to a minimum-ionizing energy deposit in the calorimeter.



(a) (b)

Figure 4.2: Prompt muon efficiency at the various working points compared to muons from light hadrons as a function of (a) η and (b) p_T [14]. The MS has less coverage in the region $|\eta| < 0.1$, causing a drop in efficiency.

A reconstructed muon does not necessarily come from the primary vertex. Non-prompt muons from hadron decays are a significant source of background. Muon candidates from light hadron decays are discriminated by the muon identification WPs, as they will produce lower quality tracks. The efficiency for prompt muons is compared to muons from light hadron decays in Figure 4.2. Nearly 100% efficiency for real muons and 0.1-0.2% efficiency for light hadrons is achieved. Muons from heavy flavor (bottom/charm) decays are discriminated by isolation working points. The shape of isolation distributions for prompt and non-prompt muons is shown in Figure 4.3, demonstrating that non-prompt muons tend to be less isolated than muons from top decays.

In the $bb\tau\tau$ search, muons are required to have a minimum p_T of 7 GeV and $|\eta| < 2.7$. Similarly to electrons, no cut is applied to the d_0 significance (transverse proximity to the primary vertex) because we expect signal muons to originate in τ decays, but the cut $|z_0 \times \sin(\theta)| < 0.5$ mm (longitudinal proximity to the primary vertex) is applied. Muons passing the VERY-LOOSE identification WP are passed to the overlap removal procedure. Baseline muons are muon candidates surviving OLR which additionally pass the MEDIUM identification work-

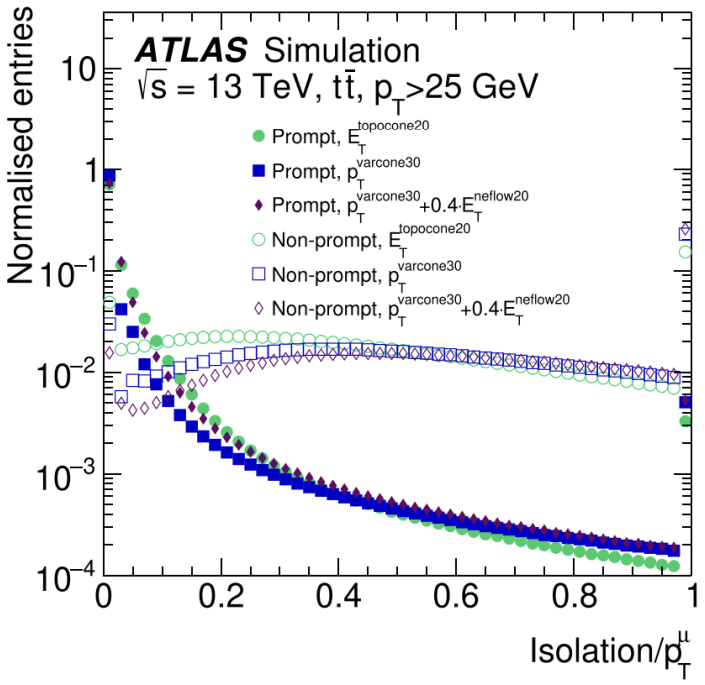


Figure 4.3: Distribution of muon isolation discriminant for prompt and non-prompt muons [14]. In this search, the discriminant $p_{\mathrm{T}}^{\mathrm{varcone30}} + 0.4E_{\mathrm{T}}^{\mathrm{neflow20}}$ was selected because it gives the best discrimination between heavy flavor backgrounds and prompt muons.

ing point. Signal muons are required to pass the isolation WP PflowLoose_FixedRad, corresponding to a cut $x < 0.16$ in Figure 4.3. Similarly to electrons, nearby leptons are excluded from isolation calculations.

4.6 Jets

"Jet" is a catch-all term for strongly interacting particles produced in a hard-scattering interaction which hadronize and deposit approximately 1/3 of their energy the EM calorimeter (via $\pi^0 \to \gamma\gamma$ decays) before depositing most of their energy in the hadronic calorimeter. Jets can be initiated by gluons, light quarks (u, d, s), and heavy quarks (c, b) (but not t quarks, which decay so quickly that they cannot hadronize). Jets from b quarks are discriminated with "flavor-tagging" algorithms, and are denoted as a b-jet for a single resolved b-tagged jet or DeXTer jet for a jet with two b-tagged sub-jets.

Fakes are mostly a concern for jets faking other objects, but not for other objects, say hadronic taus, faking jets. This is largely due to the fact that jets are produced with a large cross section at the LHC, so even a small rate of jets faking some other object can cause a sizeable background. The main concern with jet modeling in this search is in modeling the jet multiplicity. MC generators can often model the hard scatter jets (and perhaps one or two extra jets, depending on the order of the calculation), but the presence or absence of additional jets is difficult to model, and can bias the N_{jet} distribution. In this search reweighting schemes are used to compensate for some jet multiplicity mis-modeling in MC, and are discussed in more detail in Chapter V.

Jets are reconstructed using the anti-k_t [57] algorithm with $R = 0.4$. Jets are required to have a minimum p_{T} of 15 GeV and $|\eta| < 2.5$. The TIGHT Jet VerTex (JVT) WP is used. b-tagging is performed using the DL1r [70] tagger at the 85% efficient WP. b-jets have the same p_{T} and η requirements as jets.

The DL1r tagger aggregates information from several other tagging algorithms [69, 71, 72, 73, 74]. DL1r outputs p_b, p_c, and p_{light}, the probability for the jet to be a b-jet, c-jet, and light jet, respectively. The DL1r score is calculated as

$$D_{\text{DL1r}} = \ln \frac{p_b}{f_c p_c + (1 - f_c) p_{\text{light}}} \tag{4.1}$$

with f_c, the charm fraction, set to 0.018. Figure 4.4 shows the DL1r selection efficiency for b-jets, c-jets, and light jets. The b-tagging efficiency degrates at low p_{T}, and we expect low p_{T} jets in our signal events. Therefore, we choose the 85% WP because it maintains the highest possible signal acceptance.

The DeXTer [60] double b-tagger is also used to enhance sensitivity in the boosted regime when the a is low mass and the two b hadrons are close to each other. DeXTer tagged jets may also be referred to as "B"-jets. The DeXTer algorithm first reclusters ordinary anti-k_T $R = 0.4$ jets, collecting nearby energy and forming a $R = 0.8$ jet. After reclustering, precisely two sub-jets are required. DeXTer then uses a deep sets architecture and assigns a set of

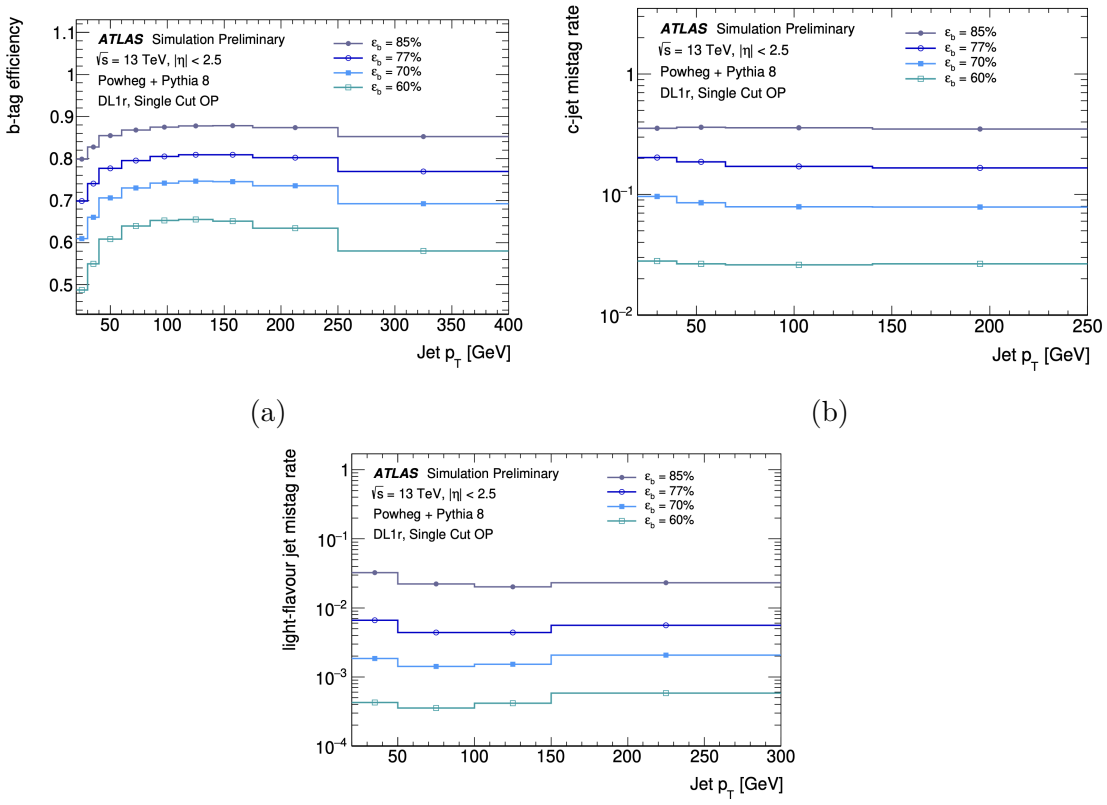


Figure 4.4: Efficiency for (a) b-jets, (b) c-jets, and (c) light jets to be accepted by DL1r in a $t\bar{t}$ MC sample. The different DL1r WPs are shown.

probability scores similarly to the DL1r algorithm. A 70% efficient WP is used for DeXTer jets. Usually the $R = 0.4$ seed jet passes the DL1r tag, so the DeXTer algorithm's primary use is not in increasing the signal acceptance. Rather, it is very uncommon in backgrounds for a second b-like object to be nearby. The large background rejection factor for $t\bar{t}$ is the primary utility of the DeXTer algorithm.

4.7 Hadronic Taus

Taus decay hadronically about 65% of the time. The decay (almost always) produces 1 or 3 charged tracks in the inner detector, which point to a displaced vertex, and the associated calorimeter cluster is usually narrow. τ_{had} candidates are seeded using jets reconstructed using the anti-k_t algorithm [57] with a distance parameter of $R = 0.4$. Candidate tracks must have a p_{T} of at least 10 GeV [75]. Tau candidates are not necessarily matched to the same primary vertex as other reconstructed objects. A dedicated tau vertex algorithm is used [65] which does not apply impact parameter requirements and finds the vertex with the largest share of the p_{T} from the τ_{had} candidate tracks ($\Delta R < 0.2$). While the tau vertex (TJVA) itself does not have any impact parameter requirements, tracks must be sufficiently close to the tau vertex, and two cuts are applied: $|d_0^{\text{TJVA}}| < 1.0$ mm and $|z_0^{\text{TJVA}} \times \sin\theta\,| < 1.5$ mm [76].

The most common sources of fake τ_{had} candidates are jets. Most jets have between 0 and 5 tracks [77], so a large number of jets contain precisely 1 or 3 tracks. Also, there are many hadrons with various lifetimes that could produce displaced vertices. Lifetimes of tau leptons and select hadrons are summarized in Table 4.2. Lifetime is not the only discriminant; for example, D^0 mesons have a very similar lifetime and mass but decay with even numbers of prongs due to its charge. Figure 4.5 shows the performance of the RNN algorithm [15] for tau identification. At the MEDIUM RNN WP, the background rejection factor is 40x for 1-prong and 300x for 3-prong tau candidates. The 3-prong tau candidates are easier to discriminate from backgrounds because there is more information for the RNN to exploit,

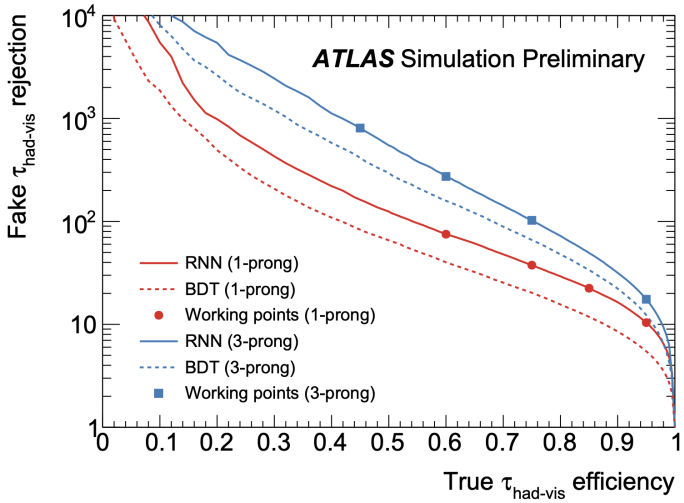


Figure 4.5: Performance of the RNN tagger for tau identification [15]. The performance is shown for 1- (red) and 3-prong (blue) τ_{had} candidates. The BDT (dashed line) and RNN (solid line) algorithms are compared. The markers show the WPs for the RNN.

Particle	Decay Length ($c\tau$, [µm])
τ	87.03
D^0	122.9
$D_s^{\pm}$	151.2
$D^{\pm}$	311.8

Table 4.2: Lifetimes for tau leptons and select hadrons [18].

compared to 1-prong tau candidates.

Additionally, leptons and photons can be a source of fake τ_{had}. Muons will naturally produce one isolated track, and a small calorimeter deposit. Photons can convert inside the inner detector and create a displaced vertex that may be identified by the TJVA algorithm. The calorimeter clusters produced by taus are very similar to electrons, so a dedicated electron BDT veto was developed by ATLAS. Without the BDT veto, electrons are misidentified as taus between 0-5% of the time, depending on p_{T} [78]. If a lepton or other particle with high OLR precedence is produced nearby a tau, for example if the same tracks were used for both candidate objects, the OLR procedure will reject the tau.

Reconstructed τ_{had} candidates are required to have $p_{\text{T}} > 20$ GeV and $|\eta| < 2.5$ with a veto

on the crack region $1.37 < |\eta| < 1.52$. Either one or three tracks are required to be associated with the τ_{had} candidate. A dedicated electron veto is applied using the ELEIDBDTMEDIUM working point. τ_{had} candidates with a minimum RNN score of 0.01 are passed to the OLR procedure. Baseline τ_{had} candidates are required to pass overlap removal. Signal τ_{had} are required to additionally pass the MEDIUM WP of the RNN tagger.

4.8 Missing Transverse Energy

As discussed in Section 2.1, stable beams circulating in a storage ring have very little transverse energy. From conservation of energy, the decay products must sum to have the same (effectively zero) transverse energy. However, not all decay products are detected, leading to what appears to be an imbalance in transverse energy. This is due to the presence of neutrinos which are effectively invisible to ATLAS. Taus decay to electrons, muons, and hadrons, but also to invisible neutrinos.

The missing transverse energy is the magnitude of the vector sum of all of the p_{T} in the event. The sum over p_{T} includes hard terms from the reconstructed objects described in the previous section and soft terms from other energy deposits not reconstructed as a high-level object. A dedicated overlap removal procedure in the $E_{\text{T}}^{\text{miss}}$ tool is used to ensure that there is no double counting.

If there are a high number of jets faking leptons, the $E_{\text{T}}^{\text{miss}}$ can be mismodeled because the fake leptons will have a high precedence in overlap removal, and these jet-to-lepton fakes are calibrated as if they were leptons. A possible solution is to tighten the isolation WPs of leptons not in the object selection but only in the $E_{\text{T}}^{\text{miss}}$ building. A change in the the WPs of the leptons was explored, but did not improve $E_{\text{T}}^{\text{miss}}$ and transverse mass modeling.

4.9 Overlap Removal

The objects described are not required to use mutually exclusive tracks or calorimeter deposits. As such, it is possible for the majority of the constituents of one object to be shared by another. To resolve the ambiguity, an overlap removal procedure is applied to the *baseline* objects described in the sections above, with a single exception: muons participating in object removal are required to pass only the VERYLOOSE identification working point. If the muon identification is not relaxed to VERYLOOSE there are a considerable number of $Z \to \mu\mu$ events in which one of the muons passes MEDIUM identification and the other passes only LOOSE or VERYLOOSE identification. If the VERYLOOSE muons are not included in overlap removal, then the $\tau_{\text{1-prong}}$ seeded by the single muon track will not be rejected, and a single muon and one τ_{had} candidate will be reconstructed. Therefore, for the purpose of overlap removal, we relax the muon identification from MEDIUM to VERYLOOSE to avoid a pollution of $Z \to \mu\mu$ events in the $\mu\tau_{\text{had}}$ channel. Muons that pass VERYLOOSE ID but fail MEDIUM ID are discarded after the overlap removal procedure is complete and no not meet our definition of a "baseline" muon. The overlap removal criteria are reported in Table 4.3.

Table 4.3: Overlap removal procedure defined in the STANDARD working selection. The steps are performed sequentially and only surviving objects participate in subsequent steps.

Step	Reject	Against	Criteria
1	electron	electron	shared track, $p_{\mathrm{T}}^{1} < p_{\mathrm{T}}^{2}$
2	τ_{had}	electron	$\Delta R < 0.2$
3	τ_{had}	muon	$\Delta R < 0.2$
4	muon	electron	is calo-muon and shared ID track
5	electron	muon	shared ID track
6	jet	electron	$\Delta R < 0.2$
7	electron	jet	$\Delta R < 0.4$
8	jet	muon	NumTrack < 3 and ghost-associated or $\Delta R < 0.2$
9	muon	jet	$\Delta R < 0.4$
10	jet	τ_{had}	$\Delta R < 0.2$
11	DeXTer jet	lepton	$\Delta R < 0.8$

CHAPTER V

Data Analysis Strategy

The previous chapters described the motivation for collecting a sample of Higgs bosons (to test BSM theory) and the experimental setup (colliding protons and recording the products). This chapter describes the strategy for searching through the resulting *events* for an exotic Higgs decay. An *event* is the result of a triggered collision. Each event is a set containing the high level *objects* discussed in Chapter IV, one event per bunch crossing[1].

The search strategy is common to particle physics. A set of Control Regions (CRs) and Signal Regions (SRs) are defined such that the new physics signal appears in the SR and does not appear in the CRs. Backgrounds appear in proportion in both CRs and SRs, and as such the CRs are used to constrain the background prediction in the SR. By counting the events in the SR, the new Higgs decay is discovered or limits on the branching fraction are set. The main experimental challenge in the $bb\tau\tau$ decay mode is that the rate of fakes is high due to the abundance of background processes in the low p_{T} region.

The methodology is as follows:

1. Use leptons from τ decays to trigger the ATLAS data acquisition system due to the difficulty of triggering on jets and τ_{had}, especially at low p_{T}. The data and simulated MC samples used are described in Section 5.1.

2. Select events in the $e\mu$, $\mu\tau_{\mathrm{had}}$, and $e\tau_{\mathrm{had}}$ decay channels of the $\tau\tau$ system. Further

[1]Not all events are saved by the TDAQ system

categorize events based on the experimental signature observed from the bb decay: (a) a merged double B-jet, (b) a single resolved b-jet, and (c) two resolved b-jets. The heavy-flavor categorization increases sensitivity in the resolved and boosted regimes (the high and low mass extremes for the exotic pseudoscalar). Event selection and trigger strategy are discussed in Section 5.2.

3. Model events with two prompt leptons or hadronic tau decays from SM processes using MC. Control regions with high transverse and $\tau\tau$ masses are defined to estimate the contribution from $t\bar{t}$ and Z+jets, respectively. Control regions are orthogonal to signal regions and have very little signal pollution.

4. Model events with non-prompt ("fake") leptons and taus using a data-driven method. Selections on isolation and/or identification criteria are inverted to create control regions rich in non-prompt leptons and hadronic taus. Empirical transfer factors are used to determine the number of events in the signal region from the number of events in the non-prompt control regions. Both the MC and data-driven background model are discussed in Section 5.4.

5. In the signal region, defined by the presence of a heavy-flavor tagged jet, low transverse mass, and low visible $\tau\tau$ mass, a neural network discriminant is trained to discriminate the exotic Higgs signal (modeled using MC) and the backgrounds. A cut on the *minimum* neural network output score defines an exclusive signal region with high signal-to-background ratio. A cut on the *maximum* neural network output score defines a background-enriched region to validate the model's extrapolation to low $\tau\tau$ masses. Section 5.5 discusses the neural network discriminant.

6. Estimate systematic uncertainties. Systematic uncertainties are applied as nuisance parameters in the maximum likelihood fits. Systematic uncertainties cover detector response, efficiencies of identification and triggers, parameterization of the data-driven

backgrounds, and more. A summary of sources of systematic uncertainty is provided in Section 5.6. Model validation and fits are shown in Section 5.7.

7. Perform statistical analysis using a Profile Likelihood Ratio (PLR) test. The $\tau\tau$ mass is a resonant variable for the mass of the BSM pseudoscalar m_a. Using a fit to the $\tau\tau$ mass distribution we extract the upper limit on the branching fraction $h \to aa \to bb\tau\tau$ as a function of m_a. Observed limits are presented and interpreted under the 2HDM+S model in Chapter VI.

Additional validation plots for the data analysis strategy are shown in Appendix A. Moving forward, the data analysis strategy is referred to as the "analysis."

5.1 Data and Simulated Samples

5.1.1 Data

The data used in this thesis was collected during Run 2 of the LHC and corresponds to the full pp collision data set from the 2015-2018 data-taking period. This data set has a center-of-mass energy of the pp collisions of $\sqrt{s} = 13$ TeV. Only events which pass data quality requirements that ensure the stable operation of the ATLAS detector are considered. The data set after these requirements corresponds to an integrated luminosity of approximately 139 fb^{-1}. The year and corresponding integrated luminosity ($\int \mathcal{L}\, \mathrm{d}t$) are found in Table 5.1.

Year	$\int \mathcal{L}\, \mathrm{d}t$ (fb^{-1})
2015	3.21
2016	32.99
2017	44.31
2018	58.45

Table 5.1: Integrated luminosity by data-taking year.

5.1.2 Monte Carlo Simulations

Simulated MC events are used to estimate several SM backgrounds. The main SM backgrounds consist of Drell-Yan (DY) production with decays to taus $Z/\gamma^* \to \tau\tau$ produced in association with jets, top production ($t\bar{t}$ or single-top) events where the W bosons decay leptonically, and backgrounds where jets are misidentified as leptons or hadronic taus. Minor backgrounds from the Higgs boson decays to taus $H \to \tau\tau$, $DY \to \ell\ell$ decays, diboson production, and vector bosons produced in association with $t\bar{t}$ (ttV) are also considered.

The SHERPA [79] generator is used for the DY+jets (v 2.2.11), W($\to (e, \mu, \tau)\nu$)+jets (v2.2.11), and diboson (v2.2.1 and v2.2.2) background processes. The V+jets setup uses next-to-leading order (NLO) matrix elements (MEs) for up to two partons, and leading order (LO) matrix elements for up to five partons in the five-flavour scheme calculated with the Comix [80] and OPENLOOPS [81, 82, 83] libraries. In the diboson setup, fully leptonic final states and semileptonic final states were generated using matrix elements at NLO accuracy in Quantum Chromodynamics (QCD) for up to one additional parton and at LO accuracy for up to three additional parton emissions. The virtual QCD corrections for the diboson setup were provided by the OPENLOOPS library [81, 82, 83]. Samples for the loop-induced processes $gg \to VV$ were generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix elements are matched and with the SHERPA parton shower based on Catani–Seymour dipole factorisation [80, 84] using the MEPS@NLO prescription [85, 86, 87, 88] using the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs [89] was used for both setups and the V+jets samples are normalized to a next-to-next-to-leading order (NNLO) prediction [90]. In the DY+jets configurations, dedicated samples were produced where the invariant mass of the two leptons are 10 GeV $\leq m_{\ell\ell} \leq$ 40 GeV and $m_{\ell\ell} >$ 40 GeV. Similarly, the diboson events are produced with a minimum invariant mass requirement of $m_{\ell\ell} >$ 4 GeV. The DY+jets samples are additionally split according to the $\tau\tau$ decays, with dedicated samples for $\tau_{\text{lep}}\tau_{\text{lep}}$ and $\tau_{\text{lep}}\tau_{\text{had}}$.

The production of $t\bar{t}$ events was modelled using the POWHEG BOX V2 [91, 92, 93, 94] generator at NLO in QCD with the NNPDF3.0NLO [89] PDF set and the h_{damp} parameter was set to $1.5\,m_t$ [95]. The events were interfaced to PYTHIA 8.230 [96] to model the parton shower, hadronization, and underlying event, with parameters set according to the A14 tune [97] and using the NNPDF2.3LO set of PDFs [98]. The decays of bottom and charm hadrons were performed by EVTGEN 1.6.0 [99].

Single-top events are split according to the different production modes: s-channel, t-channel, and in association with a W boson (tW). All three production modes are modelled using POWHEG BOX V2 [92, 93, 94, 100, 101, 102] at NLO in QCD with the NNPDF3.0NLO [89] parton distribution function (PDF) set. For the s-channel and W-associated production, the calculation is performed in the five-flavor scheme, while the t-channel production uses the four-flavor scheme. A special diagram removal scheme [103] was used in the generation of the tW sample to remove interference and overlap with $t\bar{t}$ production. All of the single-top events were interfaced with PYTHIA 8.230 [96] using the A14 tune [97] and the NNPDF2.3LO PDF set.

This analysis considers the Higgs boson production for the signal in three main processes at the LHC: gluon-gluon fusion (ggF), vector boson fusion (VBF), and the associated production with a vector boson (VH). The ggF process has the largest expected production cross-section of 48.52 pb, followed by the VBF process with 3.779 pb and then by VH with 2.253 pb. For the $H \to \tau\tau$ process, only the ggF and VBF samples are considered.

The Higgs boson production via gluon fusion is generated at NNLO accuracy in QCD using POWHEG BOX v2 [104, 105, 94, 92, 93]. The simulation achieved NNLO accuracy for arbitrary inclusive $gg \to H$ observables by reweighting the Higgs boson rapidity spectrum in HJ-MINLO [106, 107, 108] to that of HNNLO [109]. The PDF4LHC15NNLO PDF set [110] and the AZNLO tune [111] of PYTHIA 8 [96] were used. The gluon–gluon fusion prediction from the Monte Carlo samples is normalized to the next-to-next-to-next-to-leading-order cross-section in QCD plus electroweak corrections at NLO [112, 113, 114, 115, 116, 117, 118,

119, 120, 121, 122, 122].

Higgs boson production via vector-boson fusion and in association with a vector boson was simulated with POWHEG BOX v2 [123, 94, 92, 93] and interfaced with PYTHIA 8 [96] for parton shower and non-perturbative effects, with parameters set according to the AZNLO tune [111]. The POWHEG predictions for VBF and VH plus one-jet production are accurate to next-to-leading order (NLO) and both samples use the PDF4LHC15NLO PDF set [110]. It was tuned to match calculations with effects due to finite heavy-quark masses and soft-gluon resummation up to NNLL. The loop-induced $gg \to ZH$ process was not generated and is expected to have a small contribution. The VBF Higgs MC cross-section prediction is normalized to an approximate-NNLO QCD cross-section with NLO electroweak corrections [124, 125, 126]. Similarly, the VH MC prediction is normalized to cross-sections calculated at NNLO in QCD with NLO electroweak corrections for $q\bar{q}/qg \to VH$ and at NLO and next-to-leading-logarithm accuracy in QCD for $gg \to ZH$ [127, 128, 129, 130, 131, 132, 133].

For the signal samples, the decay $H \to aa \to bb\tau\tau$ is performed using PYTHIA 8. Eight distinct mass points are simulated for ggH, VBF, and the VH production modes in the range $m_a = 12 - 60$ GeV (12, 15.5, 20, 25, 30, 40, 50 and 60 GeV). The tau decays in these samples are restricted to $e\mu$, $\tau_h\ell$, and $\ell\tau_h$ decays where $\ell = e, \mu$. These decays are separated into different samples in order to enhance the amount of statistics available for the search.

The effect of multiple interactions (pileup) is modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with PYTHIA 8.186 [134] using the NNPDF2.3LO set of parton distribution functions (PDF) [98] and the A3 set of tuned parameters [135].

All simulated event samples were processed through the ATLAS detector simulation [136] based on GEANT4 [137].

The decays and spin correlations for the τ-leptons are handled by SHERPA for all samples it generated and by PYTHIA for all other MC samples. The decays and spin correlations of the τ-lepton are included in PYTHIA since version 8.150 [138].

An overview of the MC generators used for the main signal and background samples is shown in Table 5.2.

Process	Generator		PDF set		Tune	Order
	ME	PS	ME	PS		
$H \to aa \to bb\tau\tau$						
ggF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO +NNLL
VBF	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	(N)NLO
VH	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO
Background						
ggF $H \to \tau\tau$	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	NNLO +NNLL
VBF $H \to \tau\tau$	Powheg	Pythia8	PDF4LHC15	CTEQ6L1	AZNLO	(N)NLO
V+jets	Sherpa 2.2.11		NNPDF30		Sherpa	NNLO
$t\bar{t}$	Powheg	Pythia8	NNPDF2.3		A14	NLO
Single top	Powheg	Pythia8	NNPDF2.3		A14	NLO
Diboson	Sherpa 2.2.2/2.2.1		NNPDF30		Sherpa	NNLO

Table 5.2: Overview of the MC generators used for the main signal and background samples.

5.2 Event Selection

This analysis uses events classified in three separate channels defined by the final state objects of the $\tau\tau$ system: $e\mu$, $\mu\tau_{\text{had}}$, and $e\tau_{\text{had}}$. The signal regions for these three channels are based on the number of signal leptons and τ_{had} candidates and are orthogonal to each other. The three final states of the $\tau\tau$ system ($e\mu$, $\mu\tau_{\text{had}}$, and $e\tau_{\text{had}}$) cover 52% of the branching fraction of di-tau decays. The fake lepton/tau control regions (which invert isolation and/or identification requirements to derive the data-driven backgrounds) may share events between the three channels. For example, an event with an electron, muon, and τ_{had} candidate which each pass the baseline requirement but fail the signal requirement is eligible for inclusion in all three of the fake template CRs, but none of the signal regions because there are no signal leptons or taus. Next we discuss the trigger selection and the definitions of the control, validation, and signal regions for each channel in more detail.

5.2.1 Trigger Selection

ATLAS uses a two-level trigger system [139]. At first in level 1 (L1) coarse thresholds are applied to reduce the event rate to a managable rate. Then at the second level the High Level Trigger (HLT) is used to reconstruct portions of the event using fast algorithms and apply a higher p_{T} threshold to reduce the rate further. The L1 trigger is used to reduce the number of events fed to the slower algorithms in the HLTs. ATLAS implements several triggers in a *trigger menu*, which is updated every year [140, 141, 142, 143]. In this work a further threshold is applied to symmetrize the trigger menus from different years; for example, in 2015 and 2016 data the single lepton HLT threshold was 25 GeV and in 2017 and 2018 the threshold was raised to 27 GeV. An offline cut at 27 GeV is applied to override the slight difference in data from different years.

In this analysis events are selected using two main types of HLTs: single lepton trigger (SLT) and Opposite Flavor (OF) lepton triggers (di-lepton trigger (DLT)). The HLTs used are summarized in Table 5.3. A HLT string lists the species of particle and the threshold applied, for example HLT_MU50 means that to pass this trigger there must be at least one 50 GeV muon in the event. Other properties of the reconstruction algorithm are listed next, for example HLT_E60_LHMEDIUM means that there must be an electron with 60 GeV of p_{T} which also passes the MEDIUM likelihood identification working point.

For the $e\mu$ channel, three trigger regions are defined based on which trigger is passed and the p_{T} threshold of the objects in the event. These requirements and regions are further described in Section 5.2.2. In the $\mu\tau_{\mathrm{had}}$ channel only single muon triggers are used because there is no electron expected in the final state. Similarly, in the $e\tau_{\mathrm{had}}$ channel only single electron triggers are used. The event selection requirements for the channels including a hadronic tau are further described in Section 5.2.3.

Next, the procedure for matching reconstructed objects to the triggers which fired, the application of p_{T} thresholds to trigger matched objects, and the identification of other reconstructed objects of interest is described. Each event may have precisely one of each muon,

electron, and τ_{had} candidate identified as the particle of that species which is "of interest." The procedure for identifying these particles is as follows:

1. The τ_{had} candidate with an RNN score > 0.01 is identified as the τ_{had} of interest for the event. Events with more than one τ_{had} passing this criteria are rejected.

2. All possible sets of trigger-matched leptons are constructed. Each HLT in Table 5.3 is checked to see if the run number corresponds to the active period for that trigger. Any permutation where there is an appropriate number of leptons trigger-matched and 1 GeV above the threshold defined in the HLT is allowed. In order to have consistent selections across data-taking periods, if 1 GeV above the threshold is less than 27 GeV for the single lepton trigger used, the threshold is raised to 27 GeV. In the case of the di-lepton trigger, a p_{T} maximum of 27 GeV is also applied to avoid conflict with the single lepton triggers, as shown in Figure 5.3.

Table 5.3: High level triggers used for the single muon, single electron, and di-lepton triggers as defined for this analysis. The single muon trigger is used in the $e\mu$ and $\mu\tau_{\text{had}}$ final states. The single electron trigger is used in the $e\mu$ and $e\tau_{\text{had}}$ final states. The di-lepton trigger is only used for the $e\mu$ final state.

Trigger Type	Year	Run Numbers	HLTs
Single Muon	2015	276262-284484	HLT_mu20_iloose_L1MU15 HLT_mu50
	2016-2018	297730-364292	HLT_mu26_ivarmedium HLT_mu50
Single Electron	2015	276262-284484	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose
	2016-2018	297730-364292	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0
Di-lepton	2015	276262-284484	HLT_e7_lhmedium_mu24 HLT_e17_lhloose_mu14
	2016-2018	297730-364292	HLT_e7_lhmedium_nod0_mu24 HLT_e17_lhloose_nod0_mu14

3. The sets of trigger-matched leptons are sorted using the following procedure. First, signal leptons are prioritized over baseline leptons. Next, the sets of leptons are sorted by the trigger type, with single electron triggers having the highest precedence, then signal muon triggers and finally di-lepton triggers. A tie is broken by the sum of the p_{T} of the leptons.

4. A maximum of one electron and one muon is trigger matched in the highest priority lepton pair. These are the leptons of interest.

5. In the case of single lepton triggers, there will be only an electron or a muon trigger matched. The same precedence applies to select the lepton of interest which is the opposite flavor to the trigger matched lepton (prioritize signal leptons over baseline, and break the tie with p_{T}). No p_{T} threshold is applied to leptons which are not trigger matched other than the minimum p_{T} in Table 4.1. In the case of di-lepton triggers, this step is omitted since both leptons must be trigger-matched.

Whenever kinematic variables such as $m_{e\mu}$ or $\Delta R_{\mu\tau_{\mathrm{had}}}$ are constructed, their distributions use the single electron, muon or τ_{had} candidate identified by this procedure. For jets and b-jets, the highest p_{T} candidate is used for constructing high-level kinematic variables. When a kinematic variable is defined by two jets, m_{bb} for instance, the highest two p_T jets are used. Jets aside from the leading and subleading jets are not used for event selection purposes or for the construction of any kinematic variable (other than H_T, the scalar sum of the p_{T} of all reconstructed objects). When an event is sorted into a control, validation, or signal region of a given channel, this sorting is also performed using the leptons of interest. Additional leptons (those which have not been identified as the leptons of interest) which pass the baseline but fail the signal requirement are not cause for an event to be rejected.

5.2.2 $e\mu$ Channel

In the $e\mu$ channel there is a SR and four CRs: Same Sign (SS) CR, $t\bar{t}$ CR, high mass CR, and 0 b-jet Validation Region (VR) region. Signals will tend to have low transverse masses of the tau visible decay products and the missing transverse energy. The variable Σm_T is defined as:

$$\Sigma m_T = m_T(\tau^1_{\text{vis}}, E^{\text{miss}}_{\text{T}}) + m_T(\tau^2_{\text{vis}}, E^{\text{miss}}_{\text{T}}) \tag{5.1}$$

where the pairings of the transverse masses are given by $(\tau^1_{\text{vis}}, \tau^2_{\text{vis}}) = (e, \mu)$ and is analogous for the $\ell\tau_{\text{had}}$ channels. This variable will be used in the definitions of the signal and control regions. Next we describe the $e\mu$ signal region and then the control and validation regions are described in relation to the signal region. Events in the $e\mu$ SR are required to pass the following cuts:

- Precisely one Opposite Sign (OS) electron and muon pair with each lepton meeting the signal selection criteria. Events are rejected if there are any same-flavor signal lepton pairs.
- 0 signal τ_{had} candidates and no more than 1 baseline hadronic tau candidate.
- At least 1 b-tagged jet, which can be a DeXTer jet or a DL1r jet.
- A minimum angular separation on the leptons, $\Delta R(e, \mu) > 0.1$, and a visible $e\mu$ mass cut of 4 GeV are imposed to reduce backgrounds from low-mass hadronic resonances.
- $\Sigma m_T < 120$ GeV, which orthogonalizes with respect to the $t\bar{t}$ CR.
- The visible mass of the decay products $m_{e\mu} < 45$ GeV.

Note that the cut $\Sigma m_T > 120$ GeV maintains a high purity of $t\bar{t}$ and $t\bar{t}$-initiated fakes in the $t\bar{t}$ CR without significant signal leakage into the CR. The distribution of Σm_T is shown in Figure 5.1 (in this plot the signal histogram is scaled up by ≈100x compared to $\text{Br}(h \to bb\tau\tau) < 10\%$ allowed by Higgs coupling constraints [28, 29]).

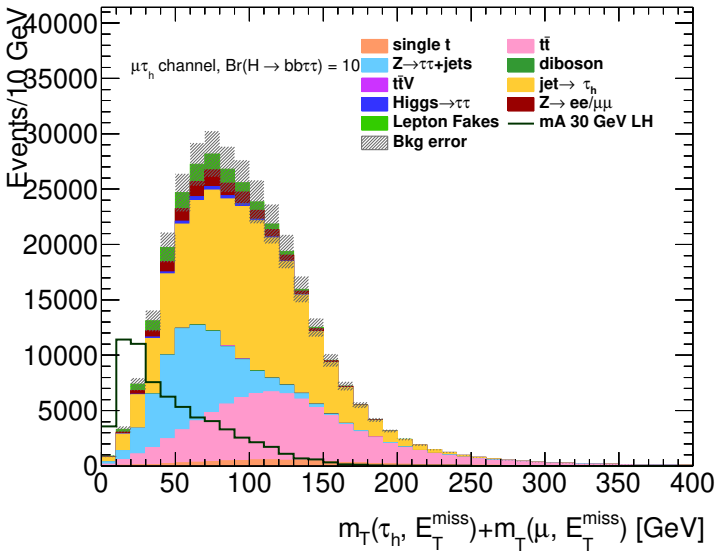


Figure 5.1: Example of the Σm_T distribution for the $\mu\tau_{\text{had}}$ channel. The $t\bar{t}$ CR is defined by $\Sigma m_T > 120$, and the SR and high mass CR both have the cut $\Sigma m_T < 120$ applied.

Next, the four control regions are described in relation to the signal region. A diagram of the DY and top CRs compared to the SR and high mass CR is shown in Figure 5.2.

1. The high mass CR inverts the cut on $m_{e\mu}$ to be $m_{e\mu} > 45$ GeV. The high mass CR is used to constrain the normalization on $Z \to \tau\tau$. Other CRs are agnostic to the $m_{e\mu}$ cut and cover the same phase space in $m_{e\mu}$ as the SR + high mass CR, as shown in Figure 5.2.

2. The $t\bar{t}$ CR inverts cut on Σm_T so that events with $\Sigma m_T > 120$ GeV populate the $t\bar{t}$ CR. The $t\bar{t}$ CR enters the maximum likelihood fit to constrain the normalization on $t\bar{t}$ and single top backgrounds.

3. The 0 b-jet VR inverts the requirement on heavy flavor jets, but maintains a minimum of at least 1 jet (in this case it will be a $R = 0.4$ jet which is not DL1r tagged and does not seed a Loose DeXTer jet). The 0 b-jet VR validates the ability of the background model to extrapolate into the low mass region. The 0 b-jet VR also covers the entire phase space of $m_{e\mu}$.

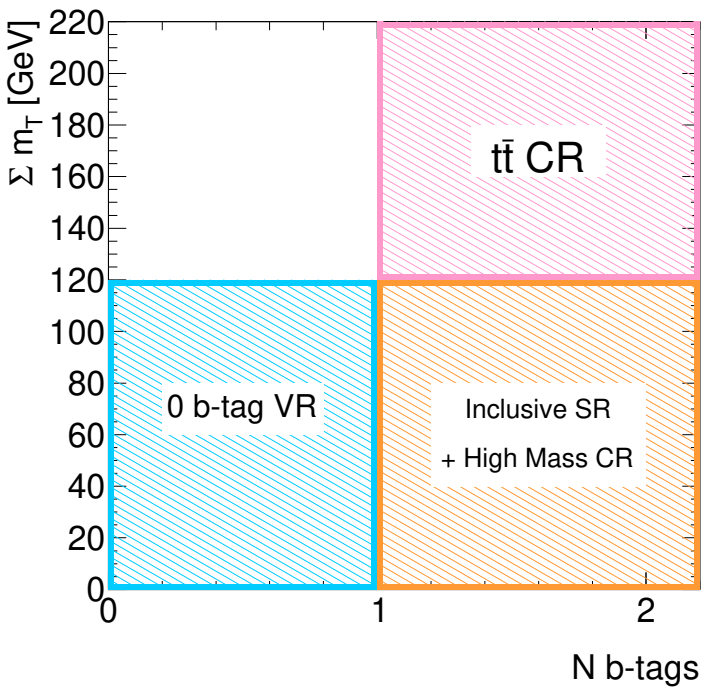


Figure 5.2: Control and validation regions used for DY and $t\bar{t}$ in all channels. There is no upper limit to Σm_T for the $t\bar{t}$ CR, nor is there an upper limit on the number of b-jets in the $t\bar{t}$ CR or signal region. In this context, DeXTer and DL1r jets are counted as a "b-tag."

4. The SS CR has an identical selection to the $e\mu$ SR but with the OS requirement inverted to SS and the low mass cut removed to cover the phase space of the SR and high mass CR in $m_{e\mu}$. The SS CR is used to derive lepton efficiencies for the non-prompt lepton estimation.

In the $e\mu$ channel (and only the $e\mu$ channel) three orthogonal trigger regions are defined. TRIG1 defines the region where the electron p_{T} is greater than 27 GeV. In this region the single electron must be trigger matched. In the region TRIG2, the electron must have less than 27 GeV of p_{T} and the muon must have $p_{\mathrm{T}} > 27$ GeV. The single muon trigger must be trigger matched in TRIG2. When both leptons have $p_{\mathrm{T}} < 27$ GeV, one of the di-lepton triggers must be trigger matched; this di-lepton trigger region is called TRIG3. The trigger regions are illustrated schematically in Figure 5.3.

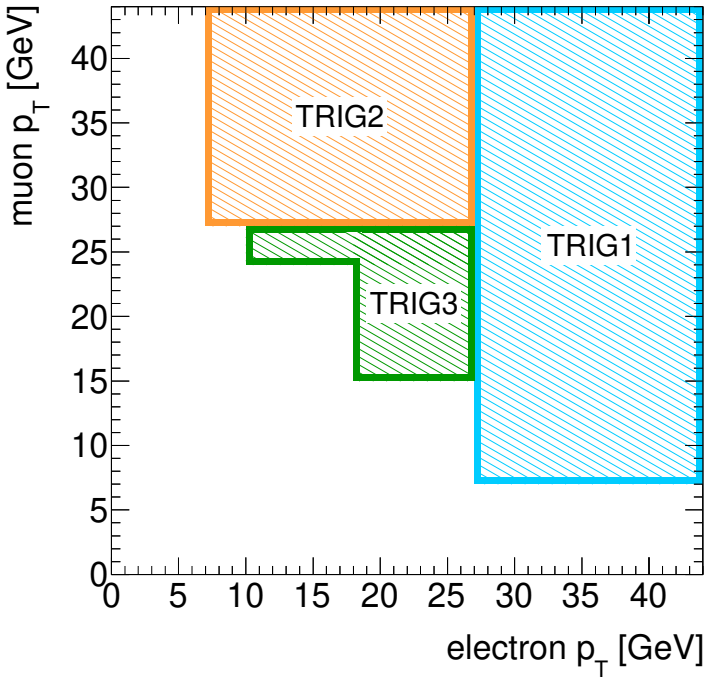


Figure 5.3: Orthogonalization scheme for the three trigger regions in the $e\mu$ channel.

5.2.3 $\ell\tau_{\text{had}}$ Channels

The two $\ell\tau_{\text{had}}$ channels have identical selections, save for the flavor of the lepton $\ell = e, \mu$. In this section we describe the event selection for both channels in parallel. In the $\ell\tau_{\text{had}}$ channels a SR, SS CR, $t\bar{t}$ CR, high mass CR, and 0 b-jet VR region are defined (similarly to the $e\mu$ channel). First the $\ell\tau_{\text{had}}$ signal region is described and then the control and validation regions are described in relation to the signal region. Events in the $\ell\tau_{\text{had}}$ SR are required to pass the following cuts:

- Precisely one OS $\ell\tau_{\text{had}}$ pair with the lepton and tau both meeting the signal selection criteria. Events are rejected if there is more than one signal lepton or more than 1 baseline hadronic tau candidate.
- At least 1 b-tagged jet, which can be a DeXTer jet or a DL1r jet.
- A minimum angular separation on the lepton and hadronic decay products, $\Delta R(\ell, \tau) > 0.2$, and a visible $\ell\tau_{\text{had}}$ mass cut of 4 GeV are applied to avoid reconstructing the same

object as both a lepton and a tau and avoid low-mass hadronic resonances. The angular separation cut is consistent with the OLR procedure described in Section 4.9.

- $\Sigma m_T < 120$ GeV, which orthogonalizes with respect to the $t\bar{t}$ CR. See Eq. 5.1 for the definition of Σm_T.

- The visible mass of the decay products $m_{\ell\tau} < 60$ GeV. In the $e\mu$ channel this cut was at 45 GeV because there is another neutrino, biasing the visible mass to be lower than the $\ell\tau_{\text{had}}$ channels.

Now the four CRs are described in relation to the SR. Note that the definitions of control regions match the $e\mu$ channel *exactly*, save for the cut on the visible decay products to be at 60 GeV for $\ell\tau_{\text{had}}$ instead of 45 GeV for $e\mu$.

1. The high mass CR inverts the cut on $m_{\ell\tau_{\text{had}}}$ to be $m_{\ell\tau_{\text{had}}} > 60$ GeV. The high mass CR is used to constrain the normalization on $Z \to \tau\tau$, $Z \to \ell\ell$, and jet$\to \tau_{\text{had}}$ fakes. Other CRs are agnostic to the $m_{\ell\tau_{\text{had}}}$ cut and cover the same phase space in $m_{\ell\tau_{\text{had}}}$ as the SR + high mass CR, as shown in Figure 5.2.

2. The $t\bar{t}$ CR inverts cut on Σm_T so that events with $\Sigma m_T > 120$ GeV populate the $t\bar{t}$ CR. The $t\bar{t}$ CR enters the maximum likelihood fit to constrain the normalization on $t\bar{t}$ and single top backgrounds.

3. The 0 b-jet VR inverts the requirement on heavy flavor jets, but maintains a minimum of at least 1 jet (in this case it will be a $R = 0.4$ jet which is not DL1r tagged and does not seed a Loose DeXTer jet). The 0 b-jet VR validates the ability of the background model to extrapolate into the low mass region. The 0 b-jet VR also covers the entire phase space of $m_{\ell\tau_{\text{had}}}$.

4. The SS CR has an identical selection to the $\ell\tau_{\text{had}}$ SR but with the OS requirement inverted to SS and the low mass cut removed to cover the phase space of the SR and

high mass CR in $m_{\ell\tau_{\text{had}}}$. The SS CR is used to derive the composition of the jet$\rightarrow \tau_{\text{had}}$ fakes (relative fraction of gluons, light hadrons, and heavy flavor jets).

In the $\ell\tau_{\text{had}}$ region only the the signal lepton trigger with corresponding flavor is used. No further orthogonalization of trigger regions is required.

5.2.4 Event Categorization

In all three channels, events are sorted into three categorizes for statistical interpretation based on b-tagging multiplicity. There are two categories for events in which 1 and 2 b-jets are individually resolved. A boosted category with the DeXTer [60] double b-tagger increases the sensitivity for $m_a < 20$ GeV, where the two b-jets merge into each other.

For the $h \rightarrow aa \rightarrow bb\tau\tau$ topology, many of the reconstructed objects will have low p_{T} and it is common for one of the b-jets to fall below the reconstruction threshold. A cut requiring events to have two b-jets would greatly reduce signal acceptance. However, large and difficult to discriminate backgrounds exist in the 1 b-jet category, namely Drell-Yan, fake lepton and jet$\rightarrow \tau_{\text{had}}$ backgrounds. $t\bar{t}$ will populate the two b-tagging category, but this background is easier to discriminate due to high-mass and high-p_{T} objects originating in top decays. To strike a balance between exploiting the additional information of a second b-tag and preserving high signal acceptance, 1- and 2- b-tag categories are defined. Additionally, low m_a signals, in contrast to background, will often have overlapping b-jets. In this case, the DeXTer double b-tagger is used to enhance sensitivity. Figure 5.4 shows the heavy-flavour categorization scheme. For the DeXTer category, precisely one double B-jet is reconstructed. For the > 1 b-jet category no upper limit is placed on the number of DL1r or DeXTer tags.

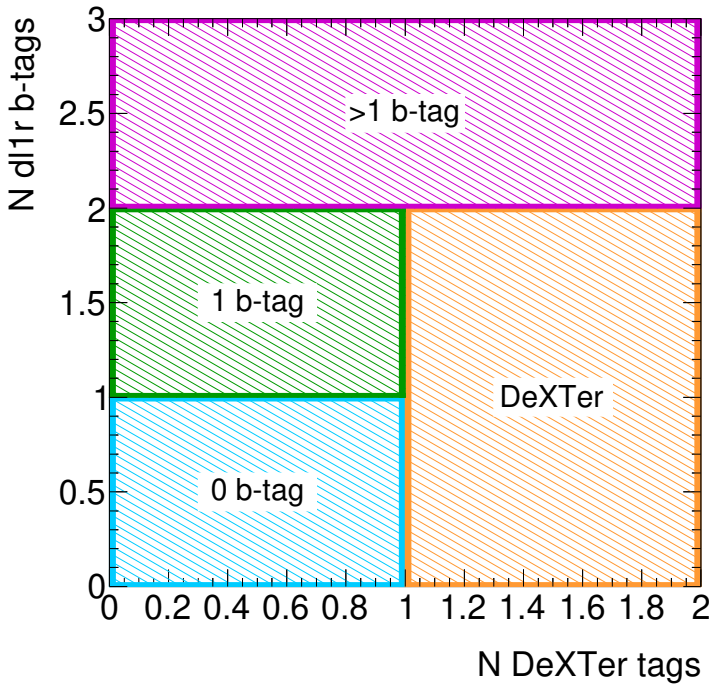


Figure 5.4: Category definitions for the three heavy flavour signal categories (1 b-jet, >1 b-jet, and DeXTer category) and the 0 b-jet CR for all channels.

5.3 Signal Modeling

A model for the signal decay mode $h \to aa \to bb\tau\tau$ is generated using MC in the ggF, VBF, ZH, and WH production modes of the Higgs boson as described in Section 5.1.2. In this section, the acceptance and kinematics of the signal are discussed.

As discussed in Section 5.2.1, lepton triggers are used for each of the analysis channels. Figure 5.5 shows the trigger efficiency across the spectrum m_a from 12 to 60 GeV for the dominant ggF production mode. The trigger efficiency is relatively democratic across the mass spectrum; however, the expected *sensitivity* is not flat across the mass spectrum. In the "low" (from 12 to ≈20 GeV) end of the mass spectrum, the signal acceptance is reduced due to the minimum ΔR cut between the visible tau decay products ($\Delta R > 0.1$ for $e\mu$ and $\Delta R > 0.2$ for $\ell\tau_{\text{had}}$). In the "high" (> 45 GeV) end of the mass spectrum background counts are much greater because the dominant SM processes ($Z \to \tau\tau, t\bar{t}$) involve higher mass objects. The sensitivity is further enhanced in both the low and high ends of the mass

spectrum by the categorization based on the tagging the jets originating from b-quarks, as discussed in Section 5.2.4. This flattens out the expected sensitivity on the branching fraction with respect to the mass of the a.

Due to collimation of the b-jets for low values of m_a, the DeXTer double b-tagger is needed to enhance the sensitivity in this region. The double tagger reduces the amount of background by using the substructure of a single $R = 0.8$ radius jet, which has two subjets that are identified as b-tagged jets. A modest increase in signal acceptance is also seen, but the background rejection is the primary driver of the sensitivity increase at low mass. On the other side of the mass spectrum, the b-jets are no longer collimated and resolved $= 1$ b-jet and > 1 b-jet categories are employed. Figure 5.6 shows the b-jet multiplicity across the signal mass spectrum in the $e\mu$ decay channel (inclusive of Higgs production mode). It is more common for higher mass signals to have a second b-jet. This channel is representative of other decay modes and demonstrates that two b-jets are resolved more often at higher masses. The b-jet categorization increases the sensitivity for higher masses because the two b-jet category is generally populated by the $t\bar{t}$ background, which is easier to discriminate from the signal process using kinematic information. This discrimination is accomplished using a parameterized neural network as discussed in Section 5.5.

Additional discrimination power can be gained from particular features of the $\tau\tau$ decays. The two taus decay with a total of three or four neutrinos in the $\ell\tau_{\text{had}}$ and $e\mu$ final states, respectively. The visible mass (the mass of the visible decay products, which excludes neutrinos) is shown in Figure 5.7 and is biased due to the missing energy from the neutrinos. The peak of each of the 8 signal models is well below the simulated mass (for example, the purple histogram is a simulation of a signal with $m_a = 60$ GeV, but the visible mass histogram peaks at 36 GeV, not 60 GeV). A Missing Mass Calculator (MMC) [144] is used to estimate the four-vectors of the neutrinos. The neutrino four-vectors encapsulate 8 unknowns in the $e\mu$ channel (two four-vectors) and 7 unknowns in the $\ell\tau_{\text{had}}$ channels (the single neutrino is known to have essentially zero mass). There are four constraints on the 7 (or 8) unknowns:

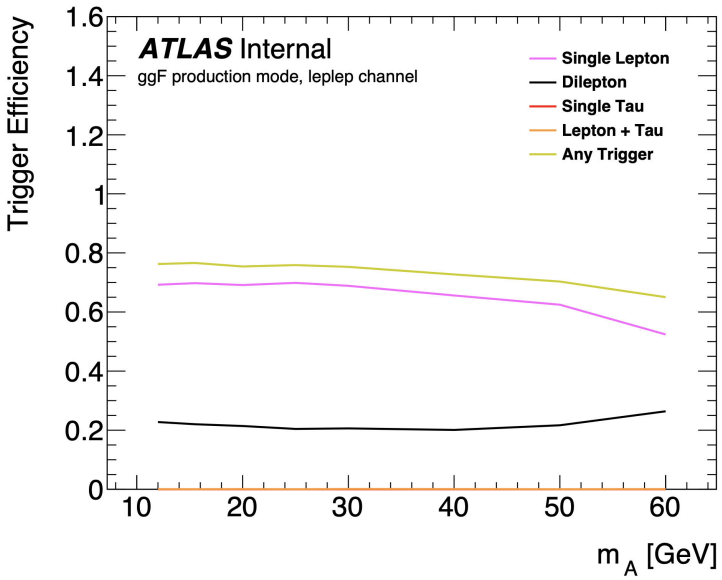


Figure 5.5: Trigger efficiency in the ggF production mode for the $e\mu$ channel

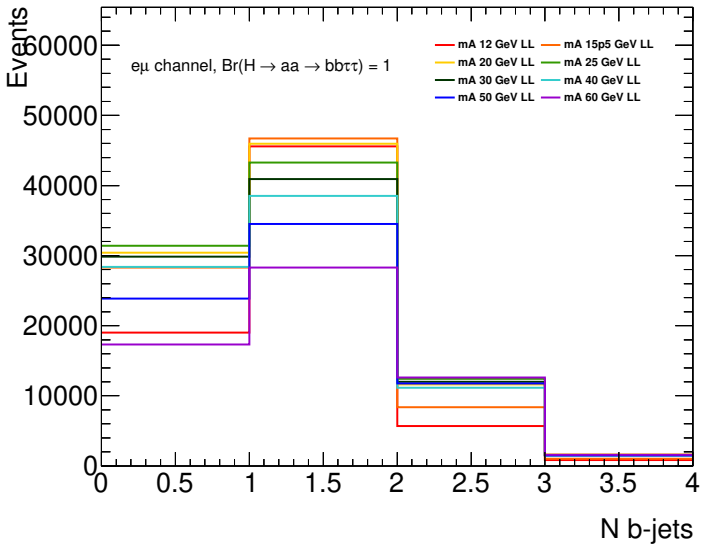


Figure 5.6: Multiplicity of DL1r tagged (anti-k_t $R = 0.4$) jets for different simulated mass samples.

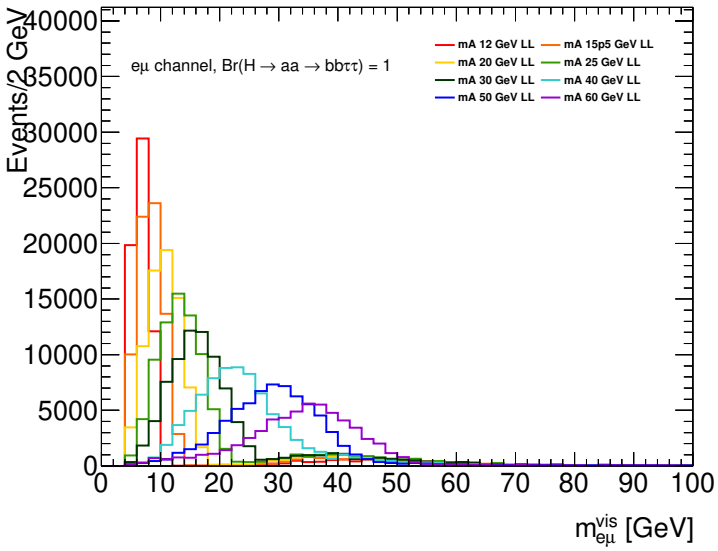


Figure 5.7: Visible mass of the $e\mu$ system in simulated signal samples in the $e\mu$ channel.

the visible + invisible tau decay must sum to the correct tau mass (1.776 GeV), and the sum of the neutrino p_{T} in the transverse plane must equal the missing transverse energy. The problem is under-constrained, but not all solutions which satisfy the 4 constraints are equally likely. The MMC uses a Monte-Carlo Markov Chain to sample from simulated distributions of the angular separation between the visible and invisible components of the tau decays. The result is a most-likely mass hypothesis for the di-tau system and a most likely neutrino four-vector for each of the two neutrino systems. Figure 5.8 shows the most-likely mass, given the visible kinematics and the $E_{\mathrm{T}}^{\mathrm{miss}}$. This demonstrates that the MMC estimates the mass in an unbiased manner because the histograms now peak at the correct mass hypothesis (although detector resolution smears the mass distributions out significantly). An example of the mass of the $\nu_e + \nu_\tau$ system is shown in Figure 5.9. The mass of the signal model is narrower than the $t\bar{t}$ background because $t\bar{t}$ contains higher p_{T} taus (due to the large t mass), and the invisible decay products are more co-linear with the visible decay products.

A further discriminating variable is D_ζ, which captures to what extent the $E_{\mathrm{T}}^{\mathrm{miss}}$ is in the same direction as the taus. D_ζ is defined as:

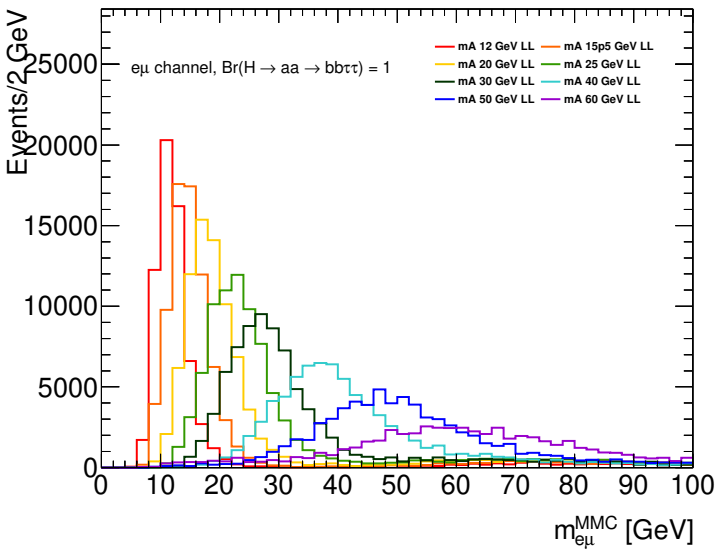


Figure 5.8: MMC most likely mass of the $e\mu$ system in simulated signal samples in the $e\mu$ channel.

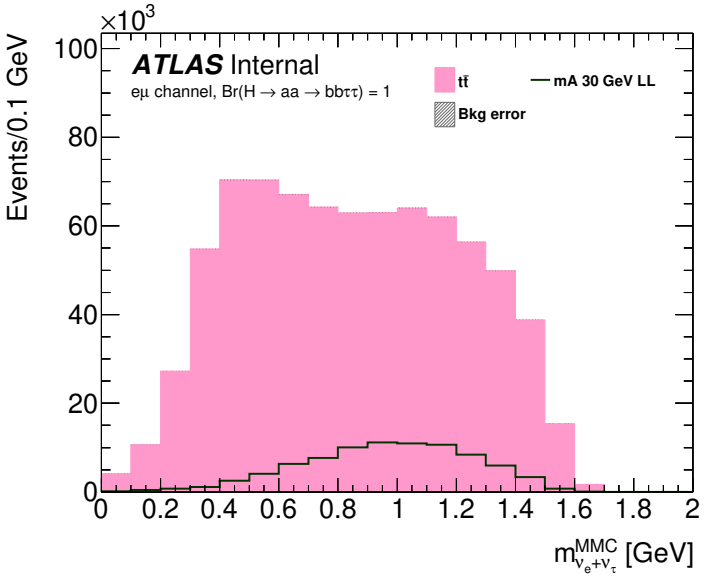


Figure 5.9: Reconstructed mass of the $\nu_1 + \nu_\tau$ system using the MMC for simulated signal samples in the $e\mu$ channel.

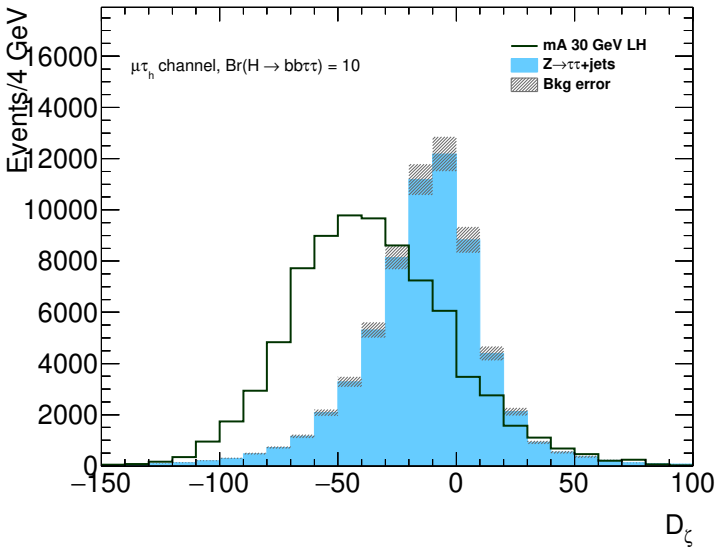


Figure 5.10: D_ζ for $m_a = 30$ GeV simulated signal sample and $Z+$ jets background, showing the discriminating power of this variable.

$$D_\zeta = \vec{E}_{\mathrm{T}}^{\mathrm{miss}} \cdot \vec{\zeta} - 0.85\,(\vec{p_{\mathrm{T}}}(\tau_1) + \vec{p_{\mathrm{T}}}(\tau_2)) \cdot \vec{\zeta} \tag{5.2}$$

where $\vec{\zeta}$ is the angle bisector between the two visible tau decay products. Figure 5.10 shows the D_ζ distribution for signal and $Z+$ jets background. In $Z+$ jets events the taus are more often back-to-back, causing their missing energy to cancel out somewhat, causing the value of D_ζ to be close to zero. In the signal $h \to bb\tau\tau$ process the two taus are on the same side of the four-body decay, so the missing energy is pointed in the same direction as the visible decay products, causing the value of D_ζ to be large and negative.

5.4 Background Estimation

This section details the background estimation strategy. A general overview for the selection of control regions for the major background processes is given in Section 5.4.1. The details related to the estimation and reweighting of the top backgrounds is given in Section 5.4.2, while Section 5.4.3 is focused on the estimation of the backgrounds from Drell-Yan. Section 5.4.4 details the estimation of the data-driven fake lepton background in the $e\mu$

channels, Section 5.4.5 describes the estimation of the background coming from jets faking hadronic taus in the $\ell\tau_{\text{had}}$ channels, and Section 5.4.6 describes the fake lepton background estimation in the $\ell\tau_{\text{had}}$ channels. Further validation plots are found in Appendix A.

5.4.1 Overview

This section discusses the control region definitions and background modeling strategy at a high level. The analysis selects events with heavy flavor, either a DL1r-tagged R=0.4 b-jet or a DeXTer double b-jet, and the decay products of two taus in the $e\mu$, $e\tau_{\text{had}}$, and $\mu\tau_{\text{had}}$ channels. The dominant backgrounds are top production from $t\bar{t}$ and single t, Z$\rightarrow \tau\tau$ "Drell-Yan" produced in association with jets, and fake leptons or hadronic taus. Due to some MC mismodeling observed in the control regions, reweightings are derived for the Drell-Yan and top MC to match the data in their respective CRs. More details on these reweightings are given in Section 5.4.2 for the top backgrounds and Section 5.4.3 for the Drell-Yan backgrounds.

The $t\bar{t}$ CR is defined using the sum of the transverse masses of the visible tau decay products with the missing E_{T} (see Eq. 5.1). The variable Σm_T is chosen because it selects the purest sample of $t\bar{t}$ and rejects both signals and non-$t\bar{t}$ initiated fakes at a high rate. The DY CR is the same as the signal selection, but with the b-tagging requirement inverted: 0 b-jets and 0 DeXTer instead of ≥ 1 b-jet and/or DeXTer jet. An illustration of the CRs considered in all analysis channels is given in Figure 5.2.

Fake leptons in the $e\mu$ final state and jet$\rightarrow \tau_{\text{had}}$ fakes in the $\ell\tau_{\text{had}}$ channel are estimated using data-driven fake-factor methods. Additionally, the modeling of fake leptons and fake hadronic taus is validated in same sign (SS) control regions, which have identical selections to the inclusive SR but the two visible tau decay products are required to have the same charge (SS).

An additional background from Z$\rightarrow \ell\ell$ events shows up as a minor background in the $\ell\tau_{\text{had}}$ channels. These events generally have one lepton correctly identified and the other

lepton fakes a hadronic tau. This results in a visible mass of the tau decay products near the Z mass, unlike Z$\to \tau\tau$ events which have missing mass from the neutrinos. A high-mass sideband in the visible di-tau mass is used to constrain the normalization of this background.

The other minor backgrounds include SM H$\to \tau\tau$, di-boson production, and ttV. These backgrounds are modeled using MC.

5.4.2 Top

5.4.2.1 Top backgrounds in the $\ell\tau_{\text{had}}$ final states

A reweighting of top backgrounds is necessary in the $\ell\tau_{\text{had}}$ final states due to mismodeling observed in the top control region that affects the N_{jets} and H_{T} distributions. This mismodeling is seen in Figures 5.11a–5.11c. The procedure as described is applied identically to the $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ channels. "Top backgrounds" refers to both $t\bar{t}$ and single t events which are modeled in MC, as well as the component of any data-driven modeling which is attributable to these MC samples. Data-driven modeling for fake leptons and taus (described in subsequent sections) involves MC subtraction in dedicated CRs where the fake templates are derived. When the MC reweighting is performed, the negative event weights from MC subtraction are taken into account insofar as they affect the total event count in the CR bins where the reweighting is defined. When the likelihood ratio fit is performed, data-driven templates are not affected by the nuisance parameters for the MC reweighting; rather, the nominal reweighting is used and a separate nuisance parameter is used for MC subtraction uncertainty in the data-driven template.

In order to correct the mismodeling between data and MC, a sequential reweighting technique is used. First, the top MC is reweighted by the number of ($R = 0.4$) b-jets and then by the number of ($R = 0.4$) jets, inclusive of the number of DL1r b-tagged jets. The reweighting by b-jet multiplicity uses bins of 1, 2, and >2 b-jets and the reweighting by inclusive N jets uses bins of 1, 2, 3, and >3 jets. There are a total of 7 nuisance parameters which are allowed to float and constrained by the $t\bar{t}$ CR for the final fit. The $t\bar{t}$ CR is

subdivided into 9 bins in the N b-jet vs N jet (excluding b-jets) plane to constrain these 7 nuisance parameters.

A separate reweighting with respect to $H_T = \Sigma p_{\mathrm{T}}$ after the N b-jet, N jet reweighting is also considered. However, no significant mismodeling with respect to H_T in the $t\bar{t}$ CR is observed, so this additional reweighting is not used.

The performance of the reweighting in the $t\bar{t}$ CR for the $\mu\tau_{\mathrm{had}}$ channel is shown in Figure 5.11 and the weights derived for each bin in the reweighting are detailed in Figure 5.12. The reweighted background model (the third row of Figure 5.11) demonstrates good agreement with the observed data. The largest deviation between MC and data is 30%, when there is a large number of jets but a small number of b-jets. Similar plots for the $e\tau_{\mathrm{had}}$ channel are shown in Figure 5.13 and Figure 5.14. The modeling and values of the weights are in good agreement with the $\mu\tau_{\mathrm{had}}$ channel.

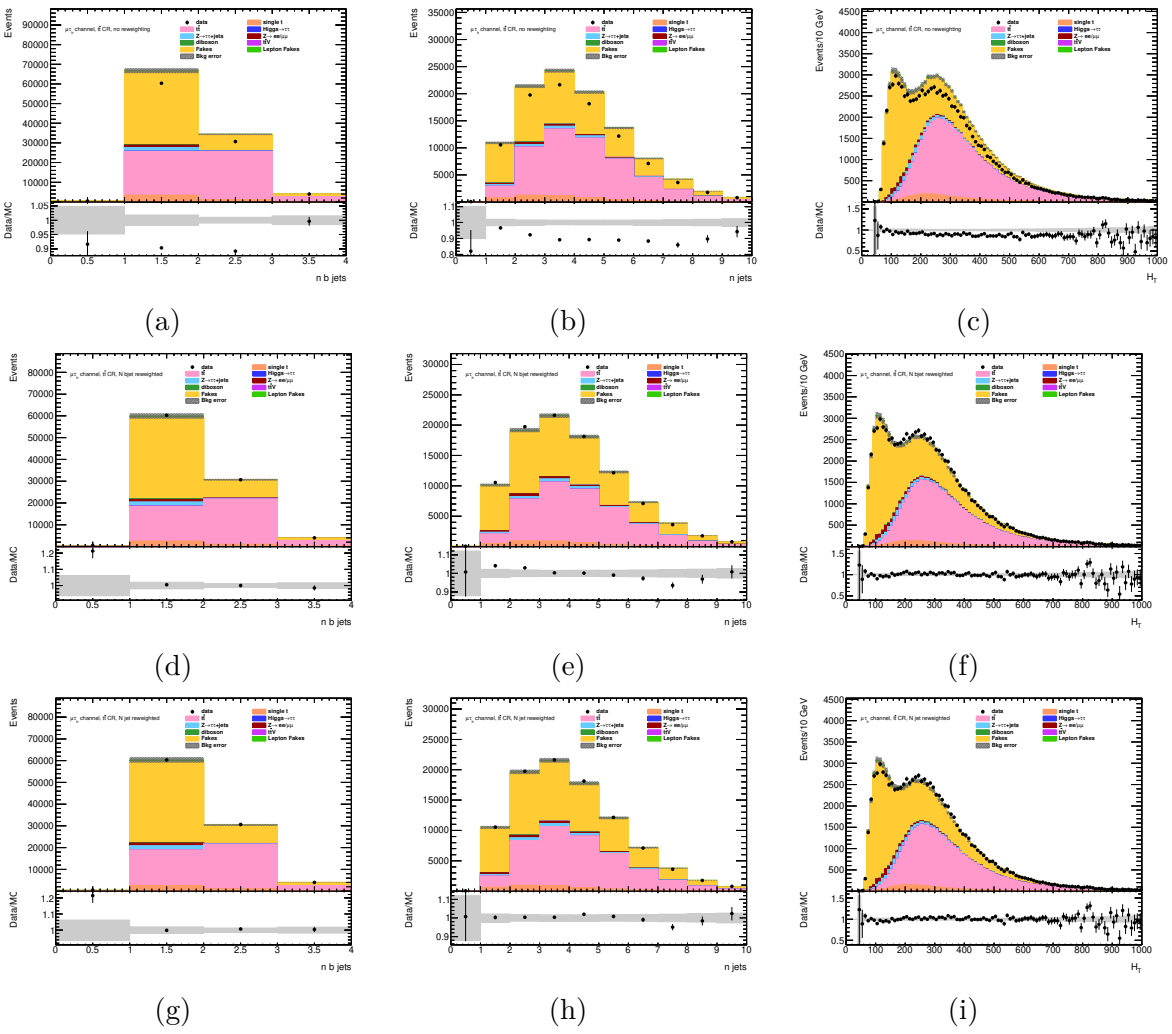


Figure 5.11: Validation plots in the $t\bar{t}$ CR for the $\mu\tau_{\text{had}}$ channel. The N b-jets, N jets inclusive, and H_T distributions are shown before reweighting (first row), with only N b-jet reweighting applied (second row), and with the full sequential reweighting applied (third row).

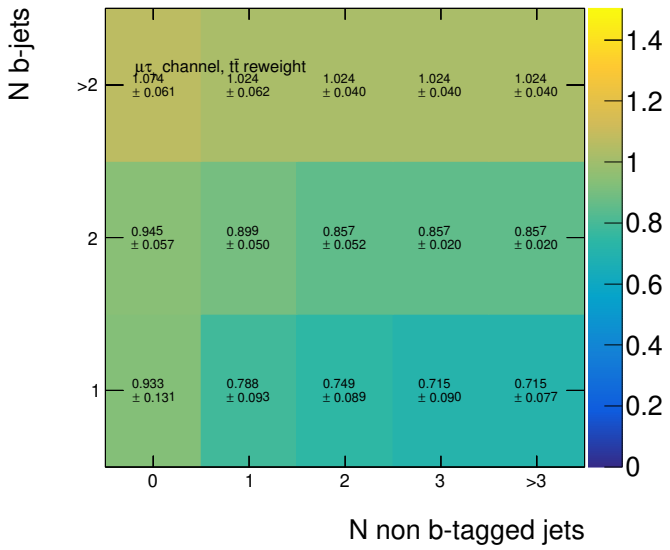


Figure 5.12: The $t\bar{t}$ weights derived during the sequential reweighting in the $\mu\tau_{\text{had}}$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template.

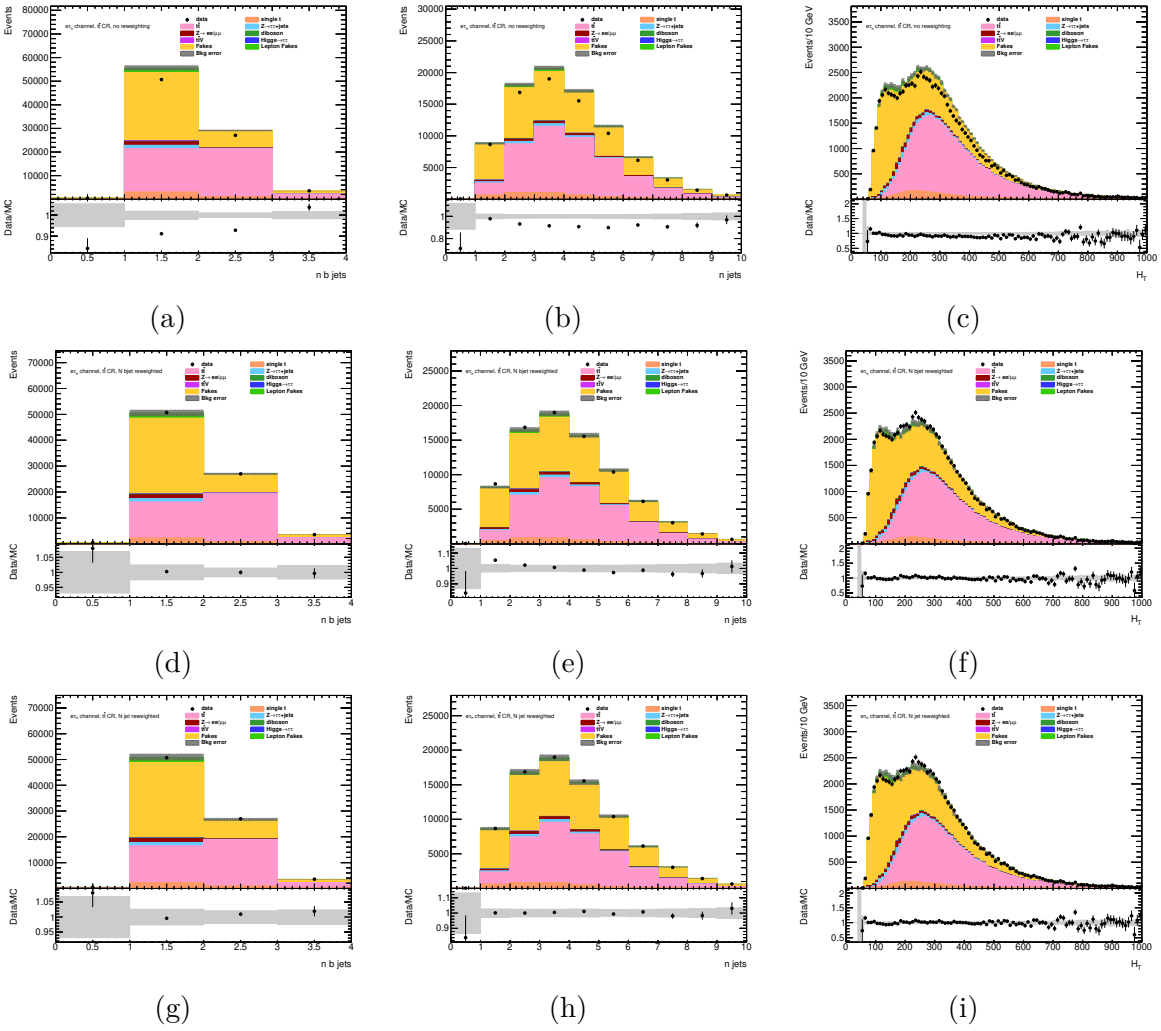


Figure 5.13: Validation plots in the $t\bar{t}$ CR for the $e\tau_{\text{had}}$ channel. N b-jets, N jets inclusive, and H_T are shown before reweighting (first row), with only N b-jet reweighting (second row), and with the full sequential reweighting (third row).

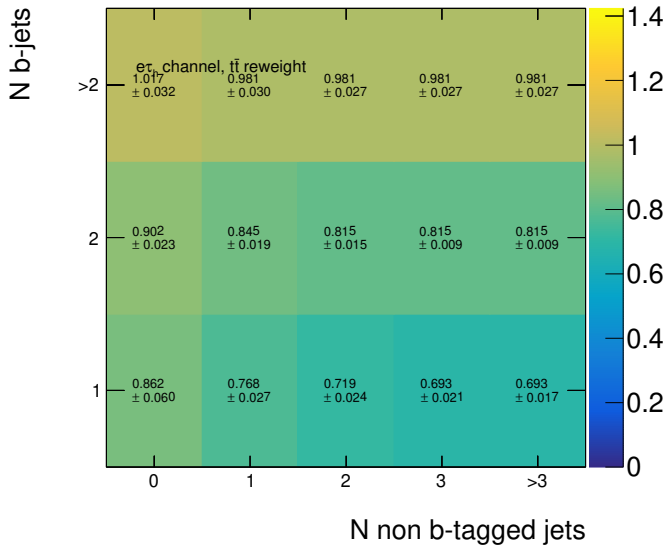


Figure 5.14: The $t\bar{t}$ weights derived during the sequential reweighting in the $e\tau_{\text{had}}$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template.

5.4.2.2 Top backgrounds in the $e\mu$ final state

A nearly identical procedure to the scheme used for the $\ell\tau_{\text{had}}$ final state is applied in the $e\mu$ final state. First, the top MC is reweighted using a sequential reweighting performed in the N b-jets and the N jets distributions, inclusive of the number of b-tagged jets. The observed deviation is far less in this channel, but the sequential reweighting is applied nonetheless, and the nuisance parameters are again left floating in the final fit. The same binning in N b-jets and N jets as the $\ell\tau_{\text{had}}$ channel is used for this reweighting and the jet multiplicity weights are shown in Figure 5.17. After this reweighting, however, a residual mismodeling is still seen in the $H_T = \Sigma p_{\text{T}}$ distribution shown in Figure 5.15, which necessitates an additional reweighting in the $e\mu$ channel,

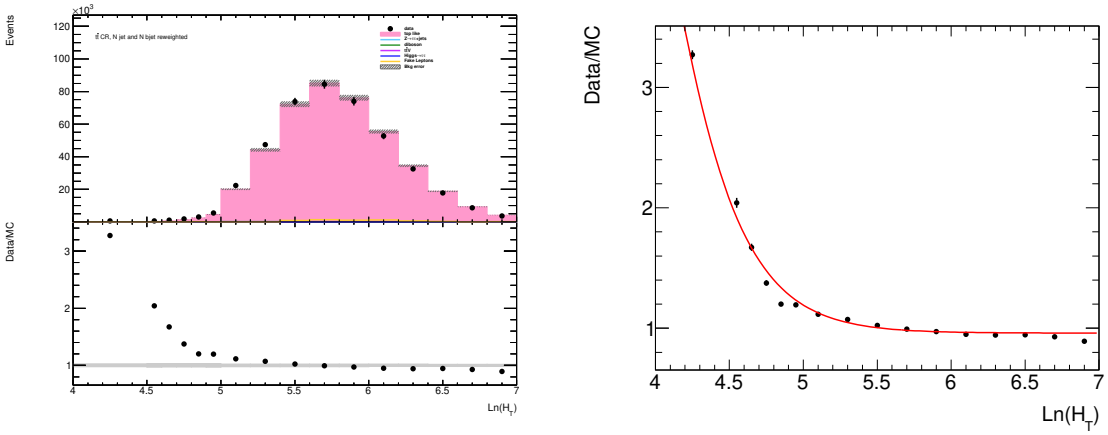
This reweighting is performed by reweighting the Data/MC fraction as a function of H_T and the functional form

$$w(H_T) = \frac{p_0}{e^{p_1(\ln(H_T)-p_2)}+1} + p_3 \tag{5.3}$$

is used for the reweighting. The best fit parameters are for this reweighting are summarized in Table 5.4 and the performance of the reweighting in the $t\bar{t}$ CR is shown in Figure 5.16.

Table 5.4: Best fit parameters to Eq. 5.3 in the $e\mu$ channel. No H_T reweighting is performed in the $\ell\tau_{\text{had}}$ channels.

Parameter	Value
p_0	10.00 ± 0.204
p_1	3.320 ± 0.048
p_2	3.873 ± 0.011
p_3	0.959 ± 0.002



(a) (b)

Figure 5.15: (a) H_T reweighting is necessary in the $e\mu$ channel, as observed in the $\ln(H_T)$ distribution. (b) Reweighting fit as a function of $\ln(H_T)$.

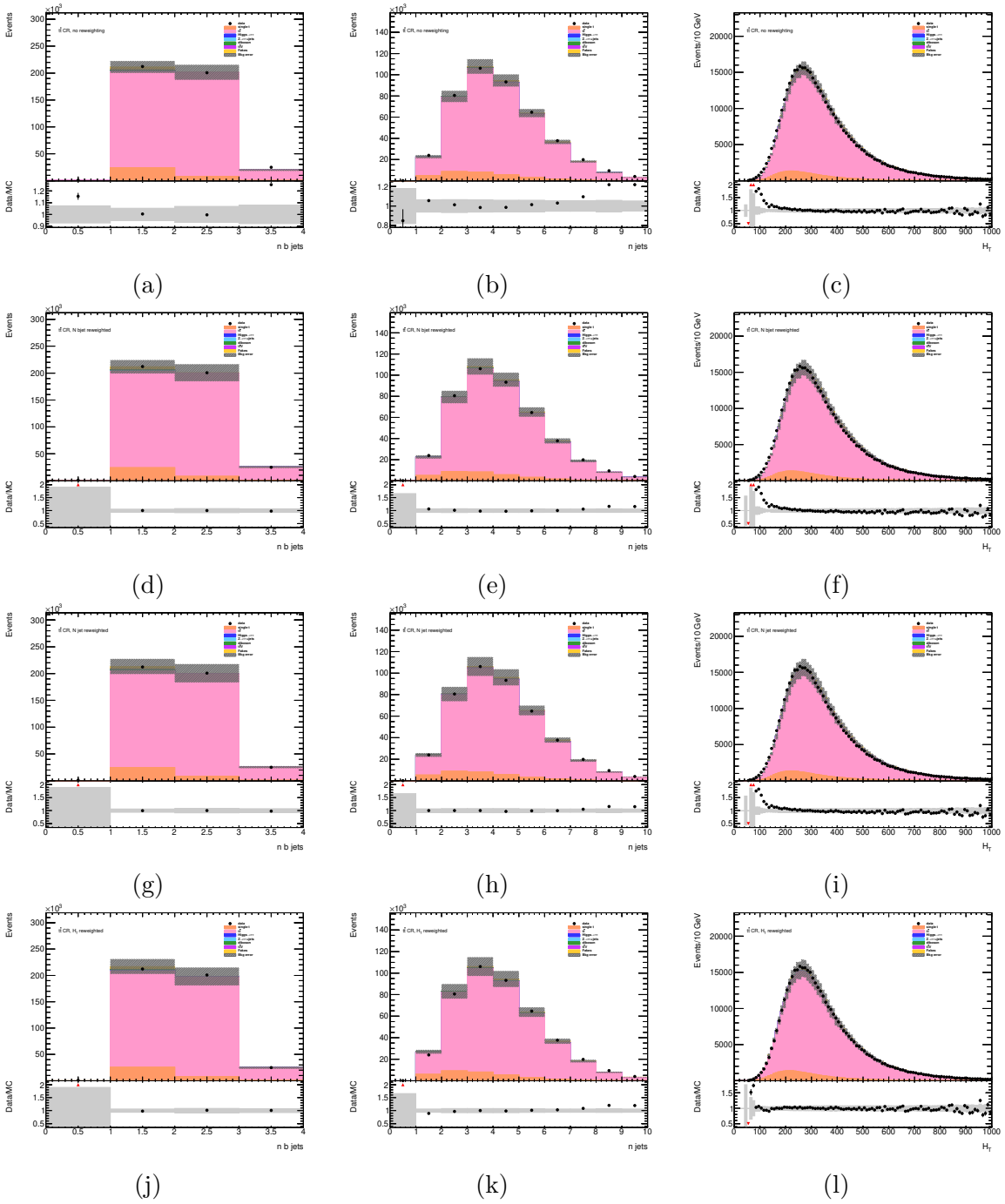


Figure 5.16: Validation plots in the $t\bar{t}$ CR for the $e\mu$ channel. N b-jets, N jets inclusive, and H_T are shown before reweighting (first row), with only N b-jet reweighting (second row), with N b-jet and N jet reweighting (third row), and with the full sequential reweighting, including H_T (fourth row).

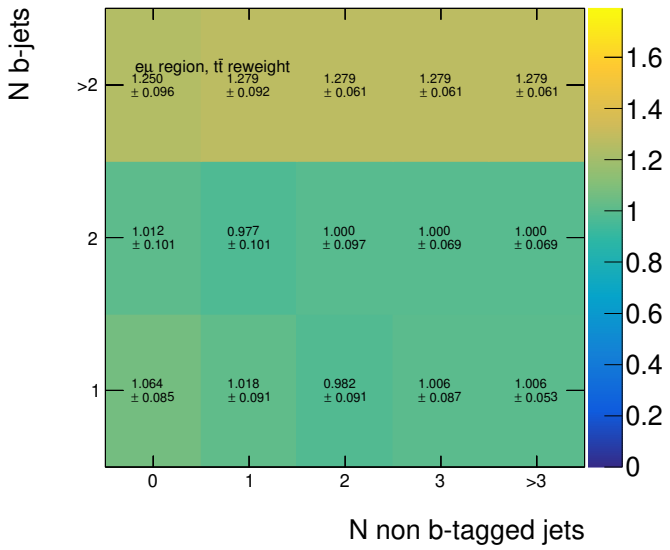


Figure 5.17: The $t\bar{t}$ weights derived during the sequential reweighting in the $e\mu$ channel. The quoted uncertainty is statistical from MC statistics and from the propagation of errors on the fake template.

5.4.3 Drell-Yan (+jets)

We here consider the background from "Drell-Yan" processes ($Z \to \ell\ell$) in association with jets. This mimics the $bb\tau\tau$ signal where $\ell = \tau$ or a $\ell \to \tau$ fake occurs and the additional jet or jets are flavor tagged (regardless of whether they originate from b quarks).

Each channel in this analysis uses a pair of objects which have a different lepton flavor, with at least one lepton and another (OF) lepton or hadronic tau. The dominant Drell-Yan final state is $\tau\tau$+jets production. However, in the $\ell\tau_{\text{had}}$ channels, there is a large enough lepton-to-τ_{had} fake rate that the analysis needs to consider the backgrounds coming from both Drell-Yan production of two taus decaying to $\ell\tau_{\text{had}}$ and DY production of $\ell\ell$ where one lepton is misidentified as a hadronic tau. In the $e\mu$ channel, $\tau\tau$ production is the only final state which has an appreciable number of events because the lepton flavor misidentification rate is sufficiently low.

In all channels a high mass sideband ($m_{\tau\tau}^{\text{vis}} > 60$ GeV for $\ell\tau_{\text{had}}$ and $m_{\tau\tau}^{\text{vis}} > 45$ GeV for $e\mu$, see Figures 5.18–5.20) is used to fit the normalization on the DY MC. The MC can generally model the normalization on the DY process very accurately; however, for the exclusive heavy flavor categories in this analysis (1 and 2 b-jets and DeXTer), the MC must be reweighted. This is expected, as it is similar to the reweighting performed as a function of the number of jets in the $t\bar{t}$ channel. In this section, we show the DeXTer category separately because it is so small, but perform the reweighting inclusively.

In the $\ell\tau_{\text{had}}$ channels, two bins are used in order to simultaneously fit the normalizations of $\tau\tau$ and $\ell\ell$ decays. The best fit normalizations are unconstrained nuisance parameters in the PLR fit.

In the $\ell\tau_{\text{had}}$ channels, the two bins used for the simultaneous fit are defined as:

- Bin 1 with (60 GeV $< m_{\tau\tau}^{\text{vis}} <$ 85 GeV)
- Bin 2 with (85 GeV $< m_{\tau\tau}^{\text{vis}} <$ 105 GeV)

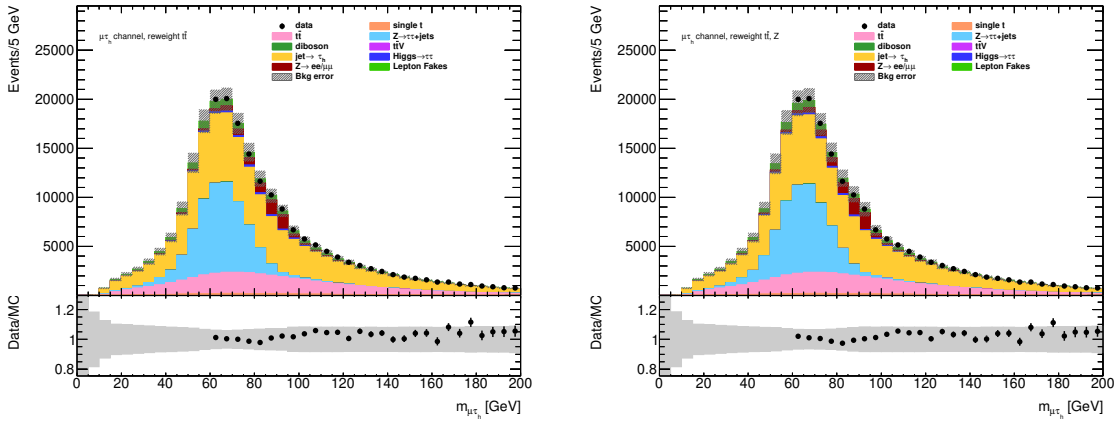
The normalizations of the DY production of $\tau\tau$ and $\ell\ell$ events are fit simultaneously using a

matrix inversion:

$$\begin{pmatrix} w_{\tau\tau} \\ w_{\ell\ell} \end{pmatrix} = \begin{pmatrix} N^1_{\tau\tau} & N^1_{\ell\ell} \\ N^2_{\tau\tau} & N^2_{\ell\ell} \end{pmatrix}^{-1} \begin{pmatrix} N^1_{\text{obs}} \\ N^2_{\text{obs}} \end{pmatrix} \tag{5.4}$$

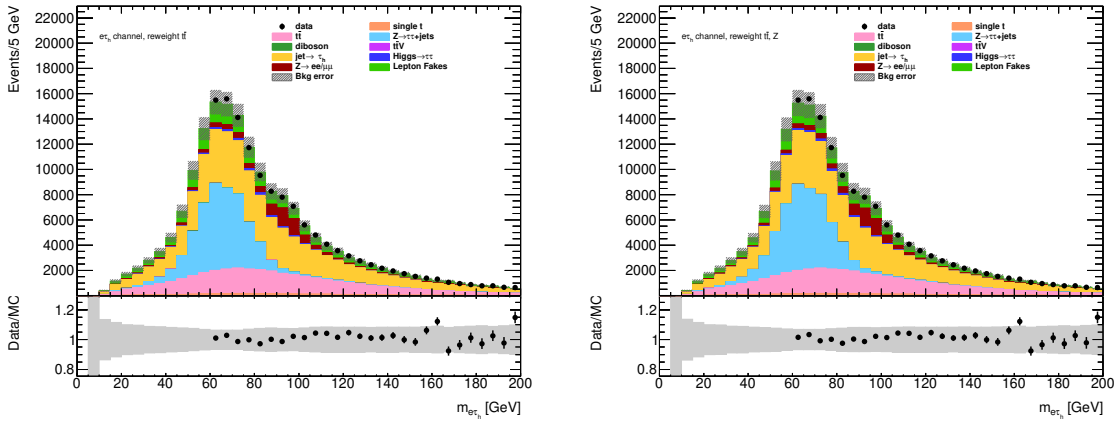
where the superscript (1, 2) denotes the counts in bin 1 or 2. N_{obs} are the observed counts, $N_{\tau\tau}$ and $N_{\ell\ell}$ are the predictions from the MC, and $w_{\tau\tau}$ and $w_{\ell\ell}$ are the derived scale factors. A single bin fit is used in the $e\mu$ channel to derive the SF on the $Z \to \tau\tau$ MC, as there is no $Z \to \ell\ell$ contribution; the bin contains all events with a visible mass between 45 and 65 GeV. The reweighting procedure is performed in parallel for each channel and in both the 0 b-tag VR and the high mass CR. Performance before and after the reweighting is shown in Figures 5.18–5.20. In the $\ell\tau_{\text{had}}$ channels the DY reweighting has almost no affect. The DeXTer category is a small subset of the inclusive heavy flavor region, so the DeXTer region is plotted separately in Figure 5.21. The right hand plot in Figures 5.18–Figure 5.21 shows the agreement between data and the background model for the high visible mass sideband. The gray band in the ratio plot is an error estimate including MC statistics and a 25% uncertainty on the data-driven fake estimate. The data and MC are in good agreement across the mass range, including bins with highly different compositions where several different processes are dominant.

The derived event weights from the different simultaneous fits are summarized in Table 5.5. There is a consistent 20% mismodeling in the 0 b-jet category, but this category is only used for validation and does not factor into the fit (discussed in section 5.7). All categories with heavy flavor have a reweighting factor consistent with unity. Additional plots validating the performance in other kinematic variables are presented in Section A.



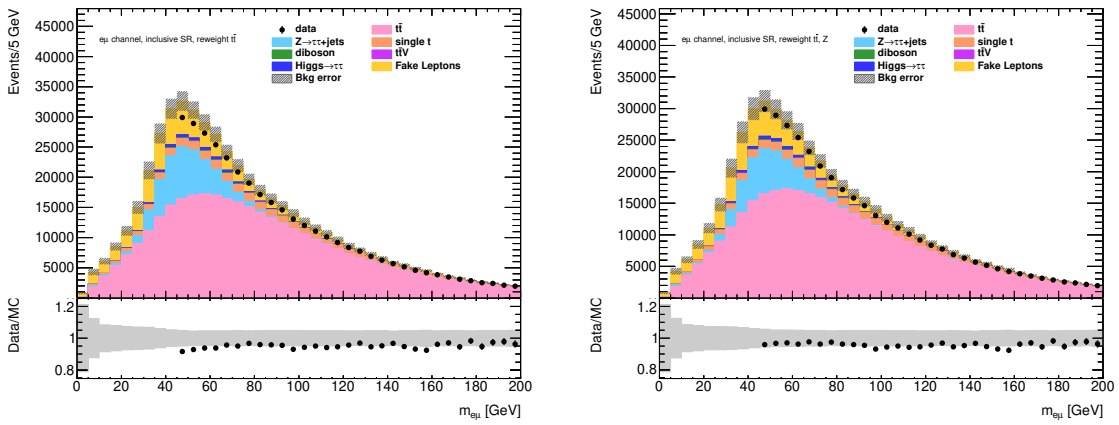
(a) (b)

Figure 5.18: Validation plots in the high mass CR for the $\mu\tau_{\text{had}}$ channel (a) before and (b) after DY reweighting.



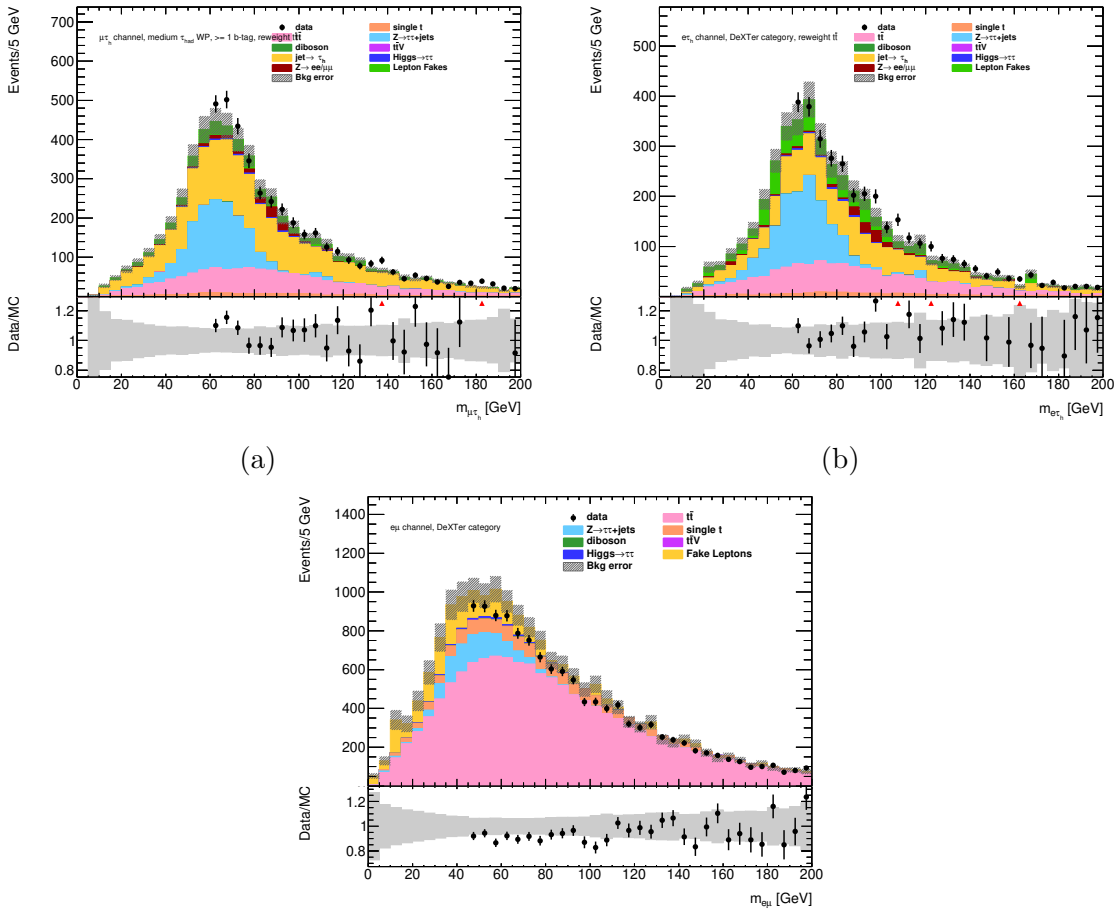
(a) (b)

Figure 5.19: Validation plots in the high mass CR for the $e\tau_{\text{had}}$ channel (a) before and (b) after DY reweighting.



(a) (b)

Figure 5.20: Validation plots in the high mass CR for the $e\mu$ channel (a) before and (b) after DY reweighting.



(a) (b)

(c)

Figure 5.21: Validation plots in the high mass CR with the DeXTer selection applied for the (a) $\mu\tau_{\text{had}}$ (b) $e\tau_{\text{had}}$ and (c) $e\mu$ channels.

Table 5.5: Summary of nominal unconstrained weights derived for DY in all channels and selections. The nominal weights in this table are applied to the DY MC, including the component of the DY MC which leaks into the data-driven CRs for construction fake lepton and hadronic tau templates. The scale factors in the > 0 b-tag selection are unconstrained normalizations in the PLR fit. The quoted error is statistical and from the propagation of errors on the data-driven fake template.

Process	Channel	Selection	Scale Factor
$Z/\gamma^* \to \tau\tau$	$\mu\tau_{\text{had}}$	0 b-tag	0.805 ± 0.047
		> 0 b-tag	0.978 ± 0.083
	$e\tau_{\text{had}}$	0 b-tag	0.883 ± 0.061
		> 0 b-tag	0.990 ± 0.087
	$e\mu$	0 b-tag	1.060 ± 0.081
		> 0 b-tag	1.026 ± 0.170
$Z/\gamma^* \to \ell\ell$	$\mu\tau_{\text{had}}$	0 b-tag	0.819 ± 0.173
		> 0 b-tag	1.110 ± 0.360
	$e\tau_{\text{had}}$	0 b-tag	0.786 ± 0.174
		> 0 b-tag	1.001 ± 0.271

5.4.4 Fake leptons in the $e\mu$ final state

In the $e\mu$ channel, one or both of the leptons considered can be fake or non-prompt. Fake events are mostly the result of $W \to \ell$+jets and QCD multijet processes, where one or more jets are misidentified as a lepton.

The Matrix Method (MM) is used to simultaneously model fake events with different fake multiplicities (one or two of the leptons is fake) with the same fake rate on the individual electrons or muons. The core assumption of the MM is that orthogonal tight (T) and loose (L) selections for the leptons are defined and that fake and real leptons will pass the tight and loose selections at different rates (or else the matrix, described below, is singular). The tight selection is the "Signal" selection described in Table 4.1 and the loose selection is the "Baseline" selection, but with a veto applied for any lepton passing the "Signal" selection. This selection enforces that the tight and loose regions as described are orthogonal, while "Signal" and "Baseline" are not, motivating the redefinition.

The lepton efficiencies are given by:

$$\epsilon_a^\ell = \frac{N_T}{N_T + N_L} \tag{5.5}$$

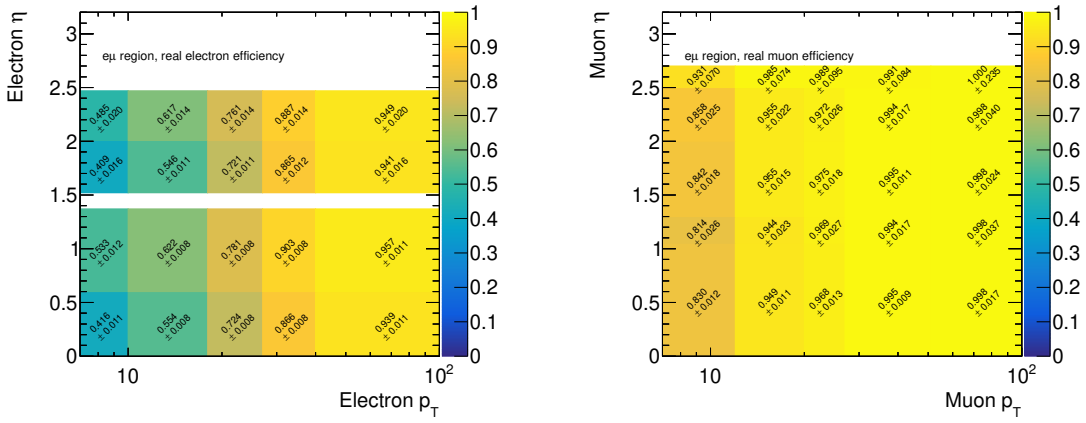
with superscripts $\ell \in (e, \mu)$ for lepton flavor and subscripts $a \in (r, f)$ for real and fake efficiencies, respectively. N_T and N_L are the number of events passing the tight and loose selections. The notation $\bar{\epsilon} = 1 - \epsilon$ is also used.

Efficiencies for real and fake leptons are derived using $Z \to \tau\tau$ MC samples and a data-driven tag-and-probe method, respectively. The real lepton efficiency is calculated using Eq. 5.5 with the $Z \to \tau\tau$ MC sample. To calculate the real electron efficiency, the muon is required to meet the signal selection; similarly, the electron is required to pass the signal selection and to calculate the real muon efficiency. Only events with two truth-matched OS leptons are considered. The leptons are also required to have a separation of $\Delta R \geq 1.4$, a cut to veto the region where the lepton isolation cones overlap and the two isolation discriminants are correlated. The real lepton efficiencies are binned in lepton $|\eta|$ and p_{T} using the binning described in Table 5.6. The observed lepton efficiencies are shown in Figure 5.22 separately for the electrons and muons. The efficiency for muons is 80-100%, rising with higher p_{T}, and the efficiency for electrons similarly rises from 40-100% with p_{T}. We use a tighter isolation WP for electrons, causing the lower efficiency at low p_{T}.

Fake lepton efficiencies are calculated in a sample of data with two leptons using a tag-and-probe method. One lepton is "tagged", placed under high isolation requirements, and the efficiency of the other is "probed." The fake lepton efficiency is estimated in the SS region using the same $|\eta|$ and p_{T} bins described in Table 5.6, but with additional bins in the trigger region and number of b-jets. The same cut of $\Delta R \geq 1.4$ is applied to veto the region where lepton isolation discriminants may be correlated. The observed fake lepton efficiencies for the heavy-flavor region are summarized in Figures 5.23–5.24. The efficiency for fake leptons is generally low, except when the fake lepton is triggering a single lepton

Table 5.6: Lepton efficiency bins. Bins in p_{T} and $|\eta|$ are used for real and fake leptons, while the trigger region and N b-jet bins are only used for fake leptons.

Flavor	Variable	Bin Edges	Comments
Electron	$\|\eta\|$	0.6, 1.37, 1.52, 2.01, 2.47	Electrons in the crack ($1.37 < \|\eta\| < 1.52$) are rejected
	p_{T}	10, 18, 27, 40 GeV	—
	Trigger	1, 2, 3	Only for fake electrons
	N b-tag	0, > 0	Only for fake electrons
Muon	$\|\eta\|$	1.05, 1.3, 2.0, 2.5, 2.7	—
	p_{T}	10, 15, 20, 25, 27, 51 GeV	—
	Trigger	1, 2, 3	Only for fake muons
	N b-tag	0, > 0	Only for fake muons



(a) (b)

Figure 5.22: Observed real (a) electron and (b) muon efficiencies which are obtained using $Z \to \tau\tau$ MC. These efficiencies are used as inputs to the matrix method to calculate the lepton fake rates in all three analysis channels.

trigger, in which case it can pass isolation at a relatively high rate, although still below the real lepton isolation efficiency. Corresponding plots for the 0 b-tag VR are shown in Appendix A (Figures A.1–A.2).

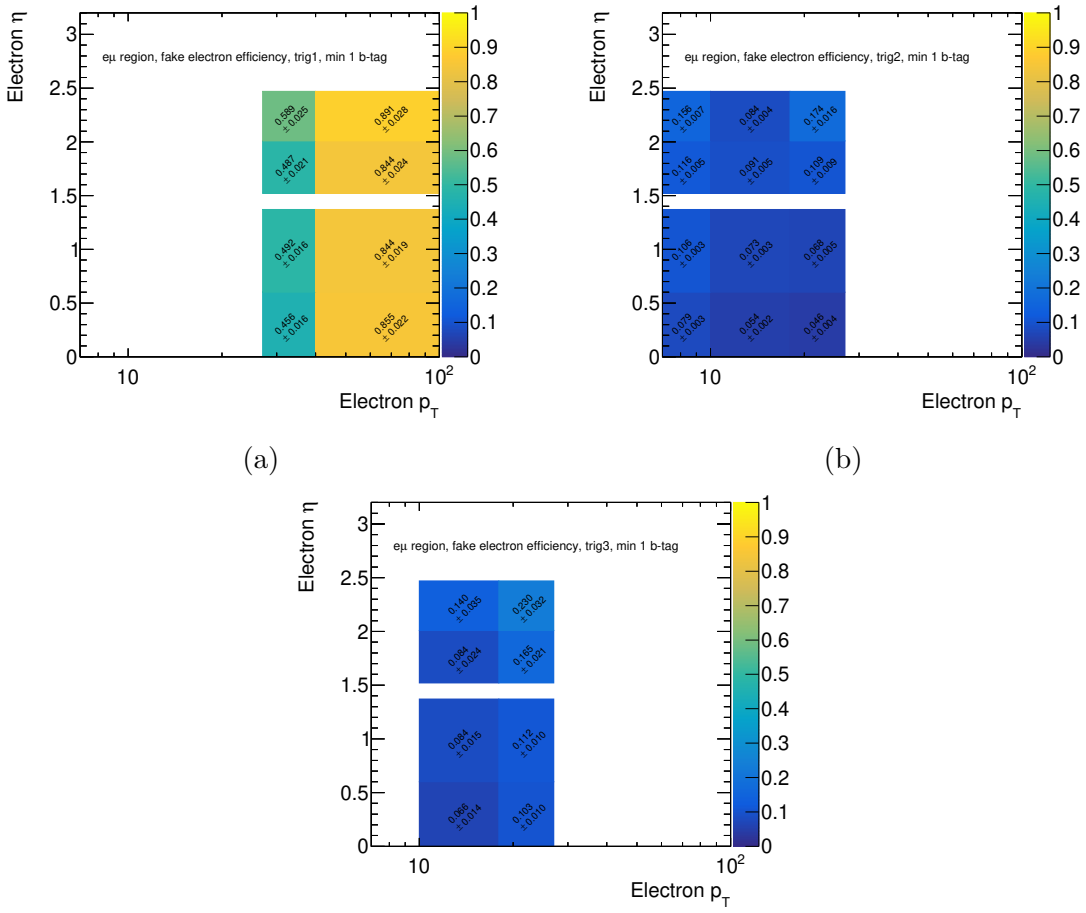


Figure 5.23: The observed fake electron efficiencies in the region containing at least 1 *b*-jet. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3.

The MM sets up a linear problem to parameterize the observed counts in terms of truth-level counts:

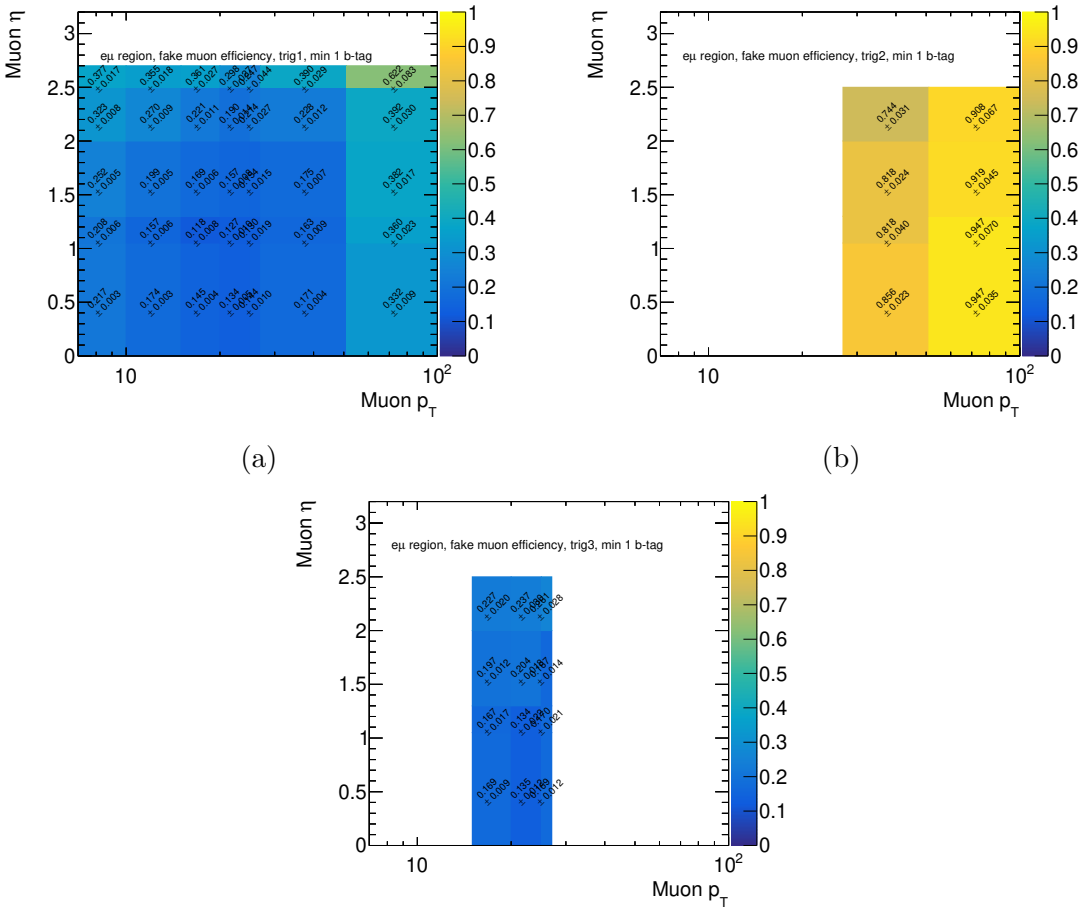


Figure 5.24: The observed fake muon efficiencies in the region containing at least 1 *b*-jet. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3.

$$\begin{bmatrix} N_{TT} \\ N_{TL} \\ N_{LT} \\ N_{LL} \end{bmatrix} = \begin{bmatrix} \epsilon_r^e \epsilon_r^\mu & \epsilon_r^e \epsilon_f^\mu & \epsilon_f^e \epsilon_r^\mu & \epsilon_f^e \epsilon_f^\mu \\ \epsilon_r^e \overline{\epsilon_r^\mu} & \epsilon_r^e \overline{\epsilon_f^\mu} & \epsilon_f^e \overline{\epsilon_r^\mu} & \epsilon_f^e \overline{\epsilon_f^\mu} \\ \overline{\epsilon_r^e} \epsilon_r^\mu & \overline{\epsilon_r^e} \epsilon_f^\mu & \overline{\epsilon_f^e} \epsilon_r^\mu & \overline{\epsilon_f^e} \epsilon_f^\mu \\ \overline{\epsilon_r^e} \overline{\epsilon_r^\mu} & \overline{\epsilon_r^e} \overline{\epsilon_f^\mu} & \overline{\epsilon_f^e} \overline{\epsilon_r^\mu} & \overline{\epsilon_f^e} \overline{\epsilon_f^\mu} \end{bmatrix} \begin{bmatrix} N_{rr} \\ N_{rf} \\ N_{fr} \\ N_{ff} \end{bmatrix} \tag{5.6}$$

whenever double subscripts are used, as in TL or fr, the first always refers to the electron and the second the muon. $\epsilon_r^e \epsilon_r^\mu N_{rr}$, $\epsilon_r^e \overline{\epsilon_r^\mu} N_{rr}$, $\overline{\epsilon_r^e} \epsilon_r^\mu N_{rr}$ and $\overline{\epsilon_r^e} \overline{\epsilon_r^\mu} N_{rr}$ are the counts in the TT, TL, LT and LL regions due to processes with two real leptons, modeled with MC. The number of fakes in the TT region is estimated by inverting the matrix and solving for $N_{TT} - \epsilon_r^e \epsilon_r^\mu N_{rr}$. The weights required on the (MC subtracted) event counts in each of the 4 regions are:

$$w_{TT} = \frac{\epsilon_f^e \epsilon_f^\mu - \epsilon_r^e \epsilon_f^\mu - \epsilon_f^e \epsilon_r^\mu + \epsilon_r^e \epsilon_r^\mu \epsilon_f^\mu + \epsilon_r^e \epsilon_r^\mu \epsilon_f^e - \epsilon_r^e \epsilon_r^\mu \epsilon_f^e \epsilon_f^\mu}{\left(\epsilon_r^e - \epsilon_f^e\right)\left(\epsilon_r^\mu - \epsilon_f^\mu\right)} \tag{5.7}$$

$$w_{TL} = \frac{\epsilon_r^e \epsilon_r^\mu \overline{\epsilon_f^e} \epsilon_f^\mu}{\left(\epsilon_r^e - \epsilon_f^e\right)\left(\epsilon_r^\mu - \epsilon_f^\mu\right)} \tag{5.8}$$

$$w_{LT} = \frac{\epsilon_r^e \epsilon_r^\mu \epsilon_f^e \overline{\epsilon_f^\mu}}{\left(\epsilon_r^e - \epsilon_f^e\right)\left(\epsilon_r^\mu - \epsilon_f^\mu\right)} \tag{5.9}$$

$$w_{LL} = \frac{-\epsilon_r^e \epsilon_r^\mu \epsilon_f^e \epsilon_f^\mu}{\left(\epsilon_r^e - \epsilon_f^e\right)\left(\epsilon_r^\mu - \epsilon_f^\mu\right)} \tag{5.10}$$

Note that when $\epsilon_r^e = \epsilon_r^\mu = 1$ the weight $w_{TT} = 0$, and the fake template is comprised entirely of the LT, TL, and LL regions. The nuisance parameters used to estimate the uncertainty on the fake template are discussed further in Section 5.6.2.1.

When the electron and muon isolation discriminants are correlated, the matrix method is no longer valid. The isolation discriminants are correlated when $\Delta R < 0.6$ and the isolation cones of the two leptons overlap. The leptons themselves are excluded from the isolation calculation, but nearby tracks not reconstructed as leptons are eligible to be counted in both isolation cones. A correction to the matrix method is derived which weights events in the

LL region within $\Delta R < 0.6$ of each other as an admixture of LL, TL, and LT. Specifically, the LL weight in the $\Delta R < 0.6$ region is modified as:

$$w_{LL}^{\text{corr}} = (1 - f(\Delta R))\, w_{LL} + f(\Delta R) \times \frac{1}{2}\,(w_{LT} + w_{TL}) \tag{5.11}$$

$$f(\Delta R) = c \times \left(\frac{2}{\pi} \arccos\left(\frac{\Delta R}{2r}\right) - \frac{\Delta R}{\pi r^2}\sqrt{r^2 - \frac{\Delta R^2}{4}} \right) \tag{5.12}$$

where $f(\Delta R)$ is the fraction of the isolation cone which overlaps, c is an arbitrary constant fit from the data in the SS CR, and $r = 0.3$ is the radius of the isolation cone.

Figure 5.25a shows the mass spectrum prior to this correction of the MM, which shows a large negative excess in the template at low mass due to the negative w_{LL}, and Figure 5.25b shows the performance of the correction. With the correction applied the modeling of the invariant mass is in good agreement with the data, up to a normalization factor of about 10-15%, which is addressed by the systematic uncertainties discussed in Section 5.6.2.1.

The MM is validated in the 0 b-jet VR and SS CR. For more information, see Appendix A.

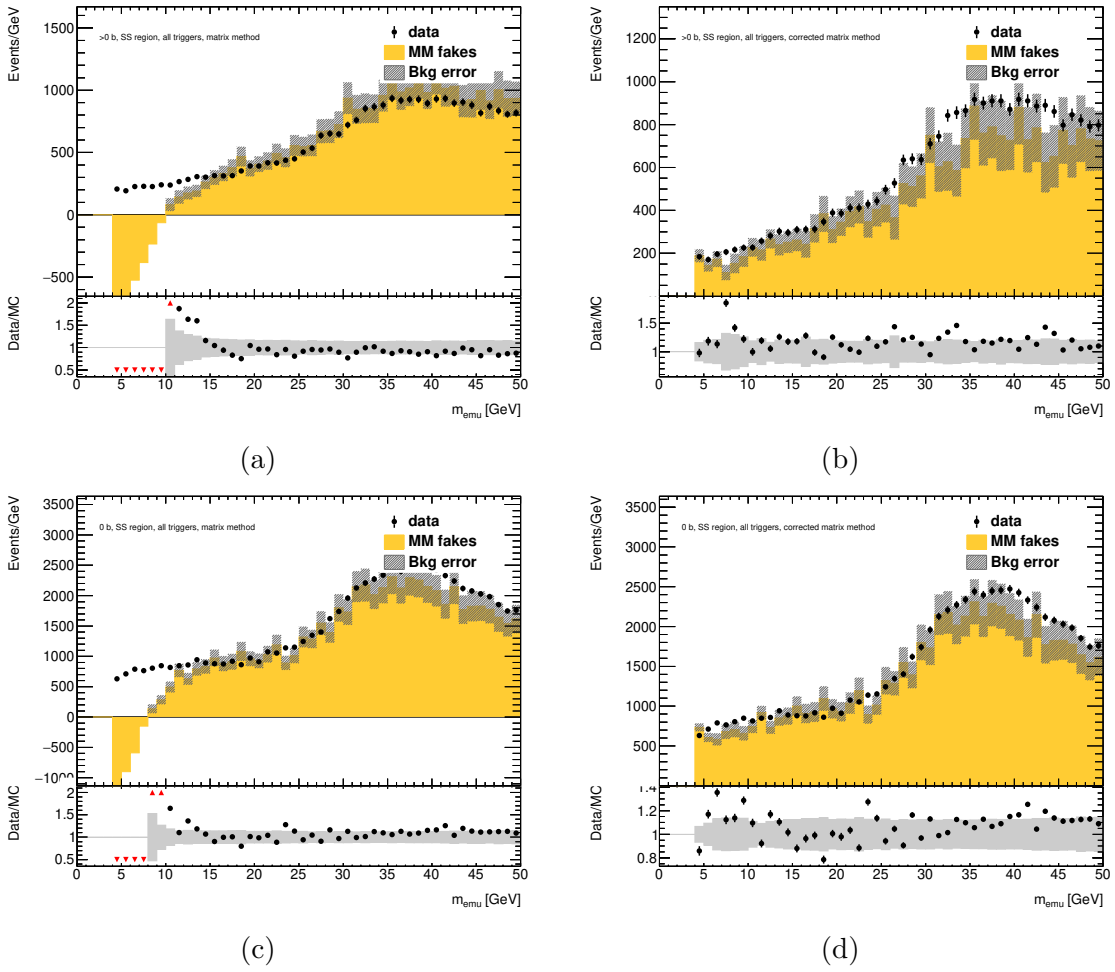


Figure 5.25: The invariant mass of the $e\mu$ system (a) before and (b) after applying the matrix method correction in the SS CR. The SS CR with 0 b-tags is also shown (c) before and (d) after applying the correction. The correction (see Eq. 5.11) adjusts the weights for events with overlapping isolation cones and two "Loose" leptons proportional to the fraction of the isolation cones which are overlapping. The error bars on the background prediction include both statistical errors and the propagation of errors on the matrix method weights.

5.4.5 Jet$\rightarrow \tau_{\text{had}}$ fakes in the $\ell\tau_{\text{had}}$ final states

Fake τ_{had} candidates in the $\ell\tau_{\text{had}}$ final states are primarily the result of $W \rightarrow \ell + \text{jets}$, $t\bar{t}$ and QCD multijet backgrounds. The Fake Factors (FFs) for jet$\rightarrow \tau_{\text{had}}$ fakes are provided by ATLAS collaborators. The FFs are binned in p_{T}, $|\eta|$, and decay mode as shown in Table 5.7.

The FFs are derived in $Z \rightarrow \mu\mu$ +τ_{had}, dijet back-to-back high JVT (HJVT) and dijet back-to-back low JVT (LJVT) event topologies. The two dijet regions are separated by a cut of 0.8 on the JVT of the seed jet of the τ_{had} candidate. The different selections provide different admixtures of light quark-, gluon-, and heavy flavor-initiated jets.

In this analysis a combined FF method is used which weights the three FFs to observations in a CR similar to the signal or validation region in question. The equation for the combined FF is:

$$FF_{\text{combined}} = \left(FF_{Z\rightarrow\mu\mu}(1 - r_{\text{QCD}}) + FF_{\text{dijet, HJVT}} r_{\text{QCD}}\right)(1 - f_{\text{LJVT}}) + FF_{\text{dijet, LJVT}} f_{\text{LJVT}} \tag{5.13}$$

where $FF_{Z\rightarrow\mu\mu}$, $FF_{\text{dijet, HJVT}}$, and $FF_{\text{dijet, LJVT}}$ are the FFs estimated in the $Z \rightarrow \mu\mu$ +τ_{had}, dijet HJVT, and dijet LJVT regions, respectively; additionally, r_{QCD} and f_{LJVT} are weights derived in dedicated CRs for this analysis, which are defined below.

The fake template is constructed using events with a nearly identical selection to the signal region, but with the τ_{had} identification inverted. To construct the fake template, events with baseline taus (RNN score > 0.01) that fail the MEDIUM RNN identification WP are weighted by the combined FFs. The kinematics of baseline taus are therefore extrapolated to the signal region. MC events in which the τ_{had} candidate are matched to hadronic tau at truth-level are subtracted off in the template construction.

The number of fake events in the region of interest (ROI $\in$\{SR, 0 b-tag VR, or SS CR\}) is

$$N_{ROI}^{\text{fakes predicted}} = FF_{\text{combined}} \times \left(N_{\text{anti-}\tau_{\text{had}}}^{\text{data}} - N_{\text{anti-}\tau_{\text{had}}}^{\text{MC,truth } \tau_{\text{had}}}\right) \tag{5.14}$$

where $N_{\text{anti-}\tau_{\text{had}}}^{\text{data}}$ is the number of data events in the anti-τ_{had} region and $N_{\text{anti-}\tau_{\text{had}}}^{\text{MC,truth } \tau_{\text{had}}}$ is the

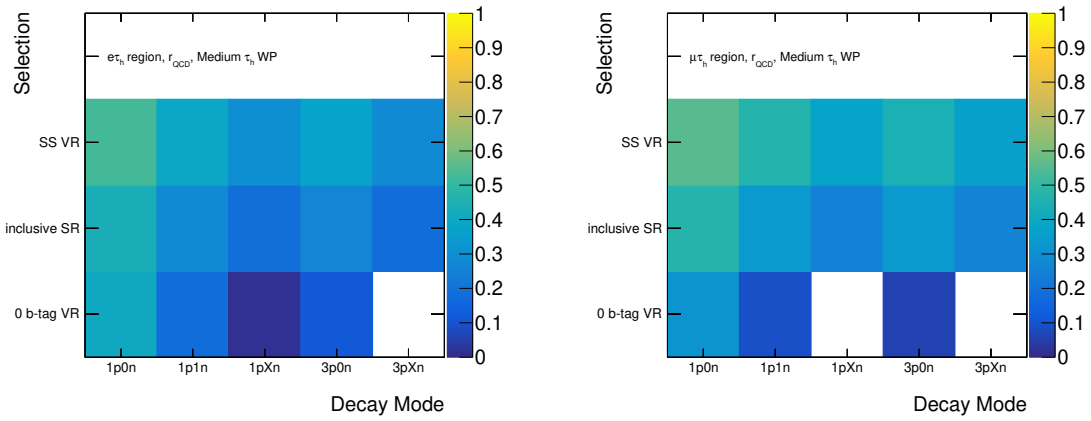
Table 5.7: Bin edges used for jet→ τ_{had} fakes in the $\ell\tau_{\text{had}}$ channels.

Variable	Bin Edges	Comments
Decay Mode	1p0n, 1p1n, 1pXn, 3p0n, 3pXn	First number is N prongs, second is N neutrals
τ_{had} $\lvert\eta\rvert$	1, 1.37, 1.52	Bin (1.37, 1.52) is defined to have FF=0
τ_{had} p_{T} (1-prong)	25, 30, 35, 40, 45, 50 60, 70, 90, 150, 300	FF=0 if $p_{\text{T}} > 300$ GeV
τ_{had} p_{T} (3-prong)	25, 30, 40, 50, 70, 150, 300	FF=0 if $p_{\text{T}} > 300$ GeV

MC prediction with a truth-matched τ_{had} candidate, which is subtracted from the number of data events. The nominal reweighting for $t\bar{t}$ and DY, summarized in Figures 5.12, 5.14 and 5.17 for $t\bar{t}$ and Table 5.5 for DY, is used when estimating $N_{\text{anti-}\tau_{\text{had}}}^{\text{MC,truth }\tau_{\text{had}}}$. Section 5.6.2.3 discusses the nuisance parameters used to capture uncertainty on r_{QCD}, f_{LJVT}, the FFs, and MC subtraction.

Fake estimates using the combined FF methodology are performed in parallel in the inclusive SR, SS CR (SR+high mass CR with sign flipped), and 0 b-tag VR shown in Figure 5.2. Additionally, a fake estimate is performed in the $t\bar{t}$ CR, but it assumes $r_{\text{QCD}} = 0$ and $f_{\text{LJVT}} = 0$. This assumption is validated in Figures 5.11c and 5.13c. These H_T distributions show a smaller low H_T peak and a larger, broader population at higher H_T. The small peak at low H_T is the contribution from W + jets and the broader part of the distribution is the contribution from $t\bar{t}$ (real $t\bar{t}$ in pink, fake taus in the $t\bar{t}$ event topology in yellow). As shown in Section 5.4.2, the $t\bar{t}$ component must be reweighted; however, the W + jets contribution is estimated entirely through the tau FFs, so does not require any reweighting. The choice of $r_{\text{QCD}} = 0$ and $f_{\text{LJVT}} = 0$ correctly models the peak at low H_T, but is not expected to generalize outside the $t\bar{t}$ CR as the high transverse mass cut that defines the $t\bar{t}$ CR biases this region away from QCD multijet fakes, motivating $r_{\text{QCD}} = f_{\text{LJVT}} = 0$.

The two fractions r_{QCD} and f_{LJVT} are used to weight the FFs. r_{QCD} is defined as:



(a) (b)

Figure 5.26: The best fit values of r_{QCD} in the (a) $e\tau_{\text{had}}$ and (b) $\mu\tau_{\text{had}}$ channel, binned by selection and decay mode. The bins with r_{QCD} =0 are shown in white.

$$r_{\text{QCD}} = \frac{N^{\text{data}}_{\text{anti-}\tau_{\text{had}}} - N^{\text{MC,truth }\ell}_{\text{anti-}\tau_{\text{had}}}}{N^{\text{data}}_{\text{anti-}\tau_{\text{had}}} - N^{\text{MC,truth }\ell\tau_{\text{had}}}_{\text{anti-}\tau_{\text{had}}}} \tag{5.15}$$

where $N^{\text{data}}_{\text{anti-}\tau_{\text{had}}}$ is the number of data events with the anti-τ_{had} selection, $N^{\text{MC,truth }\ell}_{\text{anti-}\tau_{\text{had}}}$ is the weighted MC prediction for events in which the lepton of interest is truth-matched to a weak boson, $t\bar{t}$, or other hard-scattering process, and $N^{\text{MC,truth }\ell\tau_{\text{had}}}_{\text{anti-}\tau_{\text{had}}}$ is only the subset of MC events in which both the lepton and τ_{had} candidates of interest are truth-matched. r_{QCD} is estimated for the inclusive signal region, SS CR, and 0 b-jet VR for each $\ell\tau_{\text{had}}$ channel, and is binned by the tau decay mode. Figure 5.26 show the values of r_{QCD} in the $\ell\tau_{\text{had}}$ channels. The values in the $e\tau_{\text{had}}$ and $\mu\tau_{\text{had}}$ channels are compatible with one anther, and as expected the SS region has a higher QCD fraction than the OS region.

f_{LJVT} is derived using a fit to the seed jet width distribution in the anti-tau region. The LJVT dijet sample is gluon-enriched compared to the HJVT sample. As gluon-initiated jets tend to be wider than quark-initiated jets, the seed jet width distribution is indicative of the gluon fraction in the analysis selections.

In addition to the FFs, kinematic distributions are derived for the samples used in the FFs. The kinematic distributions include p_{T}, η, and width of the τ_{had} candidate's seed jet. The

distributions are binned by the tau decay mode and pileup ($< \mu >> 40; < \mu > \leq 40$). The dijet LJVT and $Z \to \mu\mu$ seed jet width distributions are weighted to the p_{T}, η, and $< \mu >$ distributions observed in the anti-τ_{had} region for each channel ($e\tau_{\mathrm{had}}$, $\mu\tau_{\mathrm{had}}$), selection (0 b-tag VR, SS CR, SR), and decay mode (1p0n, 1p1n, 1pXn, 3p0n, 3pXn). Each observed seed jet distribution is then fit as an admixture of the weighted LJVT and $Z \to \mu\mu$ distributions. Only the LJVT and $Z \to \mu\mu$ distributions are used, because the HJVT and $Z \to \mu\mu$ jet width distributions are very similar, so the fit is ill-conditioned if all three samples are used. The estimate of r_{QCD} provides the relative fraction of $Z \to \mu\mu$ and HJVT, and seed jet width is not sensitive to r_{QCD}.

Each fit is performed over the domain [0.02, 0.25] in the seed jet width distribution to focus on the bulk and not the tails. The low and high tails of the seed jet width distribution are not well described by the template shapes, so the fits converge very poorly without the restriction to [0.02, 0.25] (especially in the 1p1n and 1pXn decay modes where f_{LJVT} is high). Figure 5.27 shows the seed jet width fit for the 5 decay modes in the $\mu\tau_{\mathrm{had}}$ channel. Appendix A (Figures A.3–A.7) shows the seed jet width fits performed for other channels and with the same-sign selection applied. Figure 5.28 summarizes the best fit values of f_{LJVT}. The gluon fraction is essentially zero in the 3-prong decay mode and can be near 100% in the 1-prong decay mode.

The combined FF method is validated in the 0 b-tagging VR and the SS CR. Figures A.11b and A.11a show the performance of the fake template in modeling the resonant variable of interest, including in the low mass regime, in the 0 b-tagging VR (after DY reweighting). Figures 5.29 and 5.31 show the SS CR. The largest mismodeling observed is the 10% normalization mismodeling in the $\mu\tau_{\mathrm{had}}$ channel, SS CR. This mismodeling is covered by the systematic uncertainties discussed in Section 5.6.2.3. No significant shape discrepancy between the model and data is observed.

Figure 5.32 shows the combined FF for the $\mu\tau_{\mathrm{had}}$ SR selection, binned by decay mode, p_{T}, and $|\eta|$ in both of the $\ell\tau_{\mathrm{had}}$ channels. The fake factor is significantly higher for 1-prong

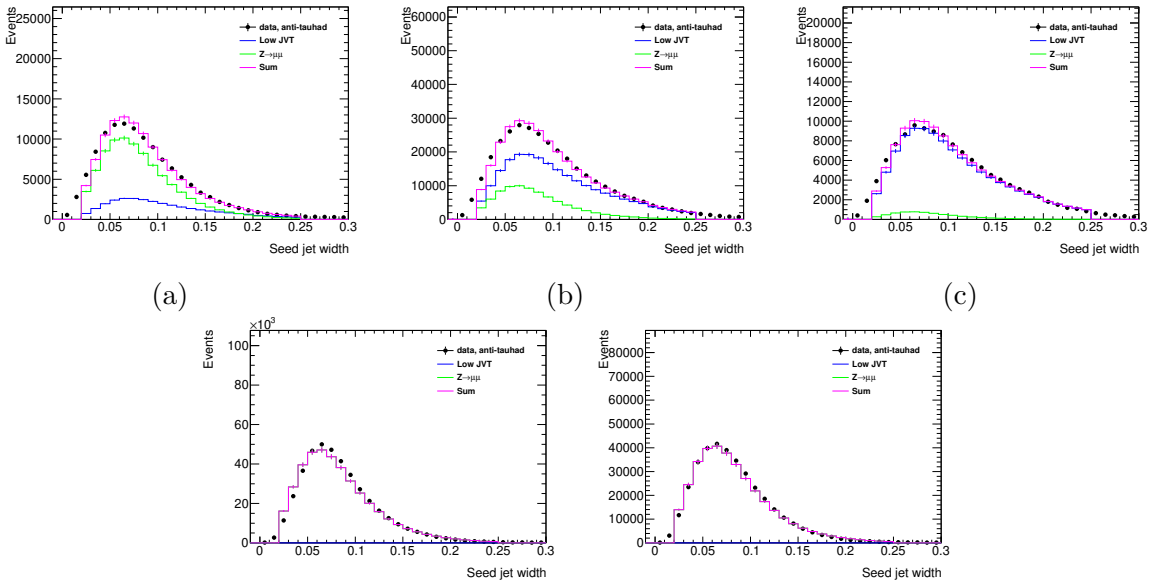


Figure 5.27: Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the inclusive SR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

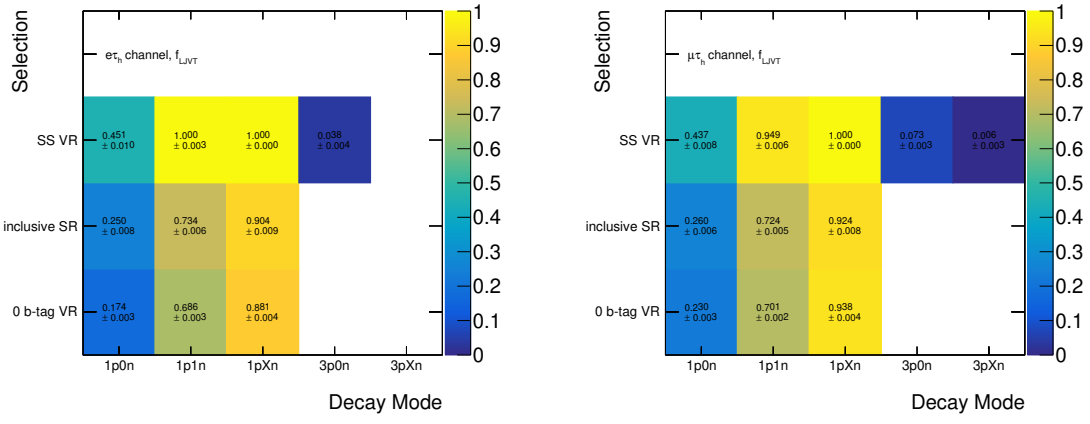


Figure 5.28: The best fit values of f_{LJVT} in the (a) $e\tau_{\text{had}}$ and (b) $\mu\tau_{\text{had}}$ channels, binned by selection and decay mode. Bins with f_{LJVT} =0 are shown in white.

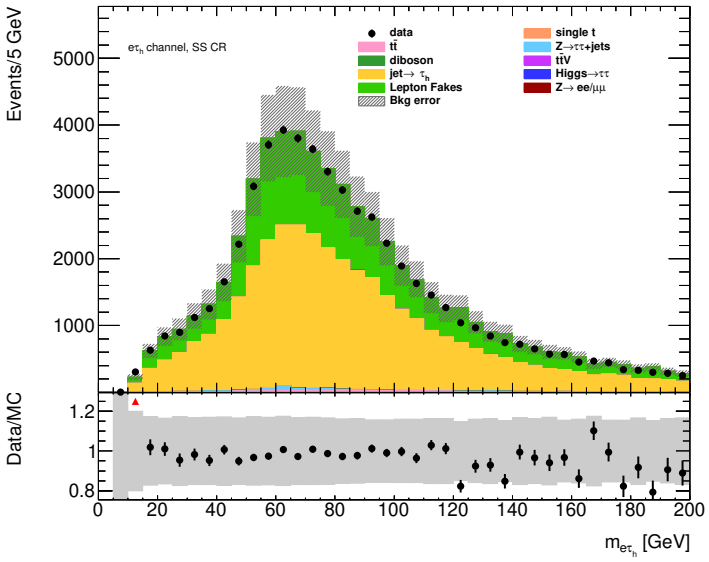


Figure 5.29: Visible mass of the di-τ system in the $e\tau_{\text{had}}$ channel for the SS CR selections. The dominant contributions are from fake τ_{had} (gold) and fake leptons (green).

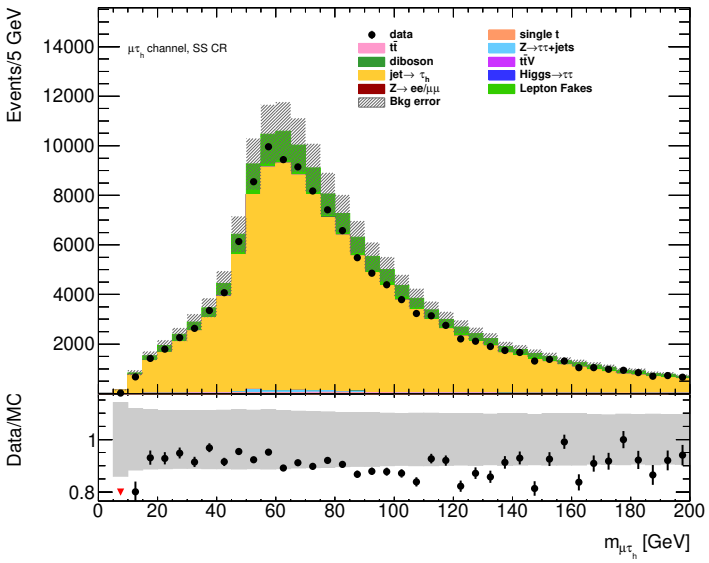


Figure 5.30: Visible mass of the di-τ system in the $\mu\tau_{\text{had}}$ channel for the SS CR selections. The dominant contributions are from fake τ_{had} (gold) and fake leptons (green).

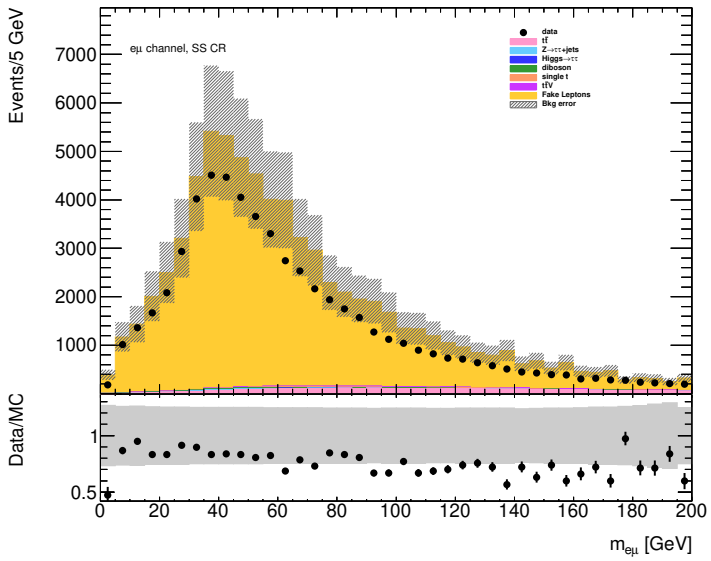
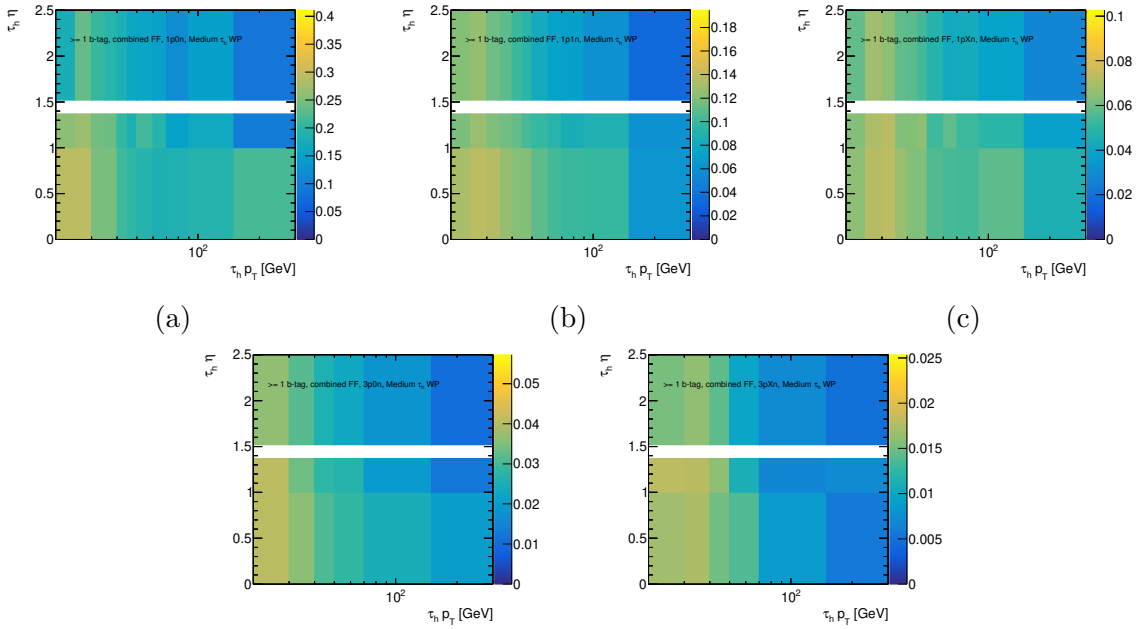


Figure 5.31: Visible mass of the di-τ system in the $e\mu$ channel for the SS CR selections. The dominant contributions is from fake leptons (gold).

tau decays than 3-prong because the 3-prong decay has more information for the RNN tau identification algorithm to exploit. For the corresponding plots in other channels and control regions, see Appendix A (Figures A.8–A.10).



(a) (b) (c)

(d) (e)

Figure 5.32: Combined FFs for the $\mu\tau_{\text{had}}$ channel under the SR selection (> 0 b-tag). Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

5.4.6 Fake leptons in the $\ell\tau_{\text{had}}$ final states

Fake leptons event counts are estimated in the $\ell\tau_{\text{had}}$ final states using a simplified version of the MM discussed in Section 5.4.4. In the $\ell\tau_{\text{had}}$ final states only one lepton is present, so the matrix is reduced to a 2x2 case and we drop superscripts for lepton flavor. The approximation that real leptons which have at least 27 GeV of p_{T} and are matched to a single lepton trigger will be correctly identified, or $\bar{\epsilon_r} = 0$, is used in this region. This approximation allows for the fake lepton template to be constructed entirely out of "Loose" leptons, without any contribution from "Tight" leptons.

The weight applied is then given by

$$w_L = \frac{\epsilon_r \epsilon_f}{\epsilon_r - \epsilon_f} \tag{5.16}$$

The MC events from prompt sources must be subtracted, as is done in the $e\mu$ channel;

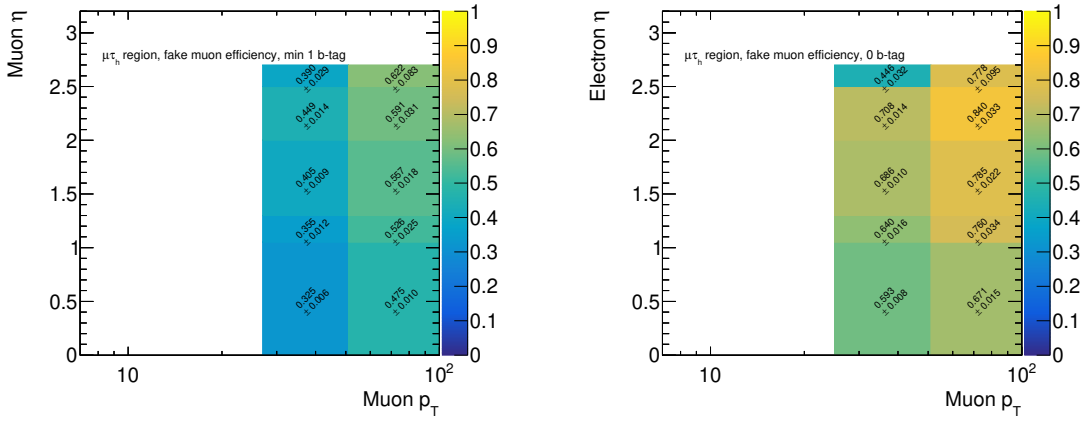
and additionally subtraction to account for the contribution due to fake taus in the "Loose" lepton region is also applied in this channel.

The number of events due to fake leptons in the region of interest is given by

$$N = w_L \times \left(N^{\text{data}}_{\tau_{\text{had}},L} - N^{\text{MC}}_{\tau_{\text{had}},L} - FF_{\text{combined}} \times \left(N_{\text{anti-}\tau_{\text{had}},L} - N^{\text{MC}}_{\text{anti-}\tau_{\text{had}},L}\right)\right) \tag{5.17}$$

where the subscripts τ_{had} and anti-τ_{had} accept or reject events with MEDIUM taus, respectively, and the superscripts "data" and "MC" refer to observed counts and truth-matched MC predictions, respectively.

The efficiencies for real leptons derived previously for the $e\mu$ channel are also applicable to the $\ell\tau_{\text{had}}$ regions. Some, but not all, of the fake lepton efficiencies are calculated in the $\ell\tau_{\text{had}}$ channel using the same phase space as the $e\mu$ channel. The efficiencies for electrons are reused in the $e\tau_{\text{had}}$ region; however, the phase space for muon p_{T} is different in the $\mu\tau_{\text{had}}$ channel than TRIG2 of the $e\mu$ channel due to the muon p_{T} cut in the TRIG2 region. Therefore, the fake muon efficiencies are derived separately for the $\mu\tau_{\text{had}}$ channel. The fake muon efficiencies used in the $\mu\tau_{\text{had}}$ channel are shown in Figure 5.33.



(a) (b)

Figure 5.33: Fake muon efficiencies for (a) 0 b-tag and (b) >0 b-tags in the $\mu\tau_{\text{had}}$ channel.

5.5 Neural Network Discriminant

A Parameterized Neural Network (PNN) [145] is used to discriminate between signal and background events. The signal MC is generated at discrete mass points and the true mass is used as one of the network inputs so that it generalizes across the mass spectrum. For the purpose of training the network, the signal MC is normalized so that each mass point has an equal sum of weights. The background is given an identical distribution of true masses, with equal weight given to delta functions at each simulated mass point (12, 15.5, 20, 25, 30, 40, 50, 60 GeV). One of the eight mass points is randomly chosen for each background event. A separate network is trained for the three channels ($e\mu$, $e\tau_{\text{had}}$, $\mu\tau_{\text{had}}$) and for each b-tagging category (1 b-jet, 2 or more b-jets and DeXTer).

The network inputs are various scalars constructed from the kinematics of the decay $h \to aa \to bb\tau\tau$, with the addition of the true mass. For the 1 b-jet category, the network inputs are the following

- $M^{\text{true}}(\tau\tau)$
- $M^{\text{vis}}(\tau\tau)$
- $M^{\text{MMC}}(\nu_e + \nu_\tau)$
- $M^{\text{MMC}}(\nu_\mu + \nu_\tau)$
- $E_{\text{T}}^{\text{miss}}$
- $M_{\text{T}}(leading\tau^{\text{vis}})$
- $M_{\text{T}}(subleading\tau^{\text{vis}})$
- $p_{\text{T}}(b)$
- $p_{\text{T}}(\tau\tau b)$
- $D_\zeta = \vec{E}_{\text{T}}^{\text{miss}} \cdot \vec{\zeta} - 0.85\,(p_{\text{T}}(\tau_1) + p_{\text{T}}(\tau_2)) \cdot \vec{\zeta}$

where $M(\tau\tau)$ is the mass of the di-tau system, for example, and the superscripts "true", "vis", or "MMC" refer to whether the mass is the true mass, visible mass, or the Missing Mass Calculator estimate of the di-tau mass. The τ^{vis} can be either an electron, muon, or hadronic tau, depending on the channel. The Missing Mass Calculator (MMC) is a tool developed by ATLAS to estimate the kinematics of the invisible neutrinos in τ decays, discussed previously in Section 5.3. It uses a Monte Carlo Markov Chain to sample the neutrino kinematics as a function of the visible kinematics and missing transverse energy. $M^{\text{MMC}}(\nu_e + \nu_\tau)$ and $M^{\text{MMC}}(\nu_\mu + \nu_\tau)$ are the MMC estimates of the masses of the neutrino systems (for $\ell\tau_{\text{had}}$, only one of these is used because by definition the hadronic tau side of the decay will have a single neutrino, which for our purposes has $m \to 0$). M_{T} is the transverse mass between the given object and the $E_{\text{T}}^{\text{miss}}$. In the definition of D_ζ [146], the vector $\vec{\zeta}$ is the angle bisector of the di-tau system. D_ζ quantifies whether the $E_{\text{T}}^{\text{miss}}$ tends to be in the same direction as the taus or not.

Additional variables are used in the network for events containing 2 b-jets or a DeXTer tag:

- subleading $p_{\text{T}}(b)$
- $p_{\text{T}}(bb)$
- $M(bb)$
- $M^{\text{vis}}(bb\tau\tau)$
- $M^{\text{MMC}}(bb\tau\tau)$

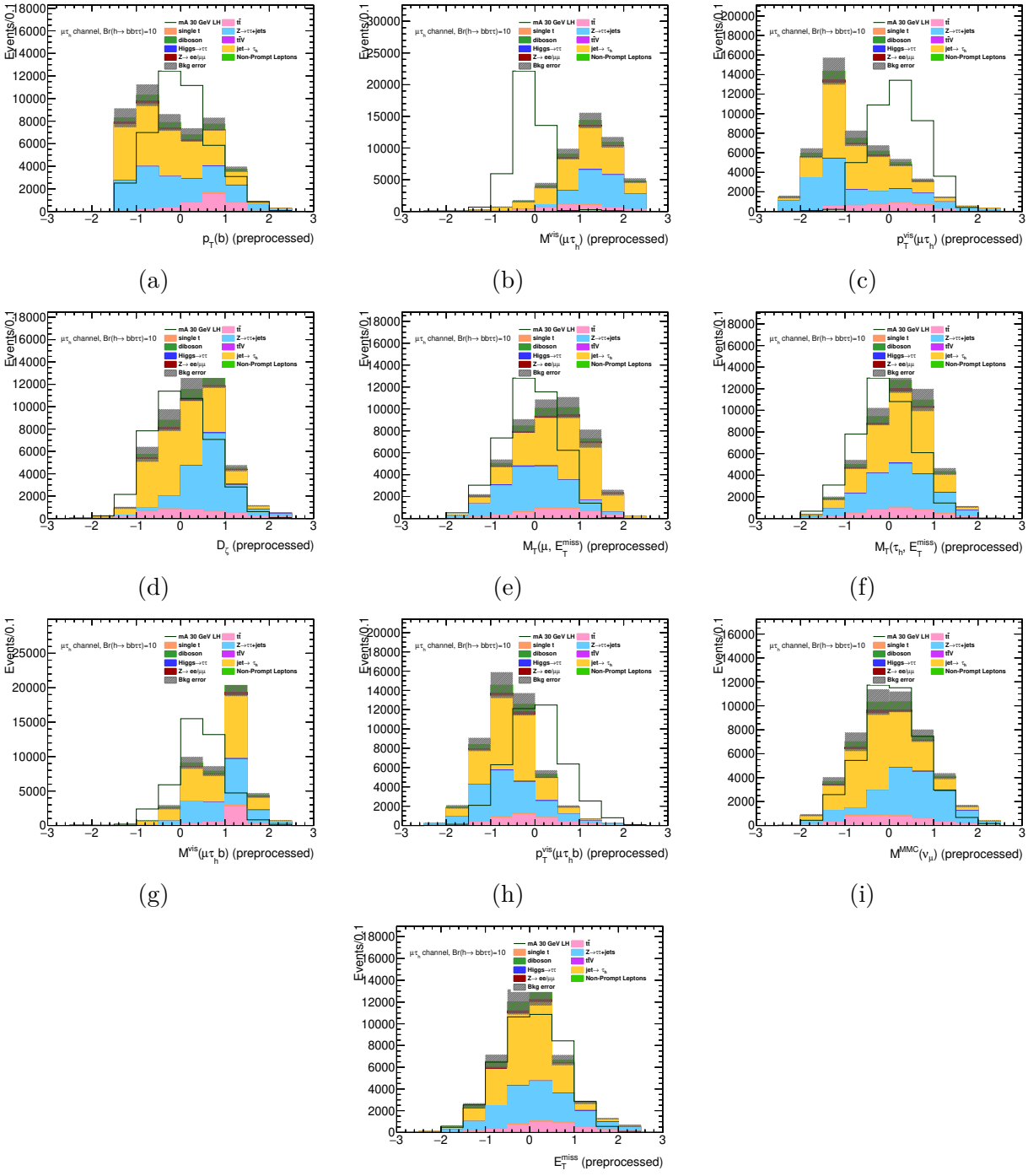
which are only defined when a second b-jet is tagged, or when a DeXTer jet is reconstructed. In the DeXTer category, each b in the definition of the input variables is replaced with the subjets of the DeXTer jet. Each DeXTer jet has precisely two subjets with a minimum p_{T} of 5 GeV.

The network is trained in the low mass region of Figures 5.18b and 5.19b (blinded region with visible mass of the di-tau system less than 60 GeV). A pre-training procedure is performed which gaussianizes the input signal distributions. The cumulative distribution function of the sum of the signal MC across all mass points is calculated for each input variable. Then the signal and backgrounds are then fed through the Cumulative Distribution Function (CDF) and inverse error function to gaussianize them in each dimension:

$$z_i^{\text{pre-trained}} = \text{erf}^{-1}\left(\text{CDF}_i(z_i)\right) \tag{5.18}$$

where z_i is the i^{th} component of the input vector, CDF_i is the CDF of the total signal MC estimate along the i^{th} input variable, and erf^{-1} is the inverse error function. The covariance of both the signal and background distributions remains nontrivial, since the pre-training is performed separately in each dimension. The gaussianization is performed with respect to the signal (and not the background) because (i) it puts the most signal-like part of the spectrum near zero, where the nonlinearity in the activation is, and (ii) better loss optimization is observed when gaussianizing with respect to the signals. The only input variable which is not gaussianized is the true mass, which is scaled uniformly to the domain [0,1]. Figure 5.34 shows the shapes of the pre-processed input variables for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category. For the distributions for each other channel and category, see Appendix A (Figures A.40–A.52). The distributions in which the total background shape is less gaussian are those which provide more discriminating power, because the pre-processing gaussianizes the signal.

Each PNN is a fully connected network with 3 hidden layers of width 15 neurons. The activation function is a leaky Rectified Linear Unit (ReLU) activation function with $a = 0.01$. The leaky ReLU activation function is given by $f(x; a) = \max(\text{x,ax})$. The network is trained against single t, $t\bar{t}$, $Z \rightarrow \tau\tau$ + jets, and fake lepton (tau) backgrounds in the $e\mu$ ($\ell\tau_{\text{had}}$) channels. The network training is supervised with signal and background classes marked.



(a) (b) (c)

(d) (e) (f)

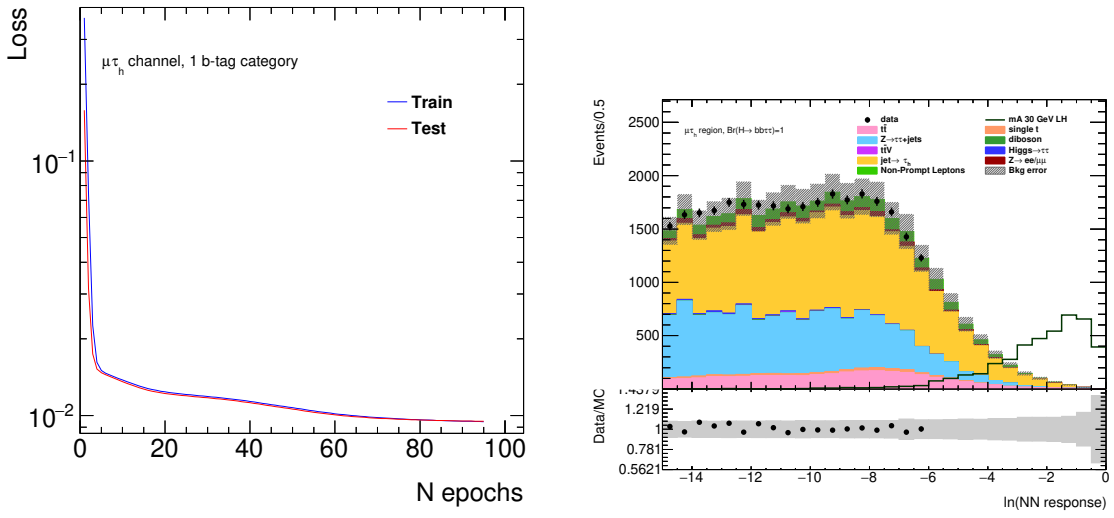
(g) (h) (i)

(j)

Figure 5.34: Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category.

The objax [147] framework is used with the Adam optimization algorithm [148] and a sigmoid cross entropy loss function. Any negative weight events are ignored in training, including negative weights due to MC subtraction. Training is performed for 300 epochs and early stopping is performed if the loss function does not improve for ten consecutive epochs. Figure 5.35 shows two aspects of the performance of the model: both the lack of over-fitting in the loss curve during training and the discriminating power of the distribution of network output scores. The background rejection power is very strong, especially considering that Figure 5.35 is only plotting the low-mass region (there is already a large amount of trivial background rejection from excluding the high mass sideband where an on-shell $h \to aa$ decay is forbidden). For the corresponding plots for each channel and category, see Appendix A (Figures A.54–A.56 for loss curves and Figures A.57–A.59 for network discriminants).

The high sideband in $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ was blinded during the developement of the analysis methodology. The blinding strategy is to ensure that $S/\sqrt{B} < 0.1$ in all unblinded bins for $Br(h \to bb\tau\tau) = 10\%$. The low $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ region is an additional CR which is kinematically closest to the SR. This region provides very strong model validation in the low mass, heavy-flavor enriched region where $bb\tau\tau$ Higgs decays would appear, just with an inverted cut on the NN score. In the > 1 b-jet and DeXTer regions the normalization is off by about 10-15% in Figures A.57–A.59 because the distributions are pre-fit (not yet constrained by the simultaneous fit including systematic uncertainties). Model validation, and in particular its ability to cover the normalization fluctuation in the double b-jet categories, is shown in Section 5.7. Figure 5.36 further validates the modeling of each individual discriminating variable in the low sideband of the $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ distribution for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category. No significant mismodeling of any network input variable is observed. See Appendix A (Figures A.40–A.52) for the distributions for the other channels and categories.



(a) (b)

Figure 5.35: (a) Cross-entropy loss function in the $\mu\tau_{\text{had}}$ 1 b-jet category. (b) Network output distribution for the $\mu\tau_{\text{had}}$ 1 b-jet category.

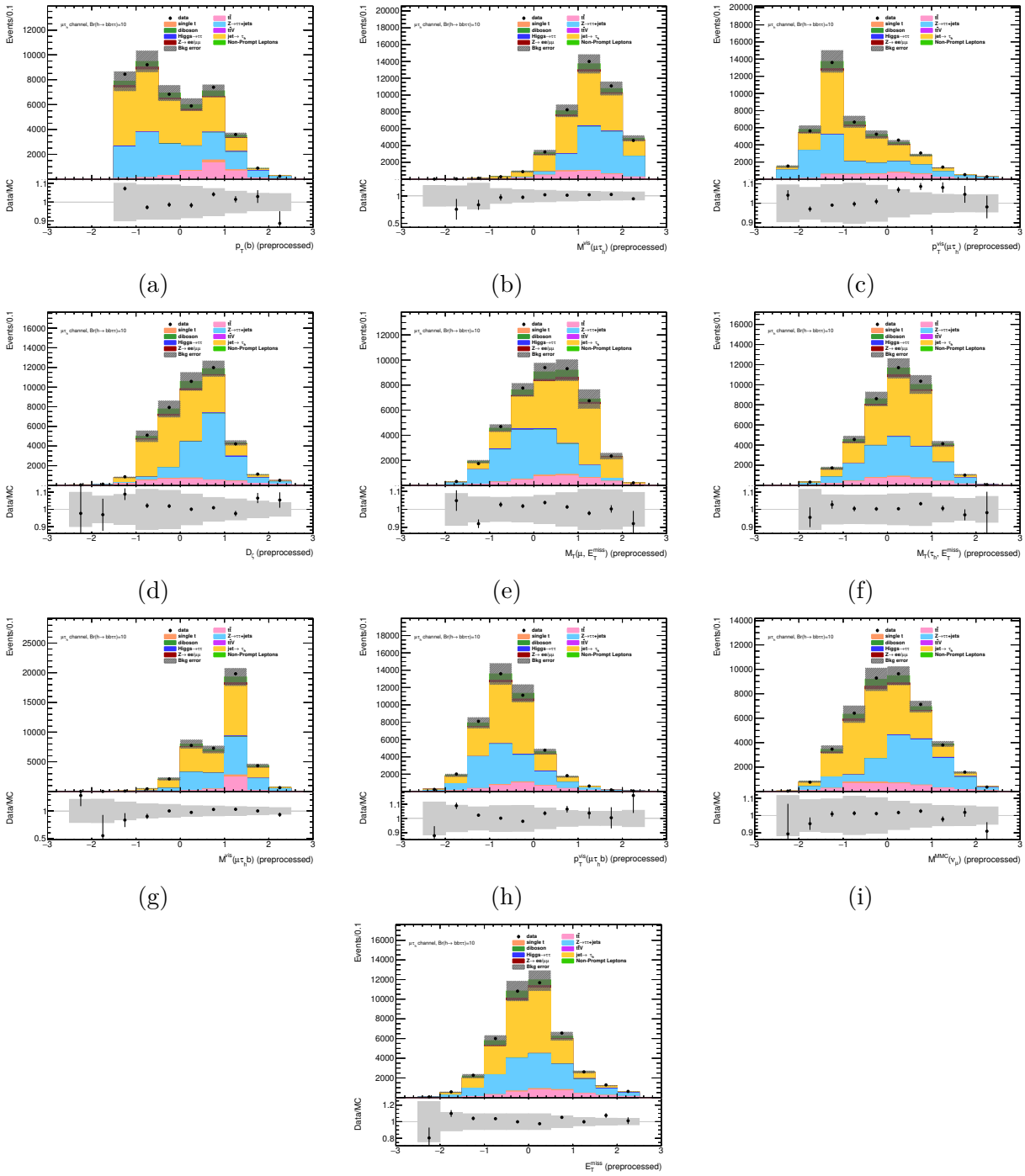


Figure 5.36: Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied.

5.6 Systematic Uncertainties

Systematic uncertainties in this analysis are divided into three categories: experimental uncertainties affecting the simulated background and signal processes, uncertainties arising from the data-driven fake estimations, and theoretical uncertainties on the simulated background and signal sample. These systematic uncertainties have the potential to affect both the yield and the shape of the final observables. A dedicated treatment of each type of systematic effect is included in the statistical framework via nuisance parameters (NPs). Each NP represents a single source of uncertainty and penalizes the likelihood with a gaussian constraint function. The NP may be *pulled* in the fit (for an example of a pull plot, see Figure 5.38) by some number of standard deviations (σ) away from its nominal value, but at a "cost" of a gaussian constraint $1/\sqrt{2\pi}e^{-1/2\sigma^2}$ multiplied into the likelihood function. However, the pulling of the NP may improve the modeling in several bins, increasing the overall likelihood. All systematic uncertainties considered are introduced as NPs in the fit of the background model to the observed data. For more information on the PLR method, see Section 5.7 and [149]. The nuisance parameters are summarized in Table 5.8.

5.6.1 Experimental Uncertainties

Most of the experimental uncertainties are estimated following the recommendations provided by the ATLAS Combined Performance groups. Additional uncertainties based on the DeXTer jet flavor-tagging and the low-p_{T} jets used in this analysis are detailed in Section 5.6.1.5. All of the experimental uncertainties for each reconstructed object are listed in the following sections.

5.6.1.1 Electrons

The uncertainties related to electrons are estimated using the techniques described in [150]. There are 7 NPs used to capture uncertainty on electrons:

Table 5.8: The number of nuisance parameters in the fit, broken down by category.

Category	Description	Number NPs
Electrons	Energy resolution and scale	2
	Reconstruction efficiencies	5
Muons	Momentum calibration	3
	Sagitta bias	1
	Reconstruction efficiencies	8
Taus	Tau energy scale	3
	Reconstruction efficiencies	14
Jets	Energy resolution and scale	15
	JVT scale factor uncertainties	2
$E_{\mathrm{T}}^{\mathrm{miss}}$	Soft term variations	2
Flavor Tagging	DL1r	19
	DeXTer	12
Luminosity	—	1
Pileup	—	1
Fake Leptons	Isolation efficiencies	4
	MC subtraction	1
	MC efficiency for real leptons	2
	Composition differences	1
	low ΔR correction	1
Fake Taus	FF stat uncertainty	3
	MC subtraction	1
	Composition differences	2
Theoretical	Minor background normalization	1
	DY SR normalization	6
Total	—	110

- Energy Resolution and Scale: All of the effects on the energy resolution and scale are considered fully correlated in η and are summed in quadrature. This simplification leads to having only 2 systematic variations with a separate variation for the scale and resolution.

- Reconstruction efficiencies: These uncertainties include the efficiencies on the electron reconstruction, identification, isolation, and trigger (2 NPs) for a total of 5 NPs.

5.6.1.2 Muons

Uncertainties related to muons are described using 12 NPs and come from the following sources:

- The muon momentum calibration considers uncertainties in the muon track momentum scale and resolution. The momentum scale is treated as a single nuisance parameter while the resolution is separated into variations on the muon spectrometer (MS) track and the inner detector (ID) track resolutions. 3 NPs are used.

- Additional momentum uncertainty associated with the sagitta bias is described using 1 NP.

- Statistical and systematic uncertainties associated with the muon reconstruction/identification, isolation, and trigger efficiencies. Each of these sources has two nuisance parameters (one describing the statistical component and one describing the systematic component). 8 NPs are used to capture these uncertainties.

5.6.1.3 Taus

The uncertainties related to taus are estimated from the techniques describe in [75] which have been updated with the full Run-2 data set. Uncertainties on taus are described using 17 NPs from the following sources:

- The uncertainties associated with the variation of the tau energy scale. The uncertainties considered describe the effects from the modeling of the tau energy (generator choice, geometry of the detector, underlying event, and pile-up), the in-situ energy correction measurement, and the detector response. 3 NPs are used to describe these uncertainties.

- The uncertainties associated with the tau reconstruction, RNN identification, and electron veto efficiencies. The RNN identification efficiency is separated into different components based on the p_{T}, number of prongs, and source (statistical or systematic) of the uncertainty on the efficiency. The reconstruction efficiency is described by a single nuisance parameter accounting for the differences between the nominal detector geometry and GEANT4 physics-lists and those coming from alternative choices. The electron veto uncertainties account for the effect on electrons (fakes) and true τ_{had} candidates and come from measurements in a dedicated tag-and-probe analysis. 14 NPs are used to describe these uncertainties.

5.6.1.4 Jets and Missing E_T

The uncertainties related to jets arise from a variety of sources and are described using 17 NPs:

- The jet energy scale and jet energy resolution uncertainties were measured using a combination of test-beam data, Run-2 LHC collision data, and simulation [151]. The jet energy scale uncertainties are described by 7 nuisance parameters: three η intercalibration non-closure uncertainties and four additional nuisance parameters which are combinations of the complete set of 100 nuisance parameters in the full jet energy scale uncertainties. The jet energy resolution uncertainty uses a reduction scheme parameterized by 8 independent sources.

- The JVT scale factor uncertainties arising from differences seen in $Z \to \mu\mu +$ jet events

is accounted for with 2 NPs.

The missing E_{T} uncertainties considered account for the variations in the scale and resolution of the soft track term. These uncertainties are parameterized by two distinct nuisance parameters, where the resolution uncertainty has separate components corresponding to the smearing of the soft term magnitude in the direction parallel or perpendicular to the direction of the hard component of p_{T} [152].

5.6.1.5 Flavor Tagging

Two different flavor tagging strategies are used: the DL1r and DeXTer taggers. A total of 31 NPs describe the uncertainty in flavor-tagging efficiency:

- The DL1r flavour tagging uncertainties come from the variation of the scale factors used to match the flavour tagging efficiencies and mis-tag rates in MC to those obtained in data. An eigenvector decomposition is performed to reduce the number of systematic uncertainties to 19 NPs.

- The use of DeXTer jets in this analysis introduced additional SFs to correct the efficiency difference between simulated samples and data. The SFs are derived and provided by the DeXTer development team [153] using 12 NPs.

5.6.1.6 Luminosity and Pileup

The uncertainty in the combined 2015–2018 integrated luminosity is 0.83% [154], obtained using the LUCID-2 detector [155] for the primary luminosity measurements, complemented by measurements using the inner detector and calorimeters. Pileup reweighting is performed using a standard procedure of 1-σ up and down variations. 2 NPs are used: one each for pileup and luminosity uncertainties.

5.6.2 Fake lepton and tau estimates

5.6.2.1 Fake leptons

The sources of uncertainty related to the matrix method fake lepton estimate are described using 9 NPs:

- Statistical uncertainty on the efficiencies. The four efficiencies (real and fake efficiencies for both electrons and muons) have statistical uncertainty as shown in Figures 5.22–5.24. Four NPs describe these four up and down variations in the $e\mu$ channel. In the $\ell\tau_{\text{had}}$ channels only two of the nuisance parameters, corresponding to the correct lepton flavour, have an effect.

- MC subtraction. A 30% uncertainty on the prompt MC subtraction yield is taken. One NP is used to cover this source of uncertainty.

- Uncertainty on the real efficiencies. The nominal real efficiency is derived using the $Z \to \tau^+\tau^-$ MC. The real efficiencies (on both muons and electrons) are derived instead using the $t\bar{t}$ MC and we take the variation in each bin as an uncertainty, using two NPs.

- Composition differences between signal and control regions. The signal region may not have the same fraction of gluons vs light jets vs b-jets, which is an uncertainty on the fake estimate. The fake efficiencies in the 0 b-jet CR (Figures A.1 and A.2) are substituted for those in the heavy-flavor enriched region (Figures 5.23 and 5.24). One NP is used for this uncertainty.

- Matrix Method correction for $\Delta R < 0.6$. The parametric correction (see Figure 5.25) is derived with a single floating normalization factor. The size of the normalization uncertainty is estimated by perturbing the normalization until the model deviated by one standard deviation in the bin with $\Delta R < 0.6$. One NP is used to cover uncertainty on the $\Delta R < 0.6$ correction.

Table 5.9: Table of fake lepton origin in the $e\mu$ channel

Region	Unknown (%)	Conversions (%)	Light (%)	Heavy (%)
electron OS SR	52.54	8.45	0.16	38.84
electron SS CR	52.58	12.02	0.52	34.88
electron SS VR	65.76	14.57	1.13	18.53
muon OS SR	2.98	0.19	0.96	95.86
muon SS CR	6.11	0.26	1.39	92.24
muon SS VR	7.77	0.01	2.36	89.86

5.6.2.2 Fake Lepton Composition Study

In this section a study of the composition of fake leptons is presented. MC is binned according to the true source of leptons which pass isolation but are not truth matched to a lepton from the decay of a W/Z boson or top quark. Four different sources are considered: light-flavor jets (u, d, s), heavy-flavor jets (c, b), photon conversions, and "unknown," which includes decays in flight and other detector affects. The relative composition from these four sources is estimated in three regions: the SS events with a b-tag (DL1r or Dexter; denoted SS CR), SS events without a b-tag (SS VR), and OS events with a b-tag (OS SR). The SS CR is where the nominal efficiency is *derived* and the OS SR is where the nominal efficiency is *applied*. Any variation between the composition in these two regions must be covered by a systematic uncertainty. The strategy of using the efficiencies derived in the SS VR to cover the variation is adopted because the variation between the SS CR and OS SR is less than that between the SS VR and SS CR. The estimated compositions are summarized in Table 5.9.

5.6.2.3 Fake taus

The sources of uncertainty related to the fake tau estimate are described using 6 NPs:

- Statistical uncertainty on the fake factors. Three NPs are used for the three types of fake factors (dijet low JVT, dijet high JVT and $Z \to \mu\mu$).

- MC subtraction. A 30% uncertainty on the prompt MC subtraction yield is taken. One NP is used to cover this source of uncertainty.

- Composition differences between signal and control regions. The signal region may not have the same fraction of gluons vs light jets vs b-jets, which is an uncertainty on the fake estimate. The fractions f_{LJVT} and r_{QCD} in the 0 b-jet CR are substituted for those in the heavy-flavor enriched signal region. The procedure is analogous to the substitution of lepton efficiencies in the heavy-flavor depleted region. See Figure 5.28 for f_{LJVT} and Figure 5.26 for r_{QCD}. Two NPs are used, one for f_{LJVT} and another for r_{QCD}.

5.6.3 Theoretical Uncertainties

There are several possible sources of uncertainty on the MC event generation, which henceforth are referred to as "theoretical uncertainties". Theoretical uncertainties can in general change the normalization (overall number of events predicted) and shape (distribution of these events into the many bins of the PLR fit) of the background prediction. Theoretical uncertainties are only estimated for major backgrounds estimated using MC: top and Drell-Yan + jets. Top and Drell-Yan backgrounds use unconstrained nuisance parameters (these NPs do not penalize the likelihood function) to fit the correct shape with respect to the jet multiplicity and normalization. Therefore, only the theoretical uncertainties with affect the shapes of other kinematic variables are considered. Minor backgrounds (including SM $H \to \tau\tau$, VV and ttV) are assumed to have a 20% normalization uncertainty, captured by a single nuisance parameter. No theoretical uncertainty is required for data-driven fake estimation.

What follows is a high-level explanation of the origin of theoretical uncertainties in the evaluation of cross sections. The p-p cross section is evaluated in n^{th} order perturbation

theory as:

$$\sigma^{(n)} = PDF(x_1, \mu_F) \otimes PDF(x_2, \mu_F) \otimes \hat{\sigma}^{(n)}(x_1, x_2, \mu_R) \tag{5.19}$$

$$\hat{\sigma}^{(n)} = \alpha_s \hat{\sigma}^{(0)} + \alpha_s^2 \hat{\sigma}^{(1)} + ... + \alpha_s^n \hat{\sigma}^{(n)} + \mathcal{O}(\alpha_s^{(n+1)}) \tag{5.20}$$

where x_1 and x_2 are the particular partons interacting and $PDF(x, \mu_F)$ is the parton distribution function, evaluated at some factorization scale μ_F. σ is the cross section and α_s is the strong coupling, renormalized at a renormalization scale of μ_R. The hat notation separates "bare" cross sections from those "dressed" with Parton Distribution Functions (PDFs), and superscripts (n) denote the order of the diagrams evaluated in powers of α_s.

The value of the strong coupling constant is determined experimentally at $Q = m_Z$ and then estimated at different interaction scales using the Renormalization Group Equation:

$$\frac{d\alpha_s(Q)}{d \log Q^2} = -b_0 \alpha_s^2(Q) - b_1 \alpha_s^3(Q) + ... \tag{5.21}$$

In a method similar to the renormalization of α_s, the PDFs are fit to data at a scale Q and then evolved using the DGLAP equations, a different differential equation which is also truncated at some order of α_s.

This scheme has *arbitrary choices* in the renormalization and factorization scales, μ_R and μ_F; *missing higher order terms* in the cross section, the evolution of the PDFs, and the renormalization of α_s; and *inputs with uncertainty* in the observation of PDFs and α_s at given scales. These many sources of uncertainty are addressed in two categories: uncertainties associated with arbitrary scales (μ_R and μ_F, as well as their affect on α_s) and uncertainties associated with PDFs.

Uncertainties associated with the factorization and renormalization scales affect the cross section $\sigma^{(n)}$. These uncertainties are estimated by varying the values of μ_R and μ_F up and down by a factor of 2 and comparing the resulting cross section to the nominal prediction. Under the systematic variations the top background normalization changes by 14%, but no

significant shape is observed with respect to the NN output variable. Similarly, the shape of the Drell-Yan background is not modified but the normalization changes by 30%. Therefore, no additional nuisance parameters are introduced for factorization and renormalization scales. Note that an uncertainty is not omitted here. The unconstrained NPs do cover this affect. We did attempt to add constrained NPs for the normalization of top and DY backgrounds, but these NPs end up fixed to a value near zero because the uncertainty is already covered by another parameter which does not penalize the likelihood.

Uncertainties associated with the renormalization of the strong coupling constant are estimated by varying the renormalization scale for α_s, affecting initial and final state radiation separately from those scales associated with the primary hard scatter process. Again the scale is varied up and down by a factor of 2. The uncertainties on α_s affecting initial and final state radiation are in general smaller than the renormalization and factorization scale uncertainties in the hard scatter process. The shape is again not modified and no new NP is introduced due to the existing unconstrained NPs ability to cover this variation.

Uncertainties associated with PDFs are evaluated by using 100 systematically varied PDF sets and taking the standard deviation of the predicted number of events as a function of the NN output score. The 100 variations cover the uncertainty on the input data and fits, as well as the evolution of the PDFs using the DGLAP equation. PDF uncertainties are negligible for top backgrounds (order of 0.02%), and are omitted. PDF uncertainties on DY+jets are on the order of the single percent level in signal region bins, and are negligible in CRs. Therefore the uncertainty has a significant shape with respect to the NN discriminant and an additional nuisance parameter is introduced to cover the variation. In 6 of the 9 SR bins there is a significant variation due to uncertainties on the PDF in the Drell-Yan background. The DY+jets process is much more susceptible to PDF variations because there are not tops which will naturally produce b-jets in a more easily predicted hard-scatter process. Rather, the associated heavy-flavor jets are often coming from the b content of the proton, which is susceptible to PDF variations. Six nuisance parameters are introduced to cover the variation

Table 5.10: Size of the systematic variation from PDF uncertainties for the DY background. Uncertainties with a $< 1\%$ affect on the background are omitted.

Category	Theory systematic size (%)
$e\mu$, 1 b-jet	6
$e\mu$, 2 b-jet	3
$e\mu$, DeXTer	7
$\mu\tau_{\text{had}}$, 1 b-jet	4
$\mu\tau_{\text{had}}$, 2 b-jet	< 1
$\mu\tau_{\text{had}}$, DeXTer	< 1
$e\tau_{\text{had}}$, 1 b-jet	8
$e\tau_{\text{had}}$, 2 b-jet	10
$e\tau_{\text{had}}$, DeXTer	< 1

of the DY background in 6 SR bins, in addition to the one NP used for the normalization on minor backgrounds, which are not fit using any unconstrained NPs. The size of the systematic variation for DY is shown in Table 5.10.

5.7 Measurement Technique

Previous sections and chapters have described several aspects of the *model* used to predict the data observed in the signal and control regions. The model includes a data-driven fake estimate, a set of backgrounds estimated with MC methods, reweighting of MC methods in various CRs, and an estimation of the size of many systematic uncertainties. The utility of the model is in its prediction of event counts in given bins, as shown in the many histograms presented to this point. Next, the model must be leveraged in some way to set exclusion limits on or discover the BSM process $h \to aa \to bb\tau\tau$. In this section such a measurement technique is described in full detail. Then, the model is validated via a simultaneous fit to all CRs.

The likelihood of observing a set of data $\boldsymbol{n} = (n_1, n_2, ...n_N)$ in N bins given a model which predicts signal counts $\boldsymbol{s} = (s_1, s_2, ...s_N)$ and background counts $\boldsymbol{b} = (b_1, b_2, ...b_N)$ given the nuisance parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_M)$ is

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j(\boldsymbol{\theta}) + b_j(\boldsymbol{\theta}))^{n_j}}{n_j!} e^{\mu s_j(\boldsymbol{\theta}) + b_j(\boldsymbol{\theta})} \prod_{k=1}^{M} \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta_k}{\sigma_k}\right)^2} \tag{5.22}$$

where μ is the signal strength and σ_k is the size of the uncertainty on the k^{th} NP. The first term is the product of Poissonian constraints on the observed counts, and the second term is the constraints on the nuisance parameters. Here we adopt the conventions that (i) $s_j(\boldsymbol{\theta})$ and $b_j(\boldsymbol{\theta})$ are the signal and background counts at the given point in model-space $\boldsymbol{\theta}$, (ii) that all constraints on NPs are gaussian (which is not always true, but a suitable constraint function can be substituted in the second product) and (iii) that the nominal value of each nuisance parameter is 0. A test statistic $\tilde{q_\mu}$ is used to compare the likelihood of the nominal and BSM hypotheses:

$$\tilde{q_\mu} = \begin{cases} -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0 \\ -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{5.23}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the values of μ and $\boldsymbol{\theta}$ which maximize $L(\mu, \boldsymbol{\theta})$ and $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ is the maximum-likelihood estimator of $\boldsymbol{\theta}$, given a fixed value of μ. The equation for $\tilde{q_\mu}$ is understood to describe the test statistic for an upper limit because

1. The first case $\hat{\mu} < 0$ describes the likelihood ratio for a single-sided signal strength, where if the best-fit signal strength is negative its value is capped at 0. The hypothesized signal will only add events to the bins, not take any away.

2. When $0 \leq \hat{\mu} < \mu$ the likelihoods of the best-fit model and the hypothesis μ are compared.

3. If the best-fit value of μ is larger than the value of μ for which we want to calculate the test statistic, then $\tilde{q_\mu} = 0$. This is understood as assigning maximal compatibility between data and model when the μ hypothesis is "too small" compared to the its

optimal value. We only can discover an excess, not a deficit, of events.

The distribution of the test statistic is integrated to calculate the p-value:

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu \tag{5.24}$$

where the probability density function of the test statistic $f(q_\mu|\mu)$ is obtained either via asymptotic formulae or sampling the distribution directly using so-called "toys." In this work, only asymptotic approximations are used because only p-values on the extreme tails of the test statistic distribution are significantly changed using toys [149].

Discovery or upper limits on the signal strength requires a statistical test with some test statistic and threshold on the probability, or p-value, of the BSM hypothesis given the data. For discovery a threshold of five standard deviations, or $p < 3 \times 10^{-7}$ is used, while upper limit exclusions use $p < 0.05$. When many independent experiments are performed, as in our mass scan,[2] the look-elsewhere effect must be considered [156]. However, as shown in Chapter VI, there is no significant excess for which to compute the look-elsewhere effect.

We next describe the binning scheme. Each channel has CR bins defined in the high mass and $t\bar{t}$ CRs. There are additional CR bins defined in the low $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ region as described in Section 5.5. The SR and low neural network output bins are split into three heavy flavor categories (see Figure 5.4) for each channel. Specifically, the number of bins and purpose for each set of bins is as follows.

1. $t\bar{t}$ CR bins. As described in Section 5.4.2, a reweighting procedure is applied to the $t\bar{t}$ MC. The weights are unconstrained normalization factors in the PLR fit. For $t\bar{t}$ 7 weights are used: 3 b-jet multiplicity weights (1, 2 and > 2 b-jets) and 4 jet (inclusive of b-jets) multiplicity weights: (1, 2, 3, and > 3 jets). To constrain these 7 normalizations the $t\bar{t}$ CR is split into 9 bins per channel (there are 12 combinations but 3 of these are

[2] In which nearby points are slightly correlated but not entirely so

undefined: for example, it is impossible to have 1 total jet and 2 b-jets). A total of 27 bins are used to constrain the normalization on $t\bar{t}$.

2. High mass CR bins. The high mass CR bins are defined in the high sideband of the variable $m_{\text{vis}}(\tau\tau)$. These bins include both a set of bins to normalize $Z+$ jets and a set of bins for the $\ell\tau_{\text{had}}$ channels to normalize the data-driven fake tau estimate. Section 5.4.3 described the procedure for reweighting the $Z+$ jets MC, including the $Z \to \ell\ell$ contribution in the $\ell\tau_{\text{had}}$ channels. In the $\ell\tau_{\text{had}}$ channels two bins are used per heavy flavor category: one between 60 and 85 GeV and another between 85 and 105 GeV which constraint $Z \to \tau\tau$ and $Z \to \ell\ell$, respectively. In the $e\mu$ channel one bin is used between 45 and 65 GeV for each flavor category, for a total of 3 bins (there is no considerable $Z \to \ell\ell$ contribution). For the fake tau normalization bins, an additional bin for each heavy flavor category is used above 105 GeV. A total of 21 bins are used to constrain the normalization on $Z+$ jets and fake taus in the high mass CR.

3. The SRs are subsets of the blinded $m_{\text{vis}}(\tau\tau) < 60(45)$ GeV region for the $\ell\tau_{\text{had}}$ ($e\mu$) channels, shown in Figures 5.18b, 5.19b and 5.20b (inclusive of heavy flavor) and Figure 5.21(DeXTer selection). The SRs have cuts on both $\ln(\text{PNN}((\vec{z}^{\text{pre-trained}}))$ and $m_{\text{MMC}}(\tau\tau)$. The SR bin is defined between $0.95\times m_{\text{MMC}}(\tau\tau)-6$ and $1.15\times m_{\text{MMC}}(\tau\tau)+1$ [GeV] for $\ell\tau_{\text{had}}$ channels and between $0.75 \times m_{\text{MMC}}(\tau\tau) - 1$ and $1.25 \times m_{\text{MMC}}(\tau\tau)$ [GeV] for the $e\mu$ channel. A scan over the neural network discriminant in the high output sideband of $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ shown in Figures A.57–A.59 is performed for each channel and category. The choice which produces the best limit on $\text{Br}(h \to aa \to bb\tau\tau)$ is adopted, and these cuts are shown in Table 5.12. There are 9 SR bins, one for each category.

4. Neural Network CR bins. The low $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ CR is defined for each category and is used in the statistical model. Bins identical to the SR in $m_{\text{MMC}}(\tau\tau)$ are used. The low PNN output CR is closer kinematically to the signal region (compared to the

$t\bar{t}$ and high mass CRs) because it is defined in the $m_{\text{vis}}(\tau\tau) < 60(45)$ GeV region for the $\ell\tau_{\text{had}}$ ($e\mu$) channel. This region validates the ability of the statistical model to extrapolate to the low mass regime ($m_{\text{vis}}(\tau\tau) < 60(45)$ GeV) without unblinding the SR. There are 9 NN CR bins, one for each category.

A summary of the complete binning scheme is shown in Table 5.11.

As discussed in Section 5.5, the > 1 b-jet and DeXTer categories see some incorrect normalizations in the low $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ CR. Figure 5.37 shows the event counts in each bin pre-fit and post-fit for the $m_a = 30$ GeV mass point. The first 9 bins are the 9 signal categories and the next 9 bins are the corresponding NN CR bins. The other 48 bins are the high mass and $t\bar{t}$ CRs for each channel and category. The agreement in the low NN CR (bins 10-18) demonstrates the degree to which the model can account for the pre-fit normalization mismodeling in the $m_a < 60(45)$ GeV region. After the fit the event counts in the low NN output and other CR bins agree with the prediction. The agreement in the first 9 bins shows that there is probably not an excess at $m_a = 30$ GeV. The next Chapter presents the results and makes more quantitative statements about the absence or presence of a new physics signal.

The postfit correlation and pulls of each constrained nuisance parameter for the $m_a = 30$ GeV mass point are shown in Figure 5.38. Most nuisance parameters should have a best fit value near 0 and a post-fit error of 1. The plot ranks NPs on the size of the post-fit error bars. This is true for any NP which this experiment cannot constrain better than a dedicated analysis. Nuisance parameters associated with the fake template are derived specific to this analysis, and as such post-fit errors less than 1 are tolerable and expected. The data-driven fake estimate is specific to this phase space and no auxiliary analysis can be performed to better constrain the fake estimate. There are a minority of NPs not related to the fake template which have some moderate post-fit constraint. These are the NPs related to flavor-tagging with the largest impact on the background template. These parameters are likely constrained because of the binning scheme, which has separate bins for events with

different b-tag multiplicity, which gives a good constraint on the flavor tagging efficiency. Overall, the pull plot shows no evidence of a poorly constructed likelihood.

The correlation matrix also shows good performance of the model. In general correlations between nuisance parameters should be low, with some exceptions. One such exception is with the fake template, in which the largest correlation of nuisance parameters is observed. For example, the fake tau f_{LJVT} and r_{QCD} are correlated because they both modulate the normalization of taus, which is well constrained by the control regions. There is also some small correlation between NPs which are eigenvariations of the same flavor-tagging technique (DL1r or DeXTer). To illustrate this the correlation matrix is organized in a block-diagonal format with nearby parameters affecting the same aspect of the model. These correlations are expected and since they remain relatively small, they are not an issue. Large correlations between completely different aspects of the model would be cause for concern, but are not observed in this analysis.

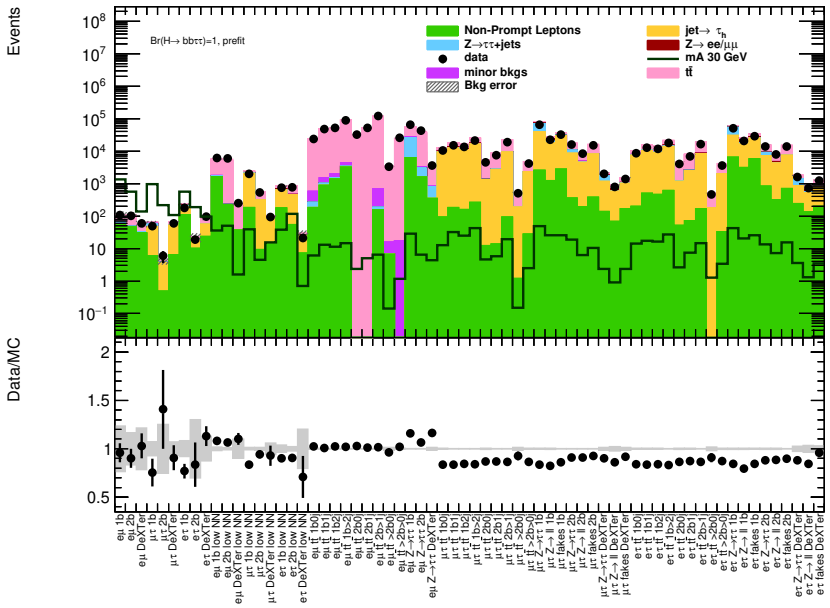
The other mass points are shown in Appendix A (see Figures A.60–A.66 for event count plots and Figures A.67–A.73 for pulls/correlation matrices). Across all 9 mass points, 68% of the 72 SR bins are within 1σ of the post-fit expectation, which is a strong sanity check of the model performance. While Figure 5.37 and its counterparts in the Appendices show the observed event counts in the signal region bins, the study of pulls, correlations, and modeling in the NN CR was performed while the analysis was still blinded to the SRs. Only after the modeling was validated were the event counts in the signal regions unblinded, and at this point the analysis methodology was already frozen.

Table 5.11: Summary of binning scheme for statistical interpretation. Note that the SR and low NN output bin definitions change depending on the mass hypothesis being tested. For more information, see text.

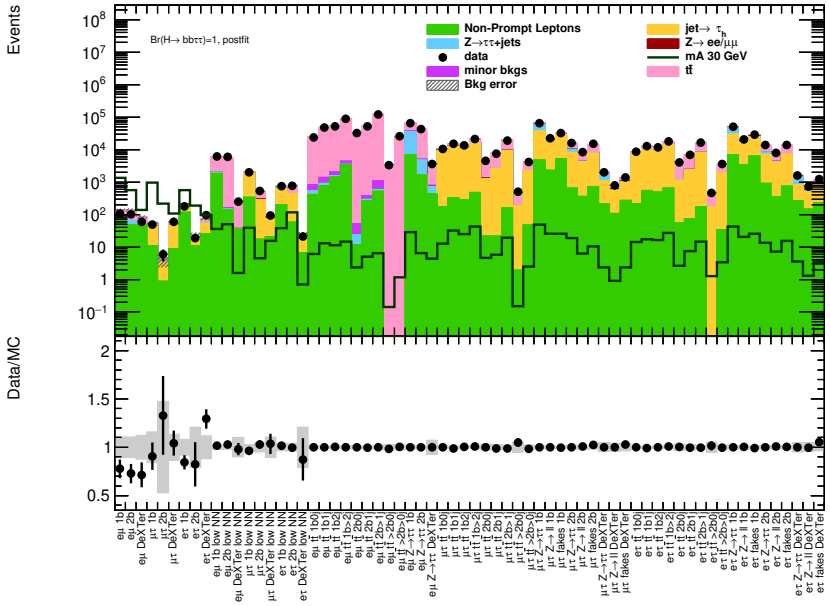
Region	Purpose	N bins
Signal	Search for BSM Higgs Decay	9
Low NN output	Validate the model in low mass sideband	9
$t\bar{t}$ CR	Constrain 7 MC weights, $e\mu$ channel	9
	Constrain 7 MC weights, $\mu\tau_{\text{had}}$ channel	9
	Constrain 7 MC weights, $e\tau_{\text{had}}$ channel	9
High Mass CR	Constrain $Z \to \tau\tau$ (all categories)	9
	Constrain $Z \to \tau\tau$ weights ($\ell\tau_{\text{had}}$ categories)	6
	Constrain Fake Tau composition ($\ell\tau_{\text{had}}$ categories)	6
Total	—	66

Table 5.12: The cut on ln(PNN($\vec{z}^{\text{pre-trained}}$)) which produces the highest sensitivity in each category is adopted. These cuts leave some region of ln(PNN($\vec{z}^{\text{pre-trained}}$)) between the CR and SR for each category.

Category	SR ln(PNN) range
$e\mu$ DeXTer	[-3.0, 0]
$e\mu$ 1 b-tag	[-1.5, 0]
$e\mu$ 2 b-tag	[-2.5, 0]
$\mu\tau_{\text{had}}$ DeXTer	[-3.0, 0]
$\mu\tau_{\text{had}}$ 1 b-tag	[-2.0, 0]
$\mu\tau_{\text{had}}$ 2 b-tag	[-1.0, 0]
$e\tau_{\text{had}}$ DeXTer	[-2.0, 0]
$e\tau_{\text{had}}$ 1 b-tag	[-7.5, 0]
$e\tau_{\text{had}}$ 2 b-tag	[-7.0, 0]

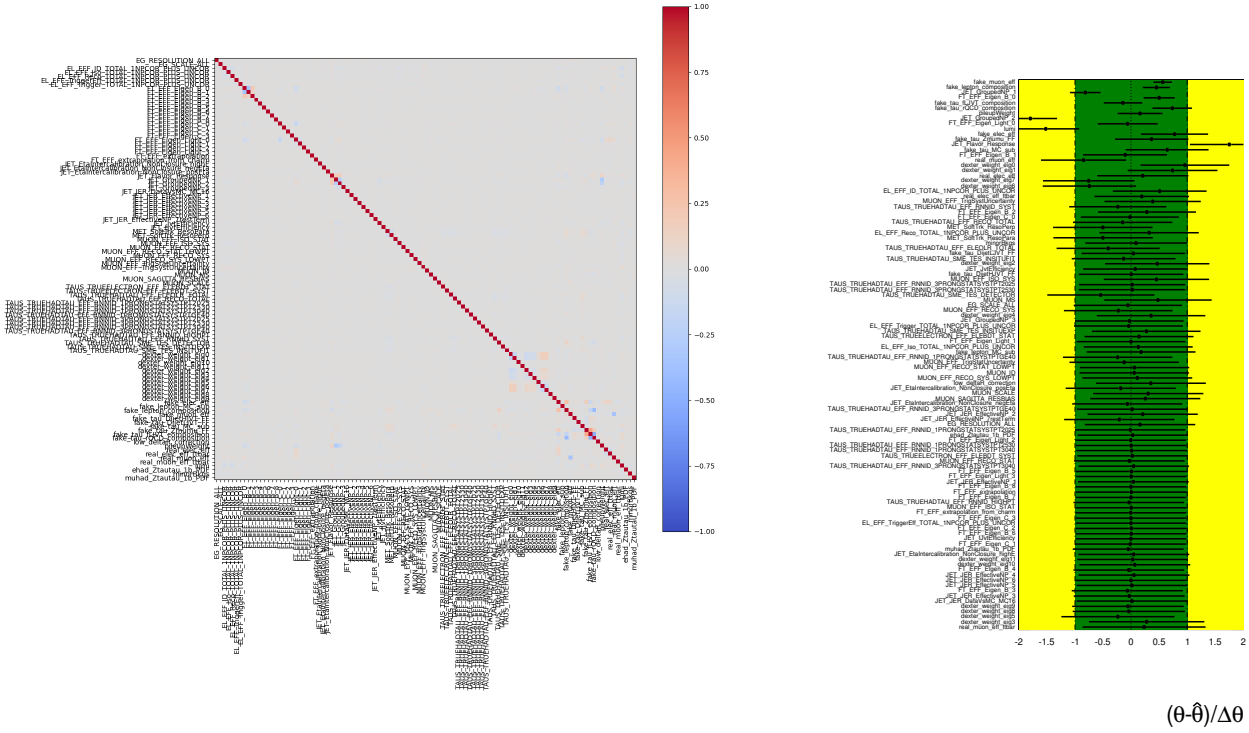


(a)



(b)

Figure 5.37: (a) Prefit and (b) postfit plots for the $m_a = 30$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.



(a) (b)

Figure 5.38: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 30$ GeV mass point.

CHAPTER VI

Results

Unblinded mass spectra for the 9 categories are shown in Figure 6.1. There is strong agreement between the expected and observed mass spectra for each channel and category. The single possible exception is in the $\mu\tau_{\text{had}}$, 2 b-tag category; however, in the following limit plots we will show that this excess is not significant. In large part the excess is not significant due to the lack of corresponding excesses of any kind in the other categories.

Next, expected and observed upper limits on the branching ratio $h \to aa \to bb\tau\tau$ are presented for each channel and category. The three heavy flavor categories (DeXTer, 1 b-jet, > 1 b-jet) are combined to create an expected limit in each channel ($e\mu$, $e\tau_{\text{had}}$, $\mu\tau_{\text{had}}$), shown in Figures 6.2–6.4. As expected, the DeXTer category drives the sensitivity in the $m_a < 20$ GeV region due to its ability to preserve signal acceptance for merged topologies and reject the SM background, in which merged b-jets are rare. In particular the DeXTer category in the $e\mu$ channel has high sensitivity because the angular separation cut between the leptons is $\Delta R > 0.1$, compared to $\Delta R > 0.2$ for $\ell\tau_{\text{had}}$ channels. Furthermore, the 2 b-jet category is the most sensitive category in the $m_a \gtrsim 50$ GeV region due to the higher likelihood of 2 resolved b-jets when $m_a \to m_h/2$. The 1 b-jet category is the most sensitive category in the intermediate mass regime. Finally, all nine categories are combined for the final expected limit, presented in Figure 6.5. The observed limits exclude any branching fraction above 1.5% for $h \to aa \to bb\tau\tau$ across the mass range [12, 60] GeV. No significant

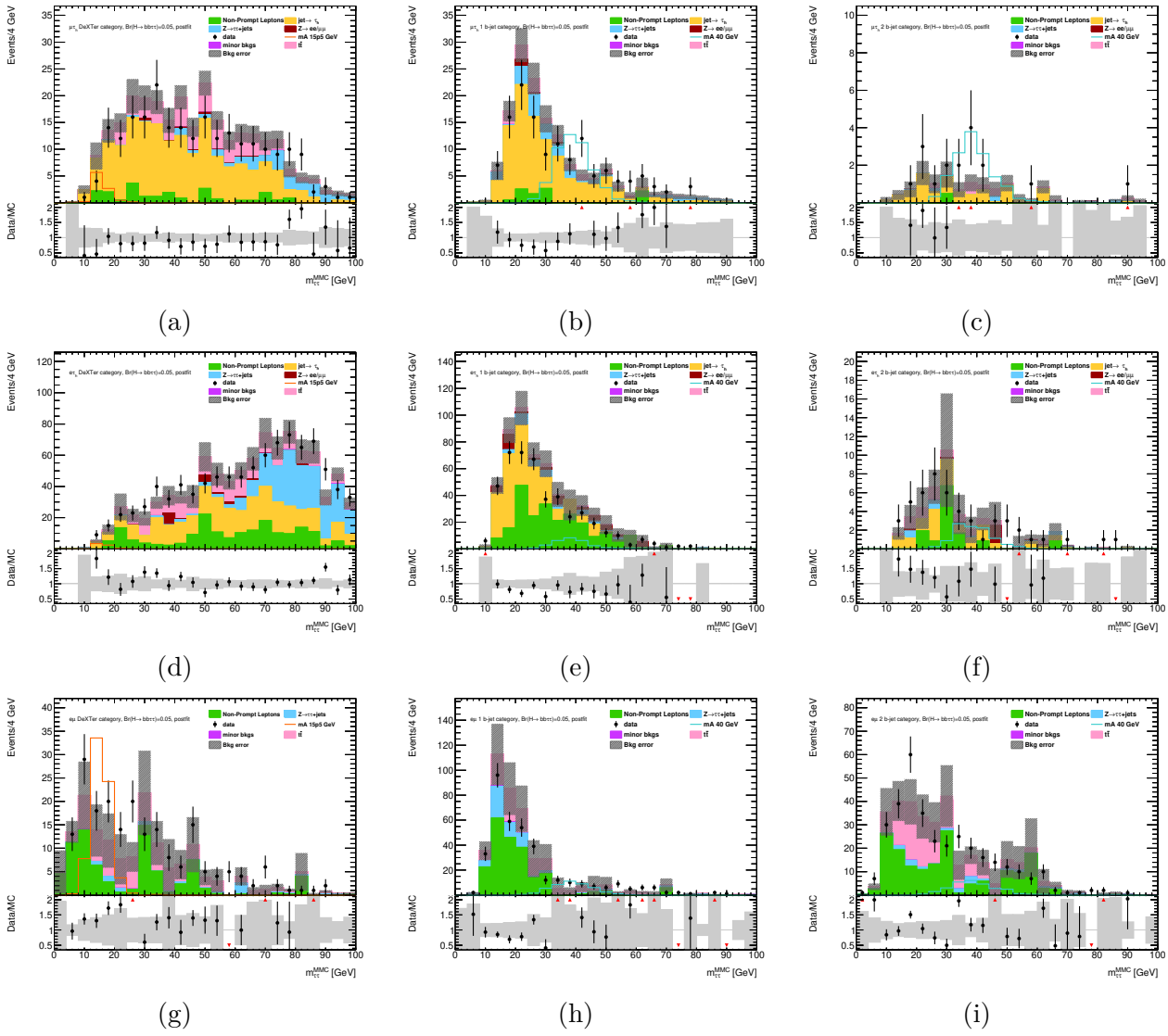


Figure 6.1: A comparison of the expected and observed mass spectra in each of the nine channels. The $\mu\tau_{\text{had}}$ channel is shown for (a) DeXTer, (b) 1 b-tag, (c) 2 b-tag categories. Next, the $e\tau_{\text{had}}$ channel is shown for (d) DeXTer, (e) 1 b-tag, (f) 2 b-tag categories. Finally, the $e\mu$ channel is shown for (g) DeXTer, (h) 1 b-tag and (i) 2 b-tag categories.

excess is observed. There is a 2 σ excess in only the $\mu\tau_{\text{had}}$ 2 b-tag category, but a lack of corresponding excesses in other sensitive channels means that in the combined limit in Figure 6.5 there is no significant excess near 40 GeV. The background only p-value of the largest upward flucuation is $p_b = 0.48$, which demonstrates that the background-only model describes the observed data without the need for a BSM particle.

The observed limit is also interpreted as an upper limit on the branching ratio $h \to aa$ in the four types of 2HDMs at various values of $\tan\beta$, shown in Figure 6.6. The Type-I case is simply the $\tan\beta = 1$ limit in Types II, III, or IV. The plots in Figure 6.6 saturate at a $h \to aa$ branching ratio of 20% because constraints from analysis of the Higgs couplings [28, 29] place limits on the order of 15% for the 125 GeV Higgs boson to decay to any exotic particles. Any phase space in which we set limits above 20% on exotic Higgs branching fractions are "uninteresting" for the $bb\tau\tau$ decay mode in the sense that this analysis does not produce a competitive limit. The reason for this is that these regions of phase space have a suppressed coupling of the a to either down-type quarks or leptons.

The limit in Figure 6.5 is the most sensitive limit on the $h \to aa \to bb\tau\tau$ process observed to date. A competing limit from the CMS collaboration [157] is less sensitive for three reasons. (i) The DeXTer tagger increases the sensitivity for this search in the low mass regime, (ii) the PNN described in this thesis is more sensitive across all channels than the NN used by CMS, and (iii) lower lepton p_{T} thresholds are employed in this analysis, increasing the signal acceptance. While backgrounds are also increased by the low p_{T} cut, the more sensitive PNN makes up for higher background counts. The limit presented in this thesis is 4.5 times more sensitive at low masses (DeXTer dominated region), 2 times more sensitive at intermediate masses (1 b-tag dominated region) and 3 times more sensitive at high masses (2 b-tag dominated region).

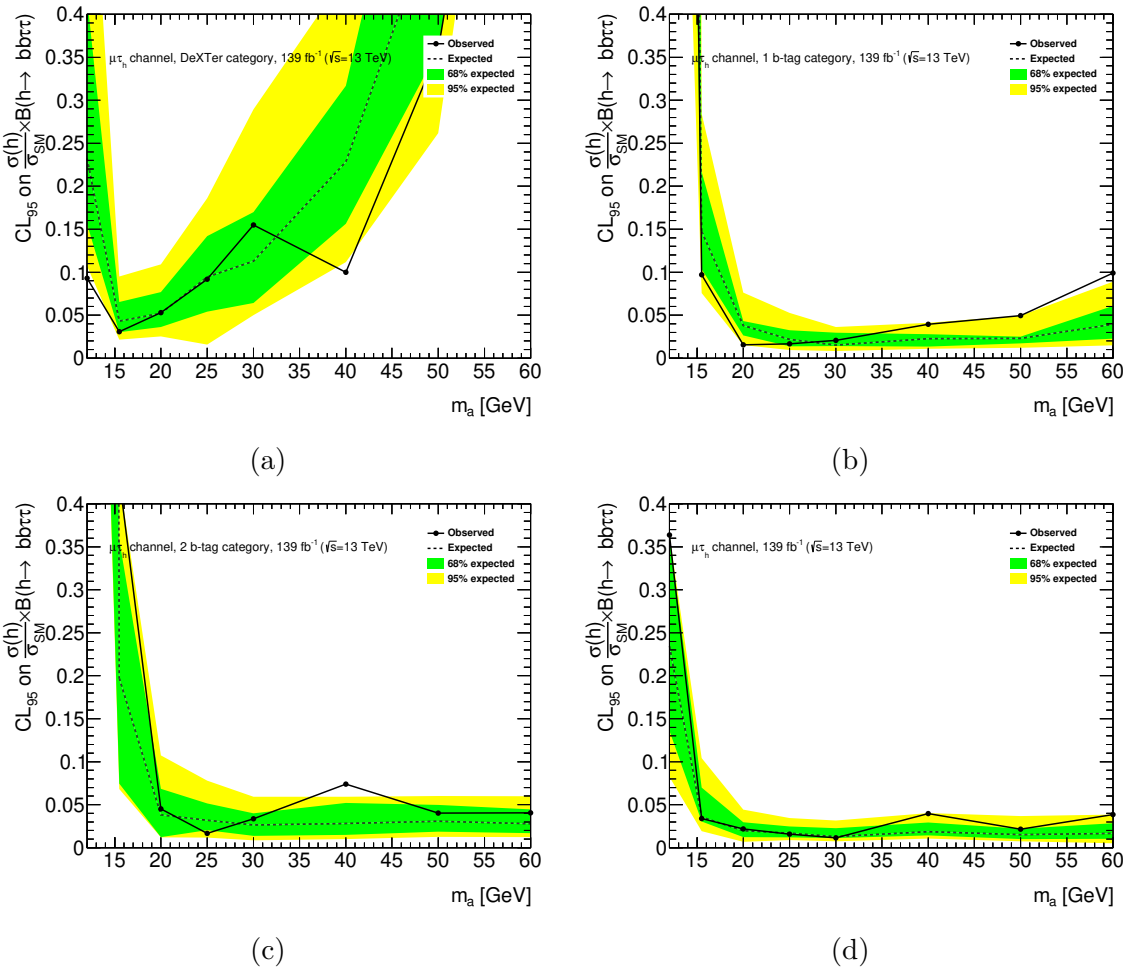


Figure 6.2: Expected limits in the $\mu\tau_{\text{had}}$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d).

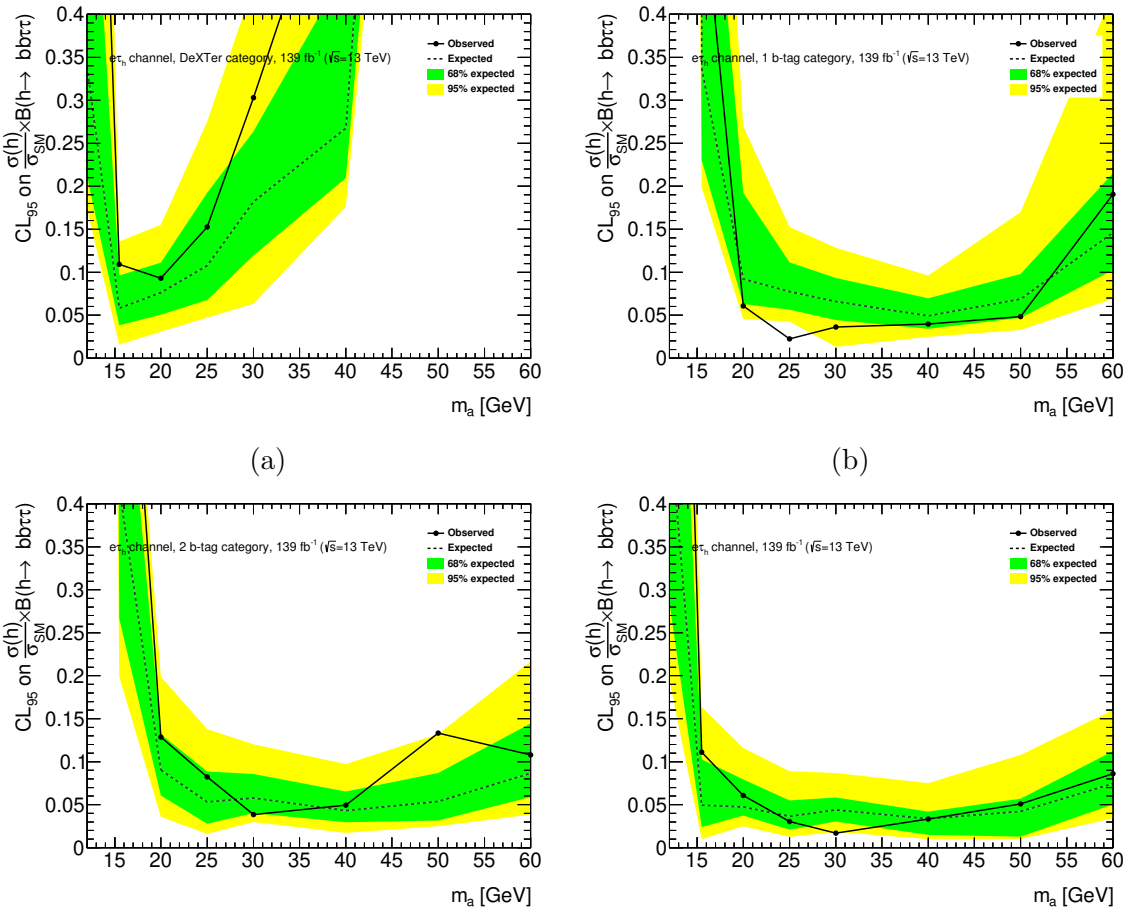


Figure 6.3: Expected limits in the $e\tau_{\text{had}}$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d).

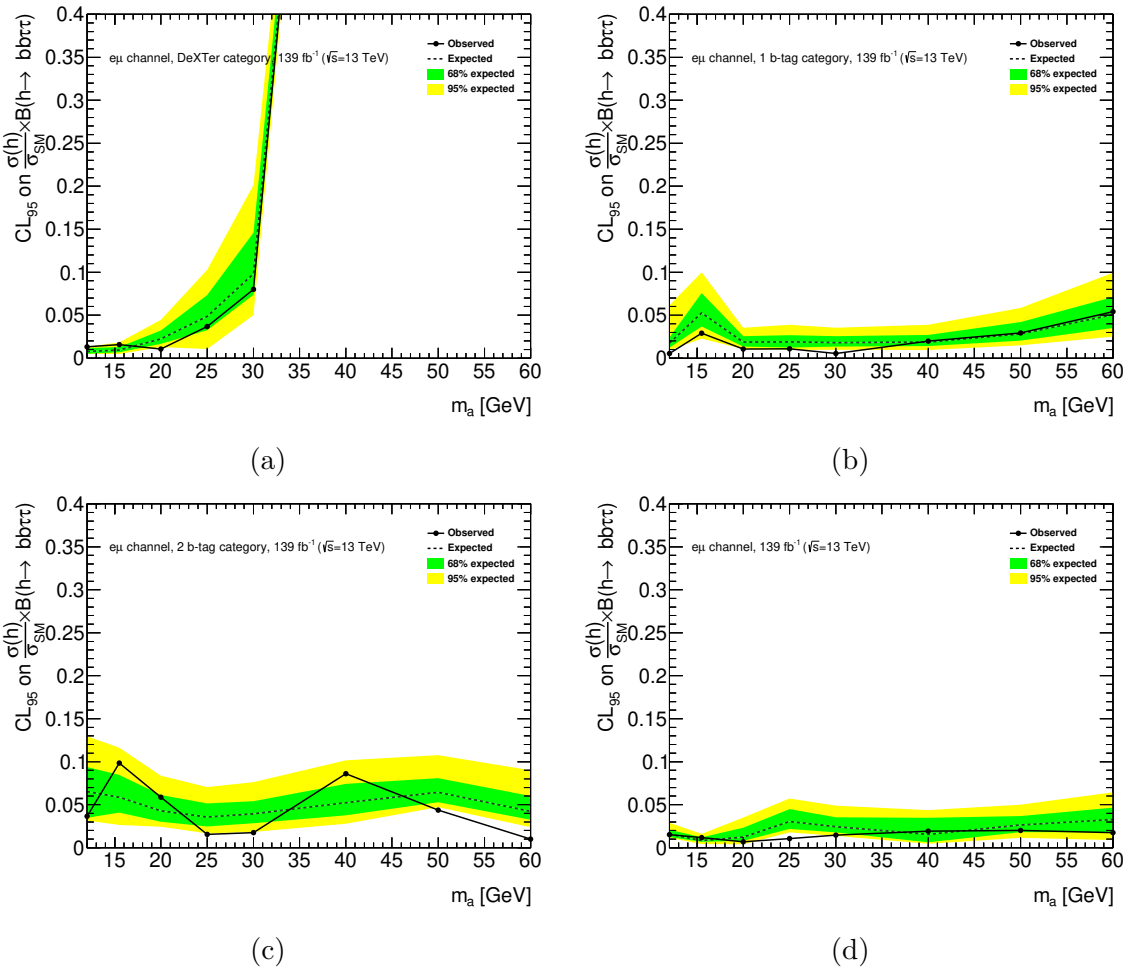


Figure 6.4: Expected limits in the $e\mu$ channel for the (a) DeXTer (b) 1 b-jet and (c) > 1 b-jet categories. The limits are combined in (d).

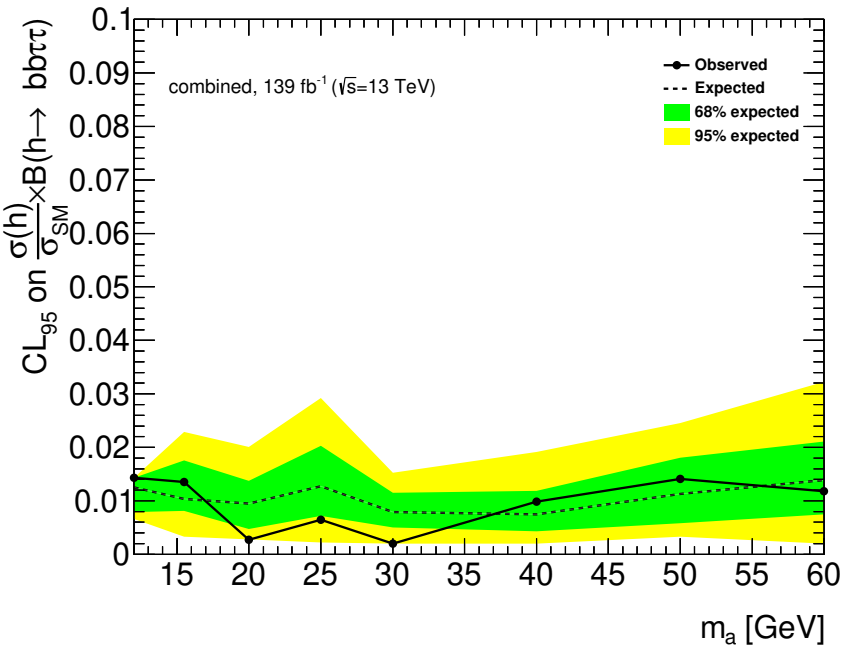


Figure 6.5: Combined expected limits across all channels and categories.

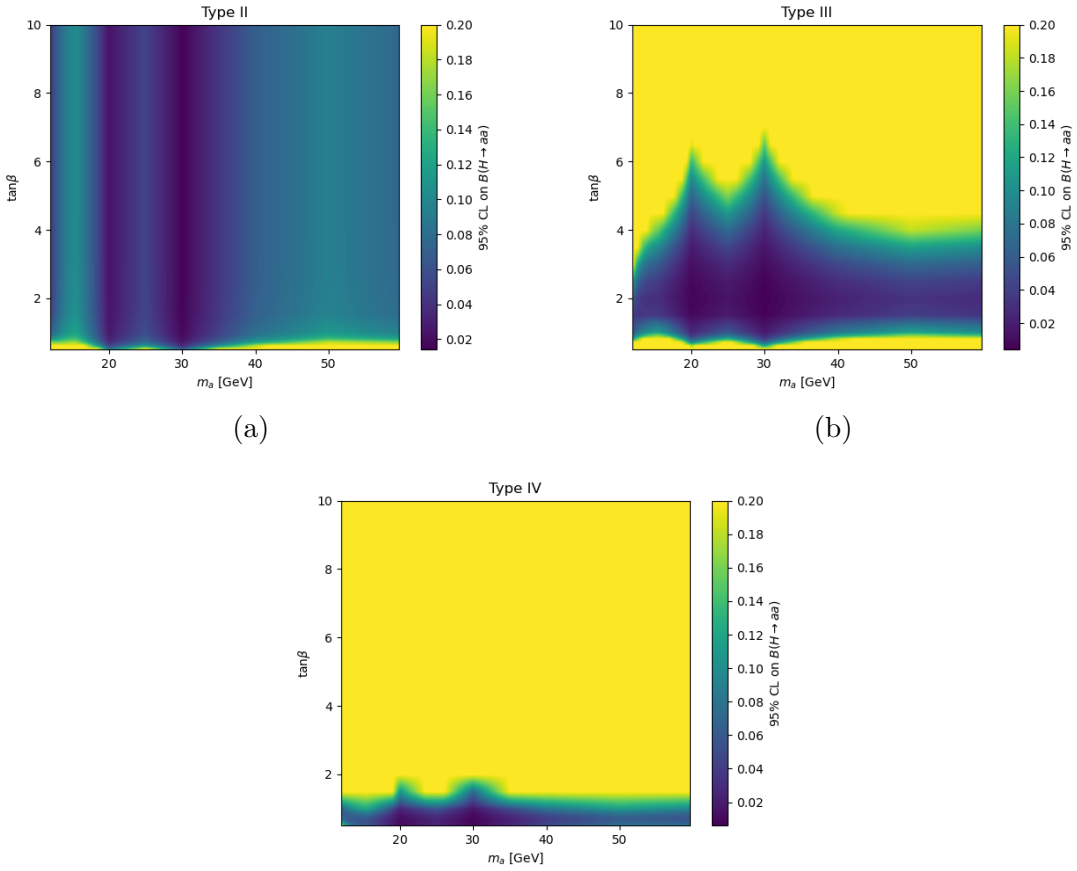


Figure 6.6: The observed limits are interpreted under each of the types of 2HDMs. Each of Types II, III, and IV reduce to a Type-I case when $\tan\beta$=1. The plots saturate at a branching fraction of 20% because at a branching fraction this large, analyses of the Higgs Yukawa couplings constrain the 2HDM more than this search.

CHAPTER VII

Conclusions

This thesis presents a search for decays of the SM-like Higgs boson to the $bb\tau\tau$ final state via a BSM psuedoscalar a in p-p collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. The search tests the Standard Model of particle physics against the 2HDM+S model, which adds a light psuedoscalar a and a second Higgs doublet. The a particle can decay to SM particles via its mixing with the extended scalar sector, and the addition of a second Higgs doublet modulates the coupling of the a to leptons and up- and down- type quarks via the parameter $\tan\beta = v_2/v_1$. Crucially, the model preserves the SM Higgs couplings and does not predict tree-level FCNCs. The 2HDM+S model would produce a peak in the $\tau\tau$ mass at m_a, the mass of the new psuedoscalar, which is not predicted by the Standard Model of particle physics.

The practical methodology of the search is to construct a model which, in conjunction with observed event counts in background-enriched control regions, can predict the SM $\tau\tau$ mass spectrum in an orthogonal, signal-enriched region. The model accounts for all known SM processes by pairing a data-driven and MC estimation of event counts. The possible effects of systematic uncertainties are accounted for by way of constrained Nuisance Parameters, which add flexibility to cover all reasonable variations in detector response, SM cross sections, and experimental technique. By comparing the SM-only predicted $\tau\tau$ mass spectrum to the observed mass spectrum we set limits on the branching fraction $h \to aa \to$

$bb\tau\tau$.

The observed results show no significant deviation from the SM. Observed limits constrain the branching fraction of $h \to aa \to bb\tau\tau$ to be less than 1.5% in the mass range from 12 to 60 GeV. The most significant upwards fluctuation compared to the SM is $< 1\sigma$ from the predicted upper limit, with a corresponding background-only p-value of $p_b = 0.48$. The observed limits in this search exclude the $> 3\sigma$ local excess observed by the ATLAS collaboration [21] in the $bb\mu\mu$ decay channel as resulting from $h \to aa$ and place stronger limits than the recent CMS search in the $bb\tau\tau$ decay mode [157]. The excess in $bb\mu\mu$ is not dead, however. This result excludes (in the context of the 3.3σ $bb\mu\mu$ excess) a mediator which couples to SM particles proportional to their masses, as is present in the 2HDM. If the exotic Higgs decay is mediated by a Z' boson which couples democratically to leptons, rather than $m_\tau/m_\mu \approx 17$ times more strongly to taus, then the $bb\mu\mu$ could still be a first sign of a new particle. Of course, the excess can also be a statistical fluctuation and further study is needed.

APPENDICES

APPENDIX A

Additional Validation Plots

A.1 Additional Plots for Fake Lepton and Taus

This section presents additional supporting plots for the fake lepton and tau estimation. Fake lepton efficiencies in the region with 0 b-tags are presented in Figures A.1–A.2. Figures A.3–A.7 shows each individual fit to the reconstructed tau's seed jet width distribution, which is used to fit the fraction of fake taus described by the low JVT FFs, f_{LJVT}. Figures A.8–A.10 show the combined FF for the 0 b-tag and SR selection, binned by decay mode, p_{T}, and $|\eta|$ in both of the $\ell\tau_{\mathrm{had}}$ channels. The FFs for the $\mu\tau_{\mathrm{had}}$ SR selection are shown in the main body of the text in Figure 5.32.

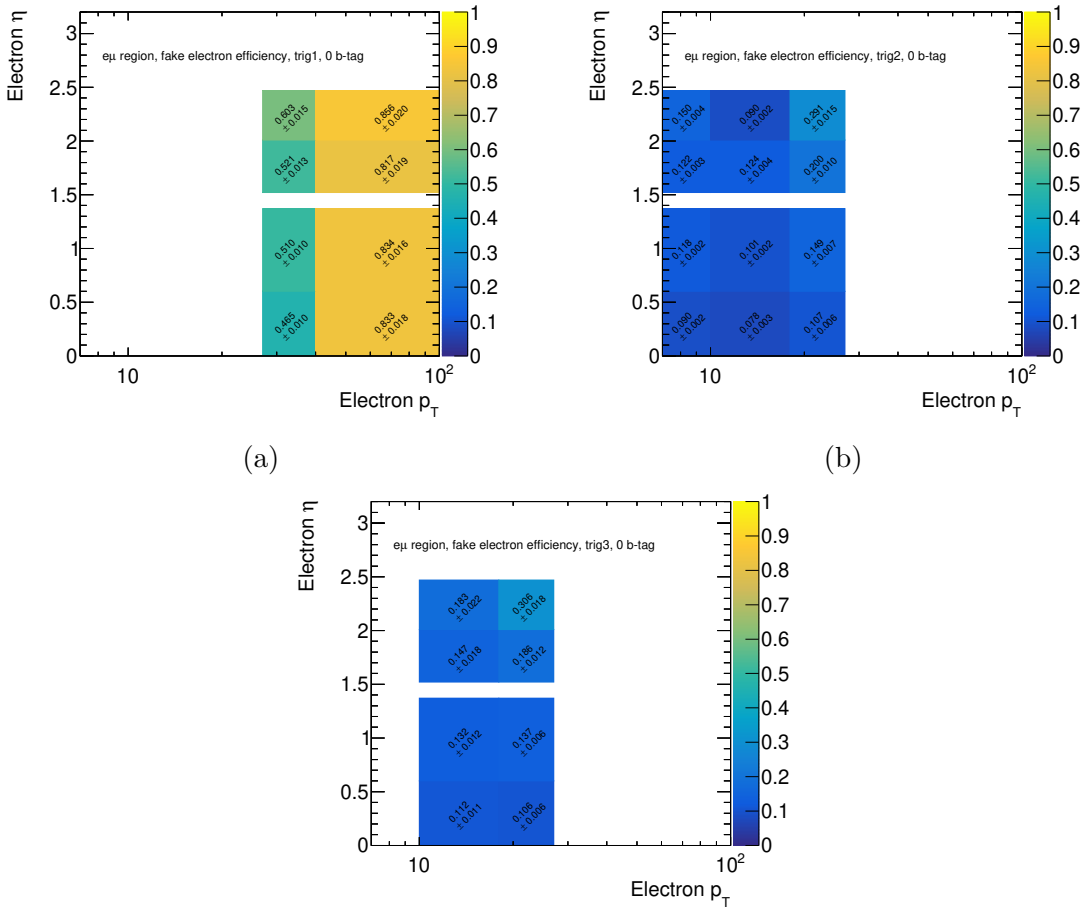


Figure A.1: The observed fake electron efficiencies in the region containing 0 b-jets. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3.

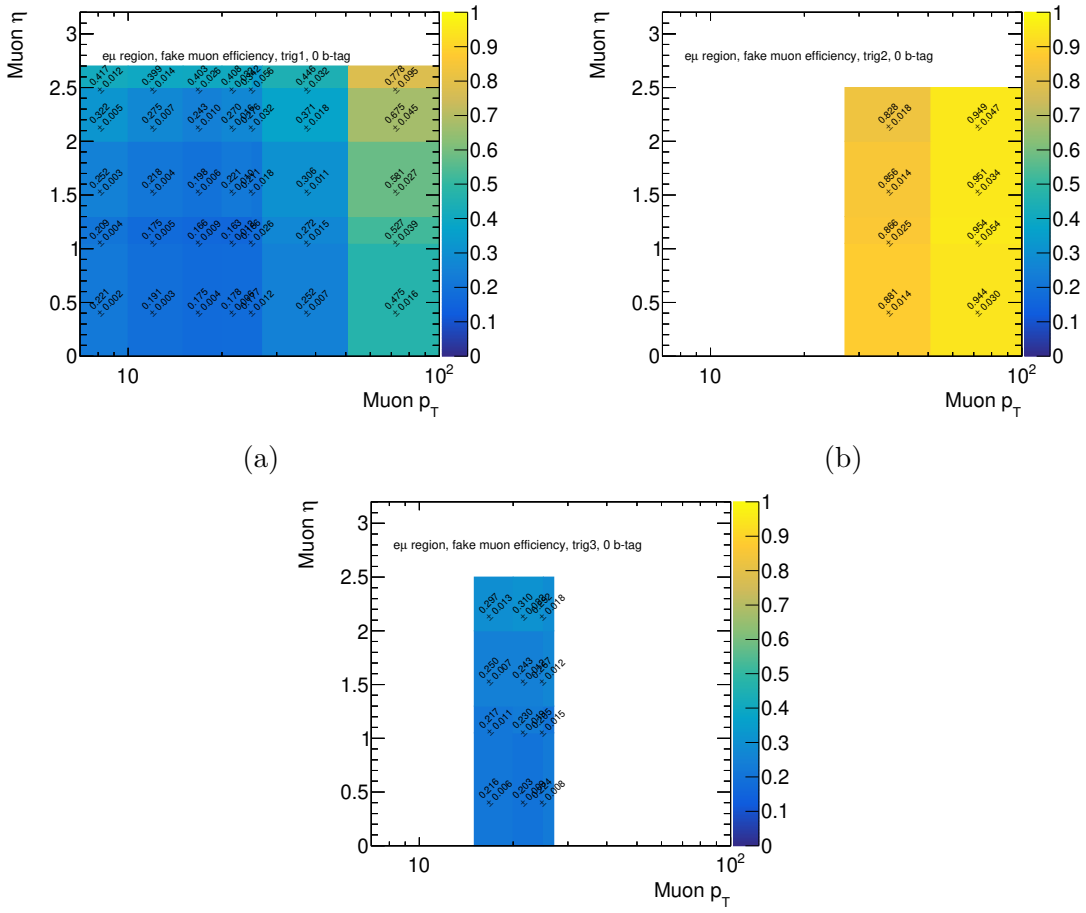


Figure A.2: The observed fake muon efficiencies in the region containing 0 b-jets. The three plots show the three trigger regions: (a) trig1, (b) trig2, and (c) trig3, as shown in Figure 5.3.

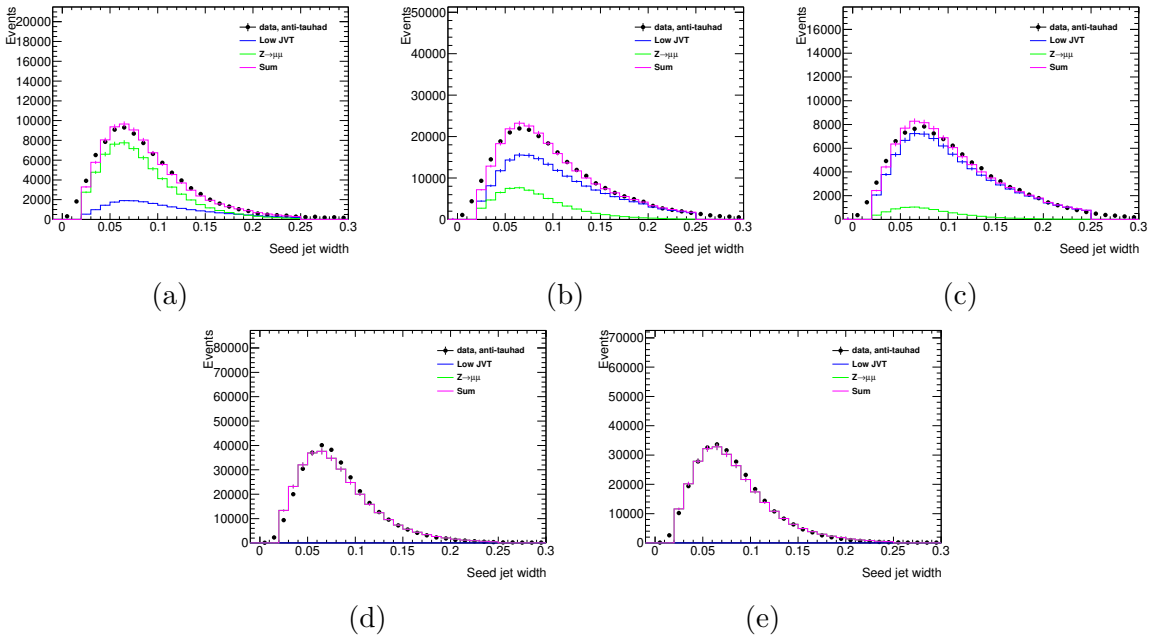


Figure A.3: Seed jet width fits in the $e\tau_{\text{had}}$ channel for the inclusive SR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

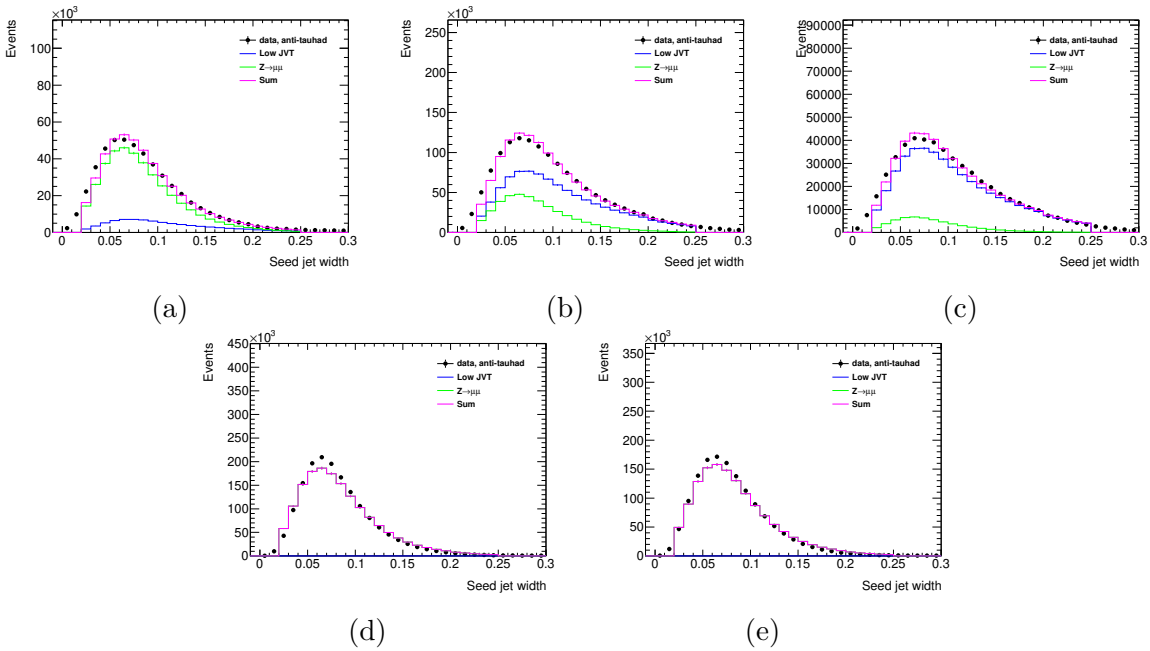


Figure A.4: Seed jet width fits in the $e\tau_{\text{had}}$ channel for the inclusive 0 b-tag VR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

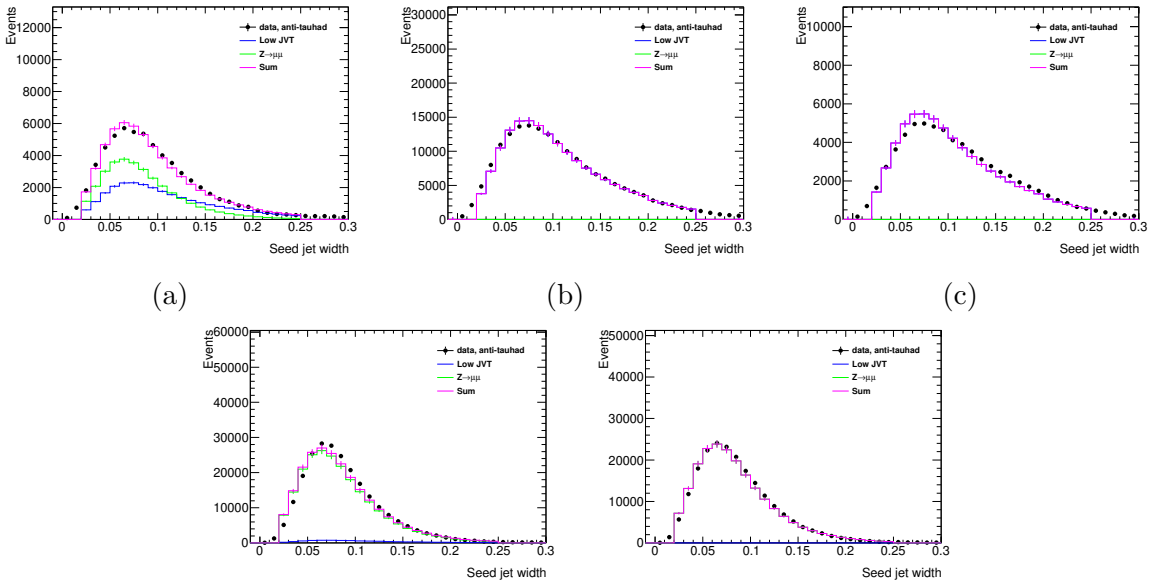


Figure A.5: Seed jet width fits in the $e\tau_{\text{had}}$ channel for the SS CR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

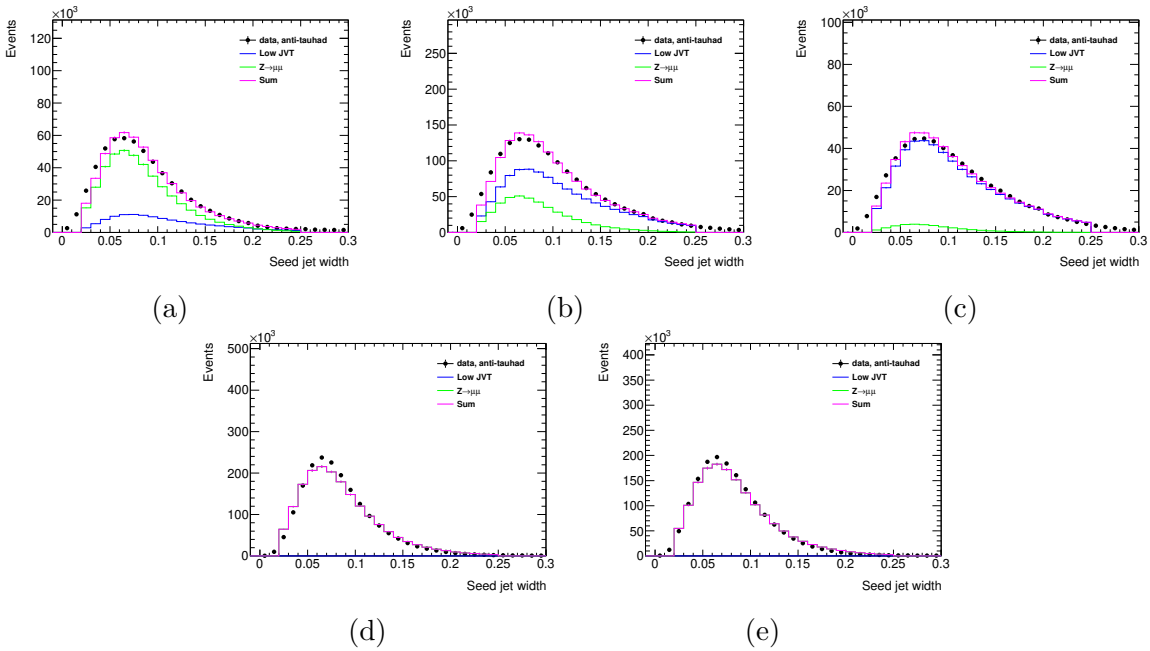


Figure A.6: Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the inclusive 0 b-tag VR selection. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

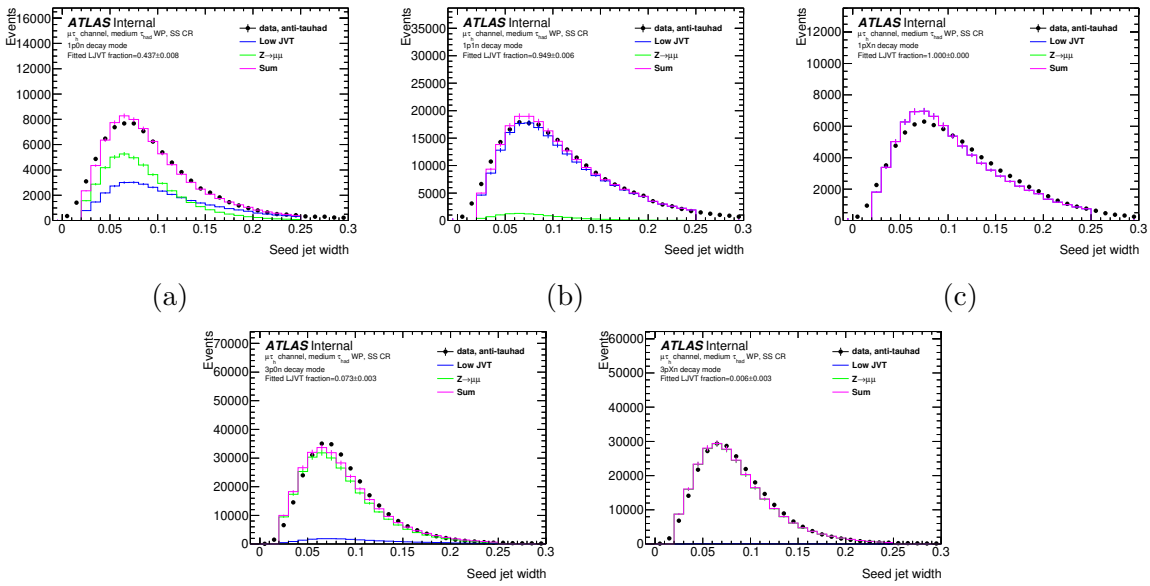


Figure A.7: Seed jet width fits in the $\mu\tau_{\text{had}}$ channel for the SS CR selections. The five tau decay modes are shown separately: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

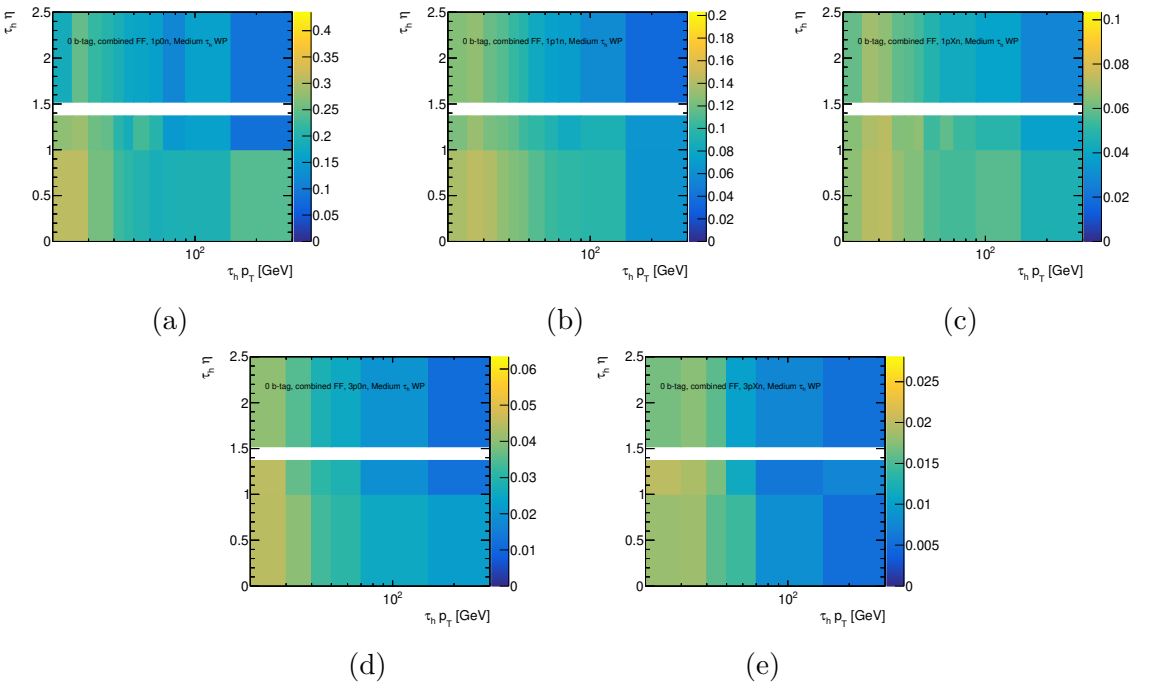
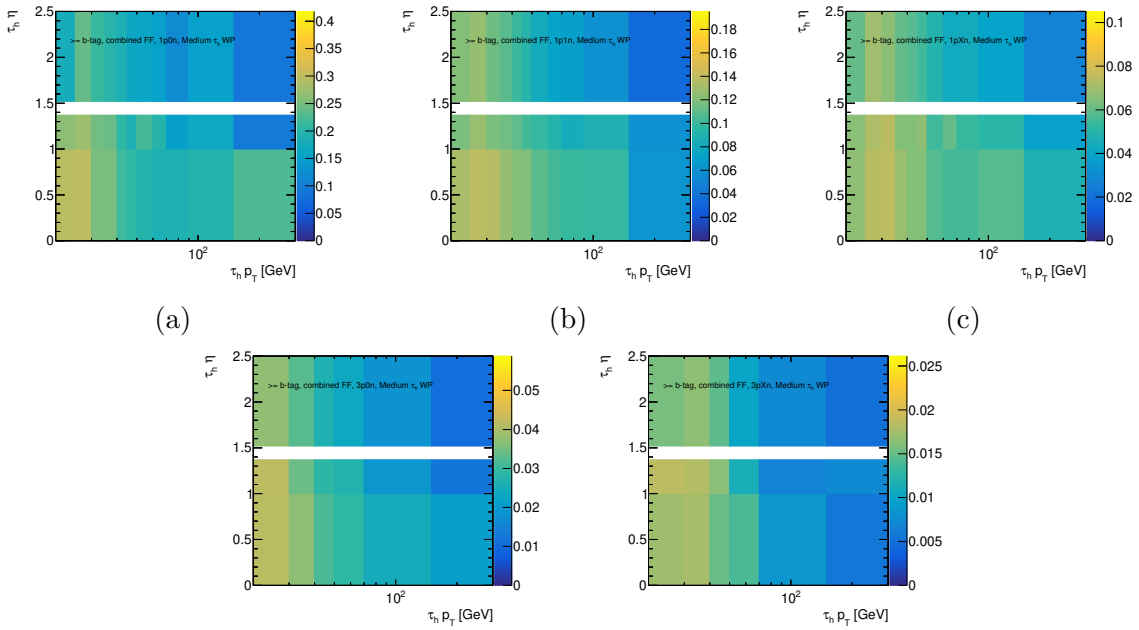


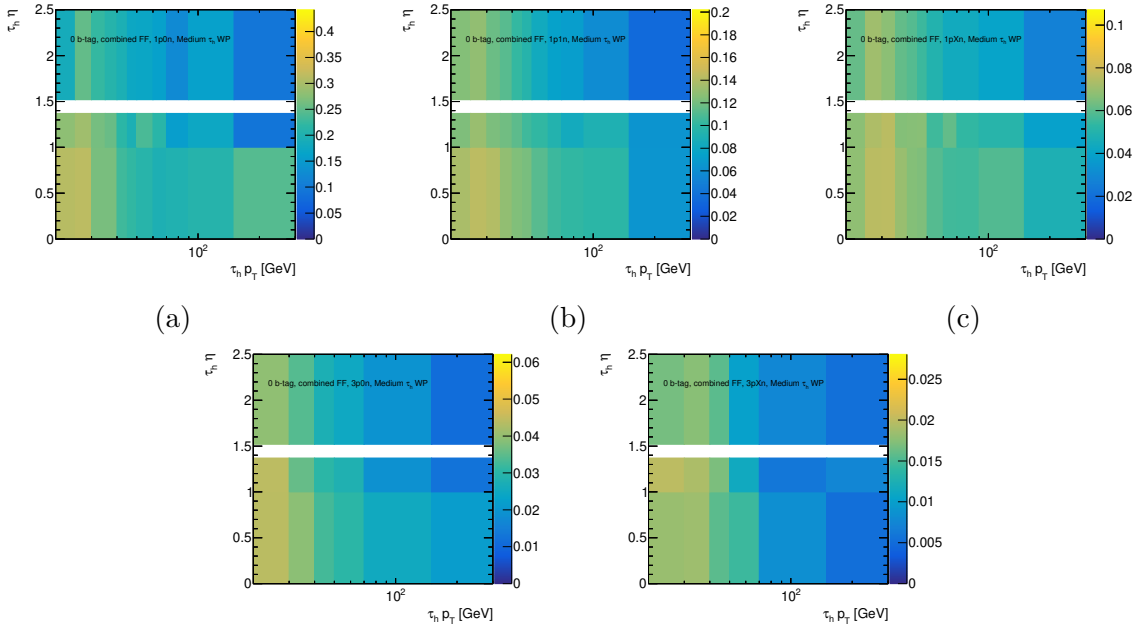
Figure A.8: Combined FFs for the $\mu\tau_{\text{had}}$ channel for the 0 b-tag VR selections. Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.



(a) (b) (c)

(d) (e)

Figure A.9: Combined FFs for the $e\tau_{\text{had}}$ channel for the SR selection (>0 b-tag). Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.



(a) (b) (c)

(d) (e)

Figure A.10: Combined FFs for the $e\tau_{\text{had}}$ channel for the 0 b-tag VR selections. Each plot shows the FF for a different tau decay mode: (a) 1p0n (b) 1p1n (c) 1pXn (d) 3p0n (e) 3pXn.

A.2 Control Region Validation Studies

The 0 b-tag region validates the extrapolation of the background model to low $m_{\tau\tau}^{\mathrm{MMC}}$, the variable which is used for limit setting and statistical interpretation. In particular, the 0 b-tag region is more pure in fake leptons and taus, as it does not have the additional requirement of a heavy flavor tagged jet. This validation region also has a very low amount of signal pollution, preserving the analysis blinding strategy.

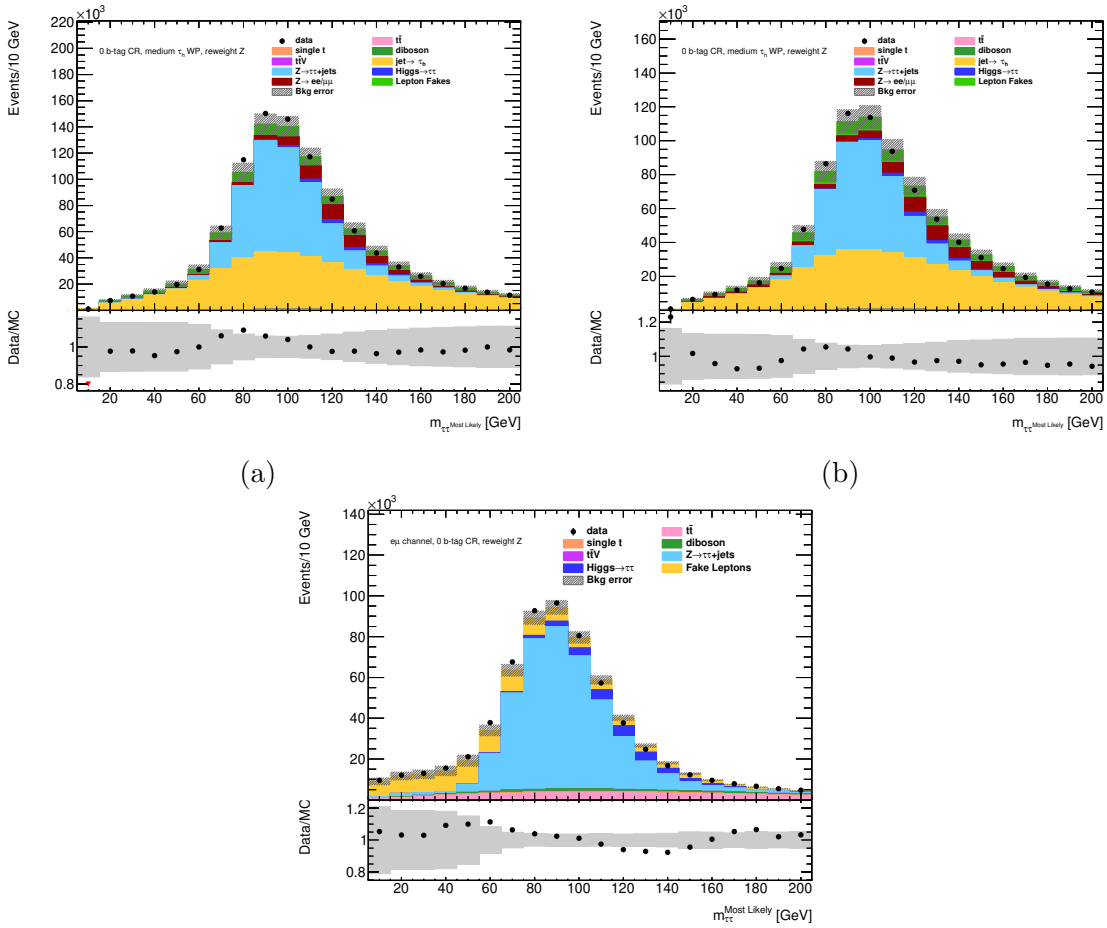


Figure A.11: Validation plots in the 0 b-tag VR for the (a) $\mu\tau_{\mathrm{had}}$ (b) $e\tau_{\mathrm{had}}$ and (c) $e\mu$ channels.

Next, additional validation plots for each channel in the high mass CR, 0 b-tag VR, and SS CR are presented. Inclusive plots in the high mass CR (Figures A.12–A.14), 0 b-tag (Figures A.15–A.16), and SS CR (Figures A.17–A.19) demonstrate the modeling in the

1 and 2 b-jet categories. The DeXTer category is a small ($\approx 2\%$) portion of the inclusive heavy flavor selection and the modeling is explicitly checked by separating out this sub-region. Figures A.20–A.22 shows the high mass CR with the DeXTer selection and Figures A.23–A.25 shows the SS CR with the DeXTer selection. In all plots a 25% error is used on fakes (two 15% systematic uncertainties dominate the fake uncertainty, so 25% is an approximate size of the error to guide the eye) and statistical error on MC and data-driven backgrounds is also included. Error from reweighting $t\bar{t}$ and Z+jets backgrounds (for example, the values in Figure 5.12) is additionally propagated to the corresponding backgrounds.

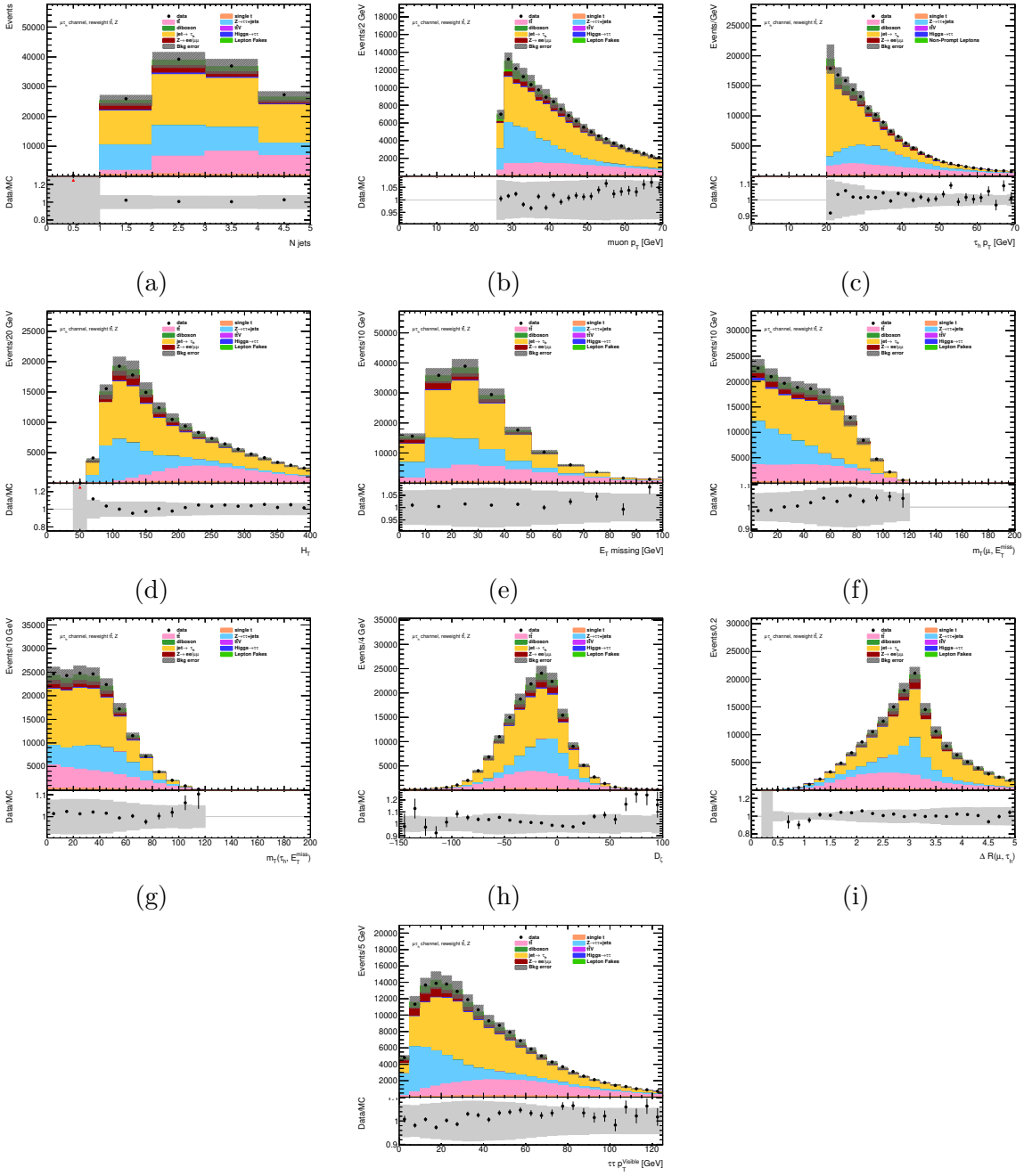


Figure A.12: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} > 60$ GeV) CR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

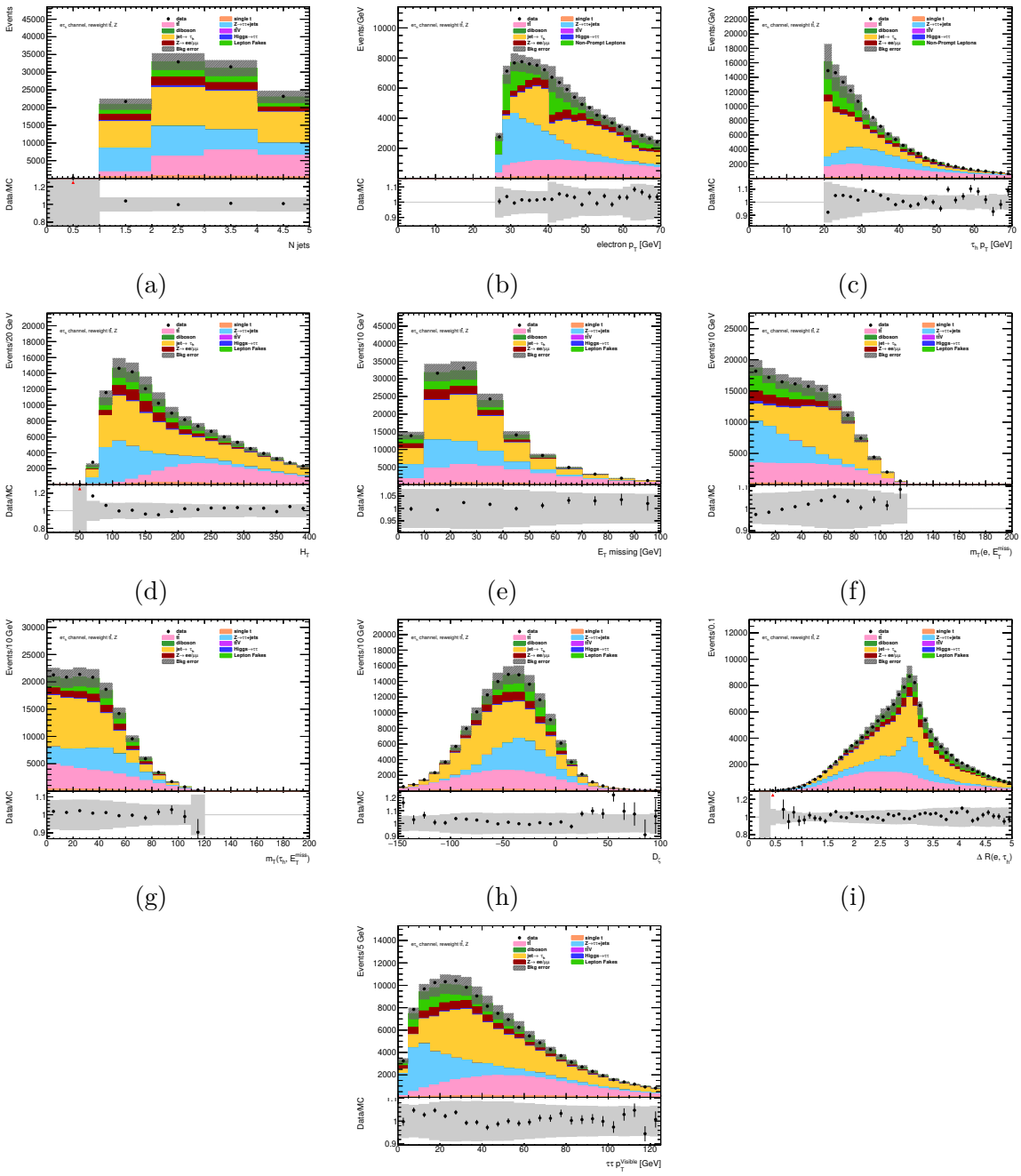


Figure A.13: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} > 60$ GeV) CR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

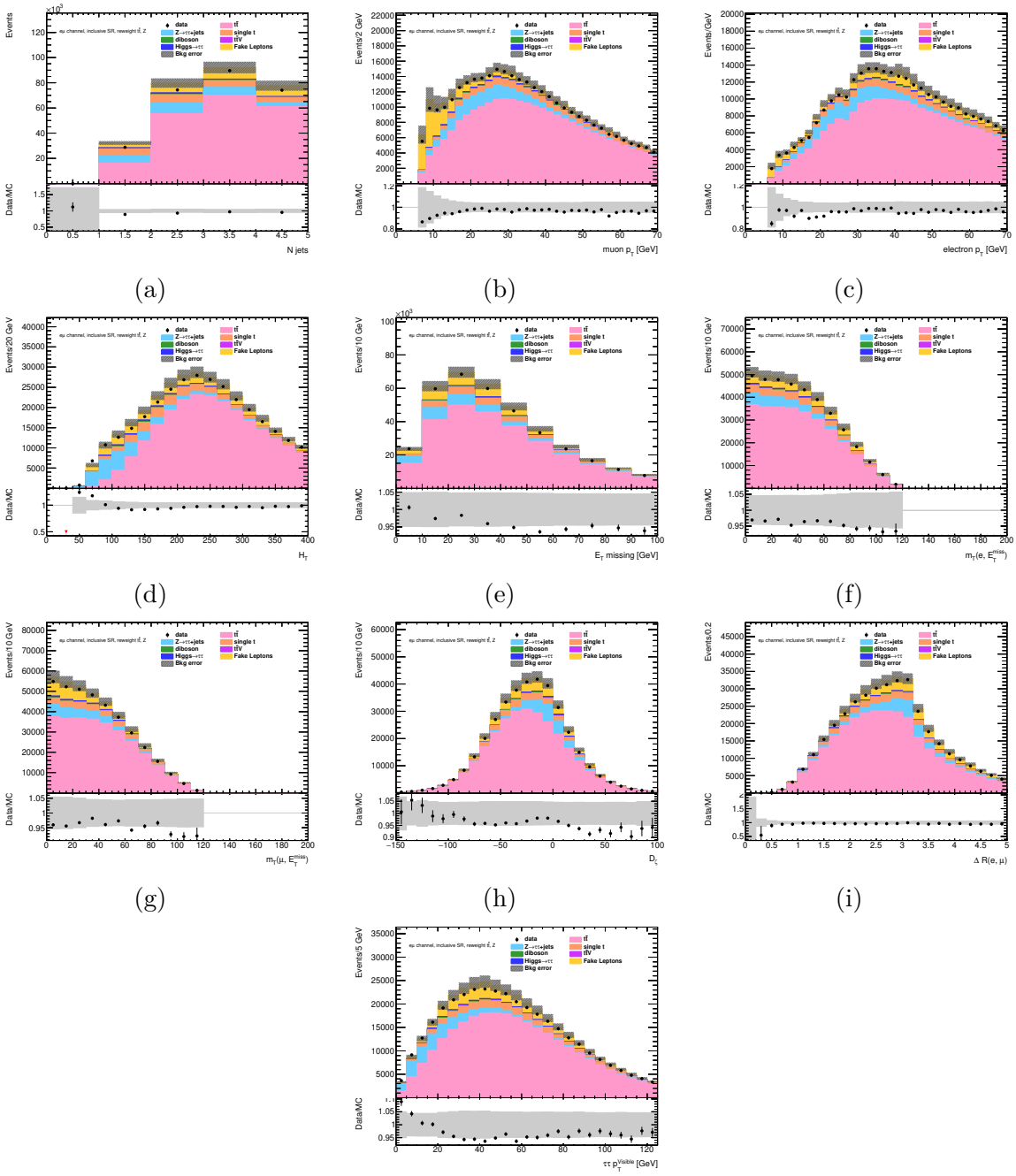


Figure A.14: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} > 45$ GeV) CR for the $e\mu$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

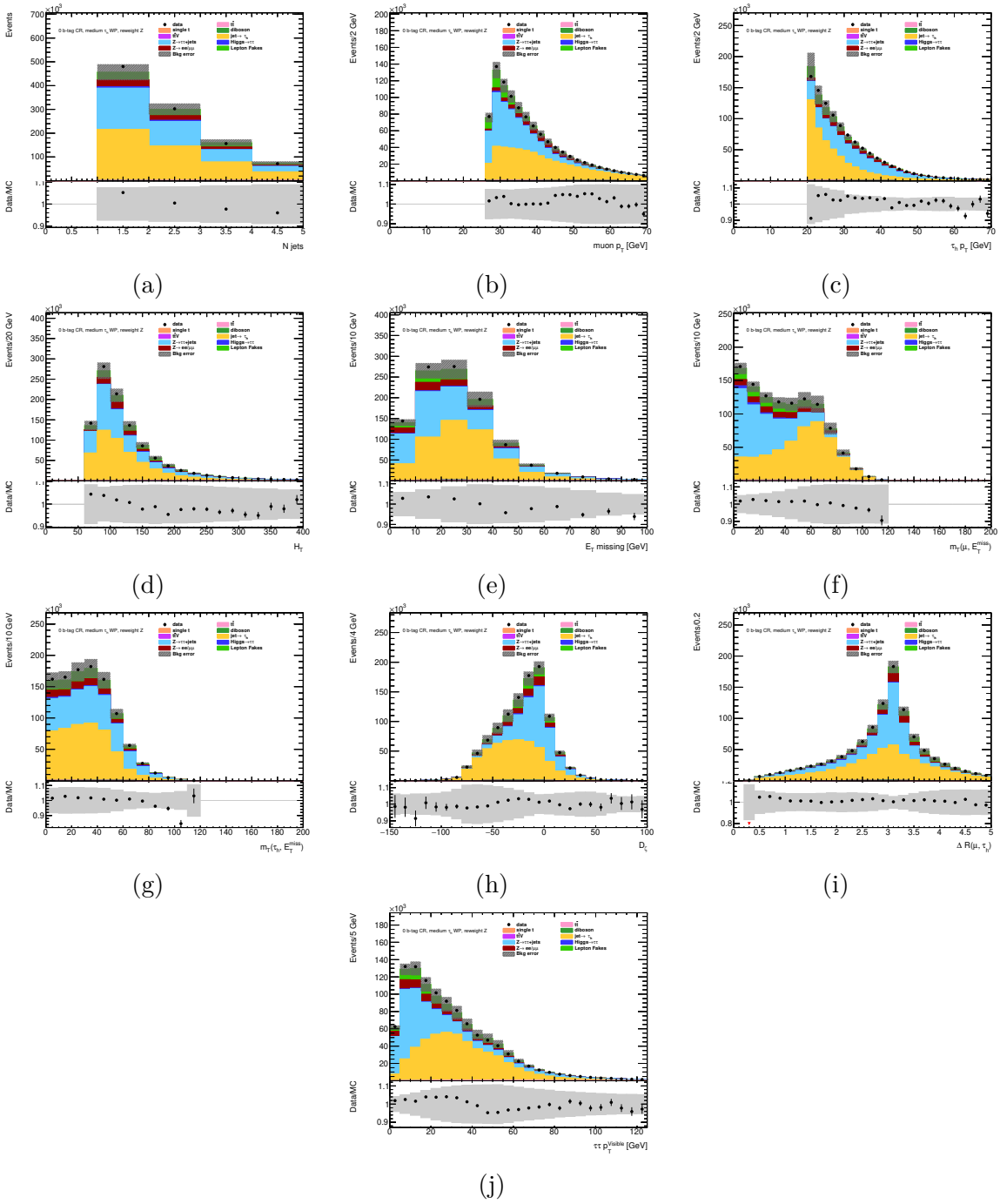


Figure A.15: Additional plots validating the kinematic modeling in the 0 b-tag VR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

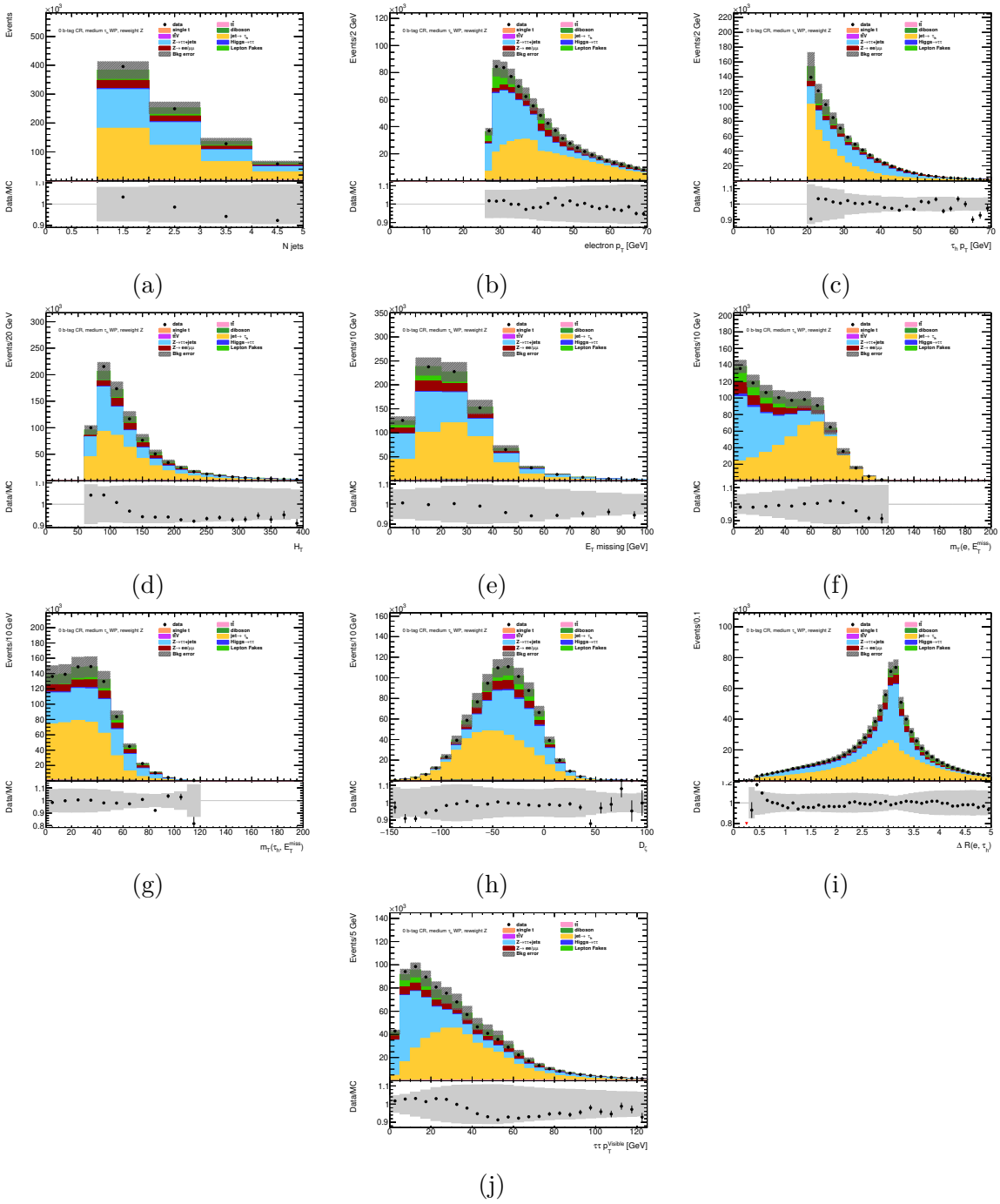


Figure A.16: Additional plots validating the kinematic modeling in the 0 b-tag VR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

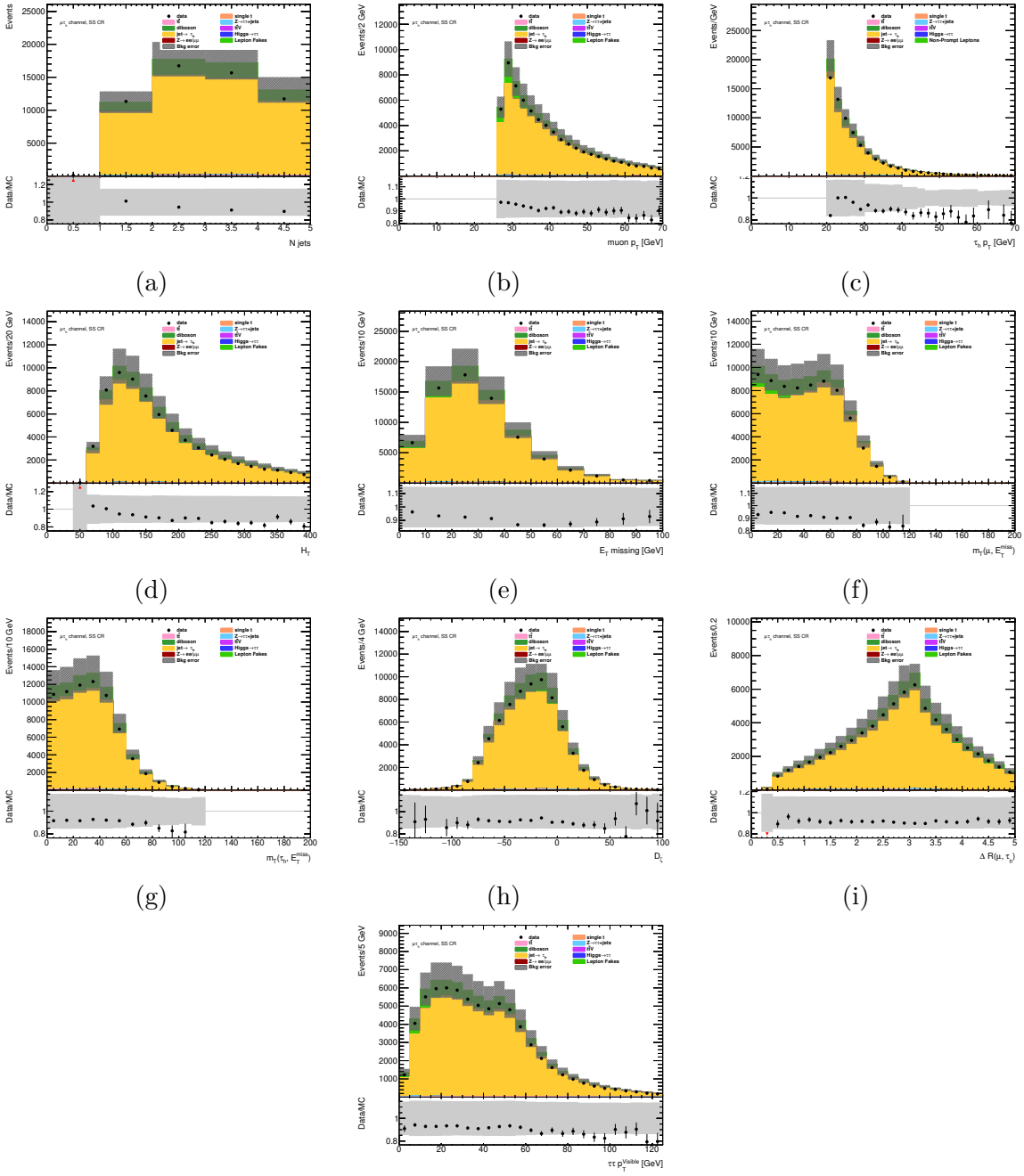


Figure A.17: Additional plots validating the kinematic modeling in the SS CR for the $\mu\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

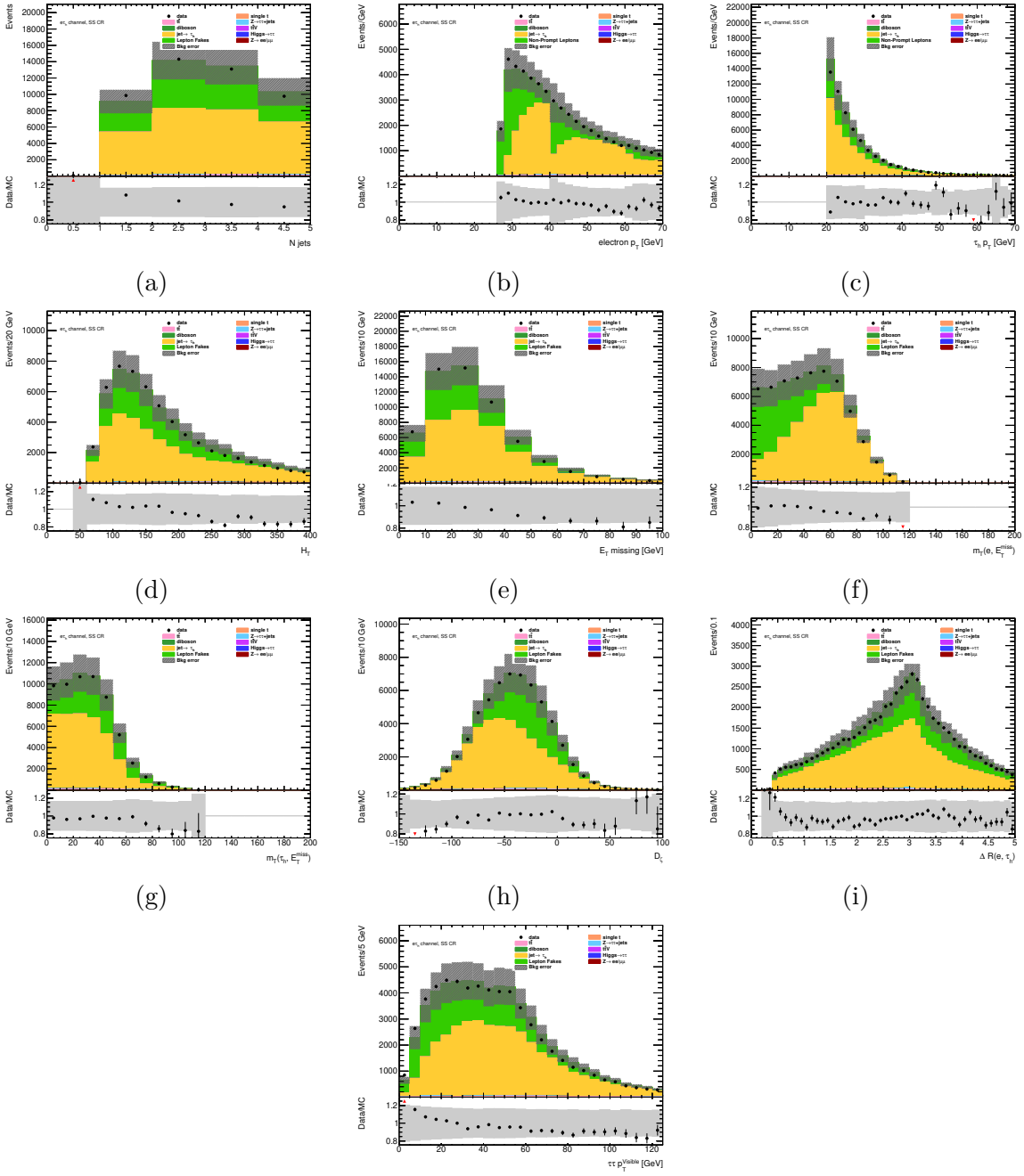


Figure A.18: Additional plots validating the kinematic modeling in the SS CR for the $e\tau_{\text{had}}$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

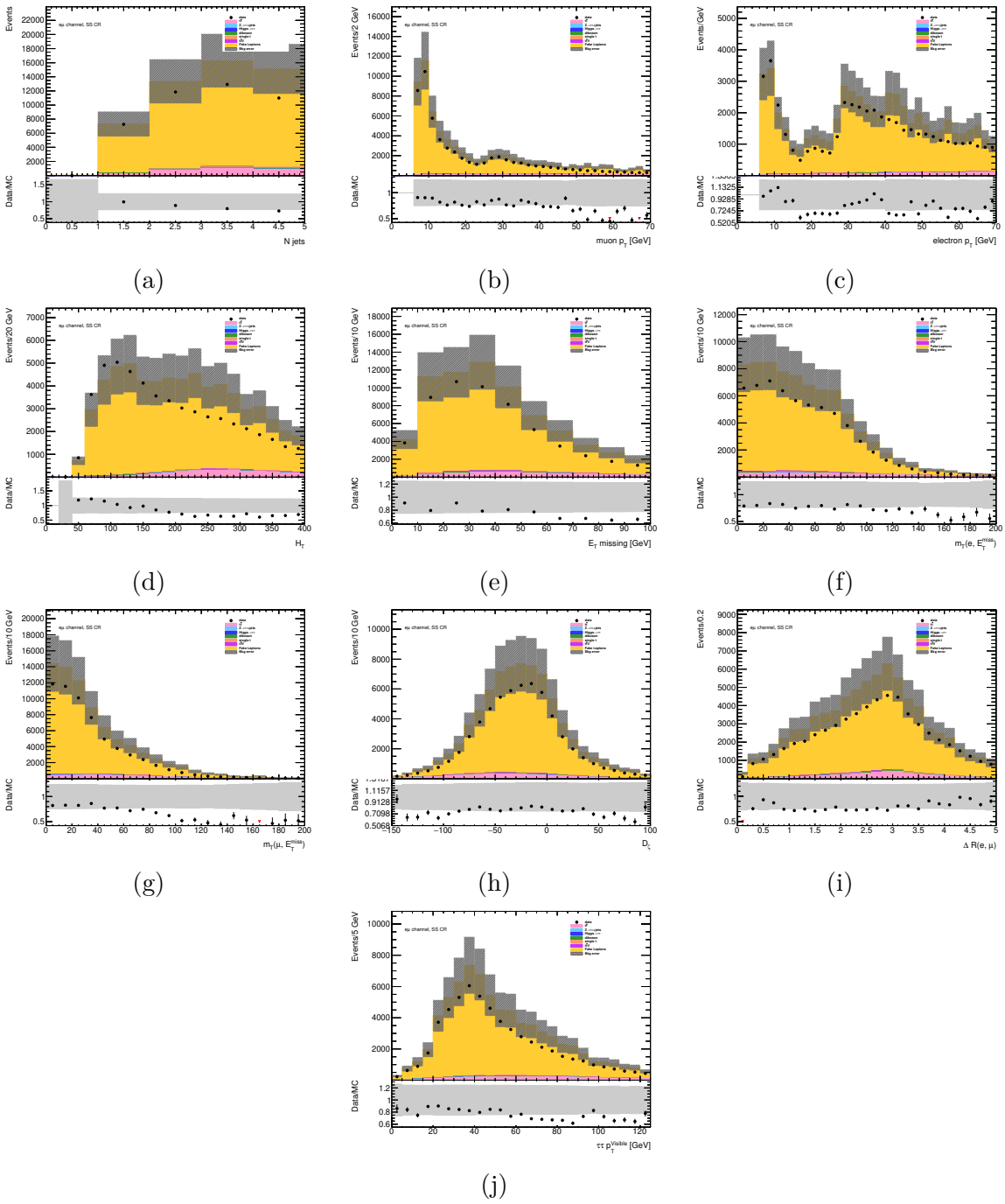


Figure A.19: Additional plots validating the kinematic modeling in the SS CR for the $e\mu$ channel. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

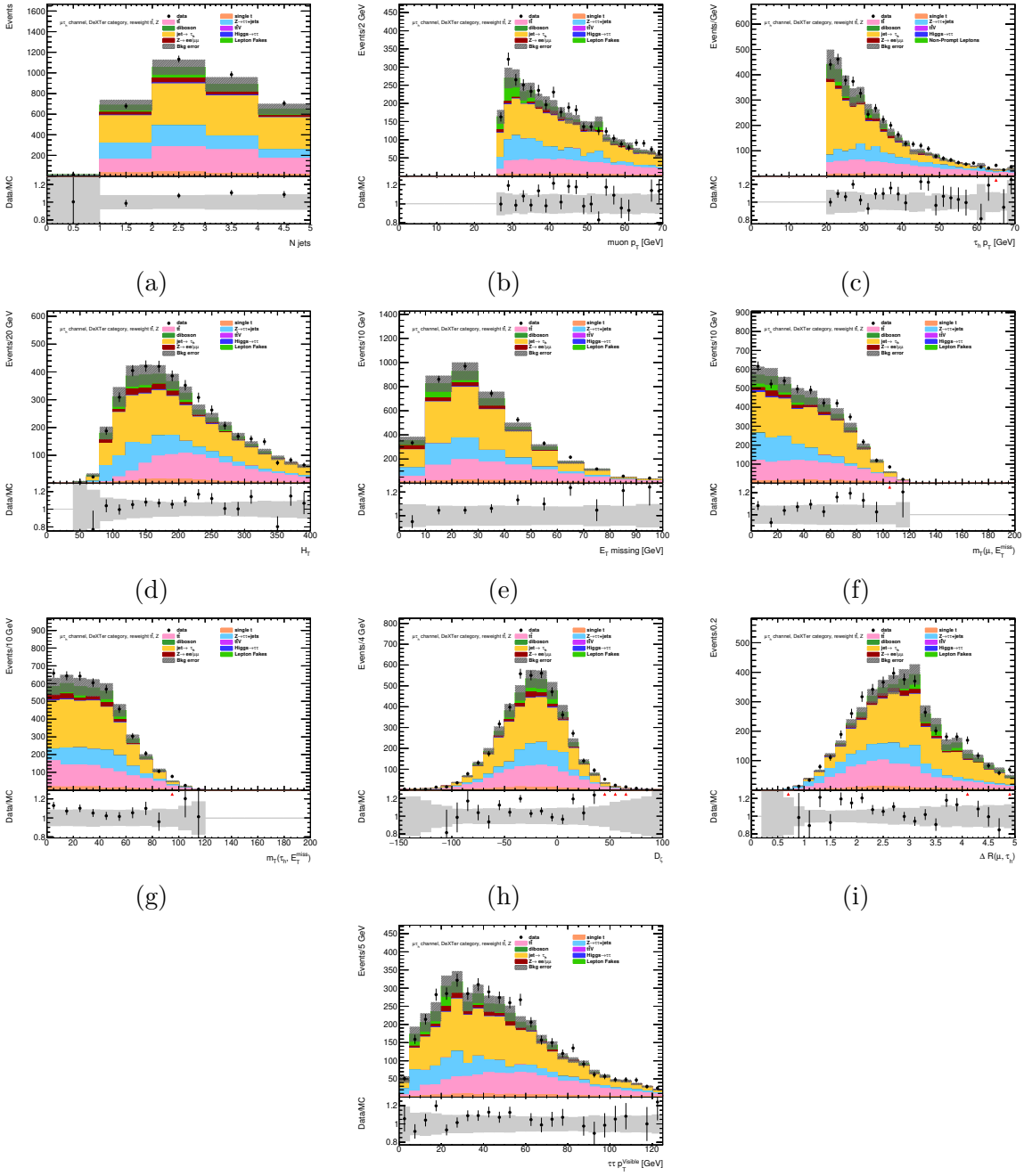


Figure A.20: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} > 60\text{GeV}$) CR for the $\mu\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

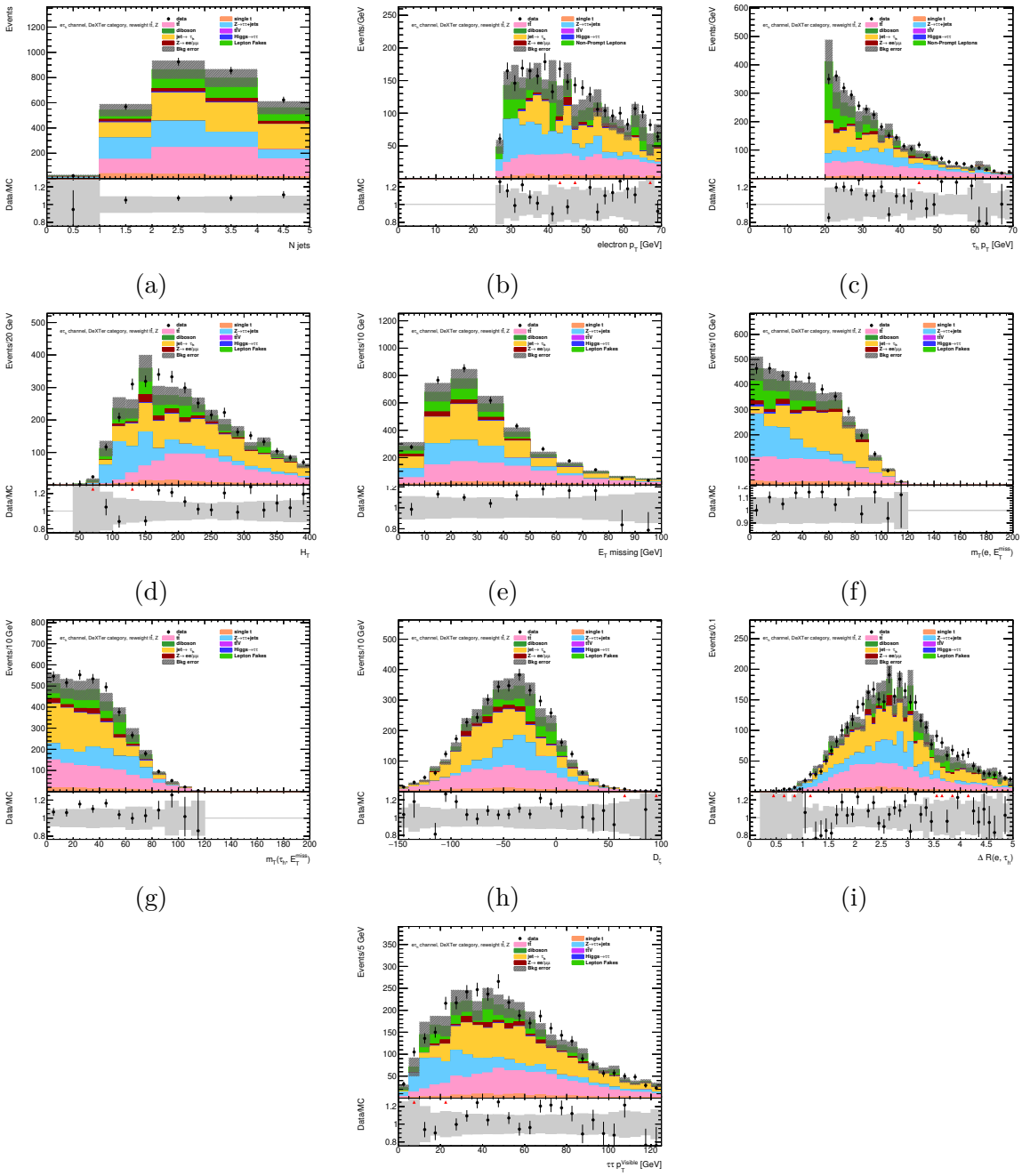


Figure A.21: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} >$ 60GeV) CR for the $e\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

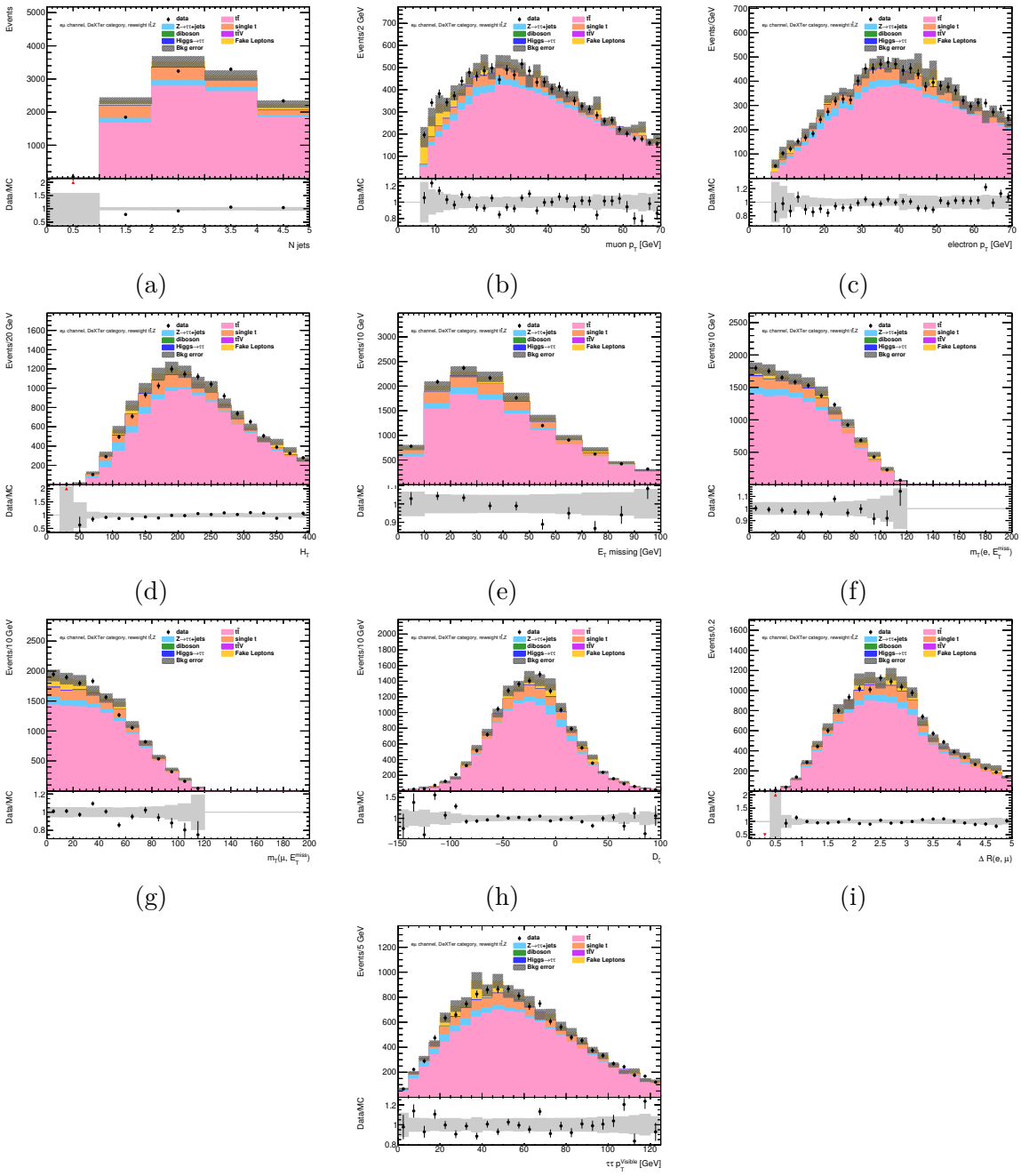


Figure A.22: Additional plots validating the kinematic modeling in the high mass ($m_{\tau\tau}^{\text{vis}} > 45\text{GeV}$) CR for the $e\mu$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

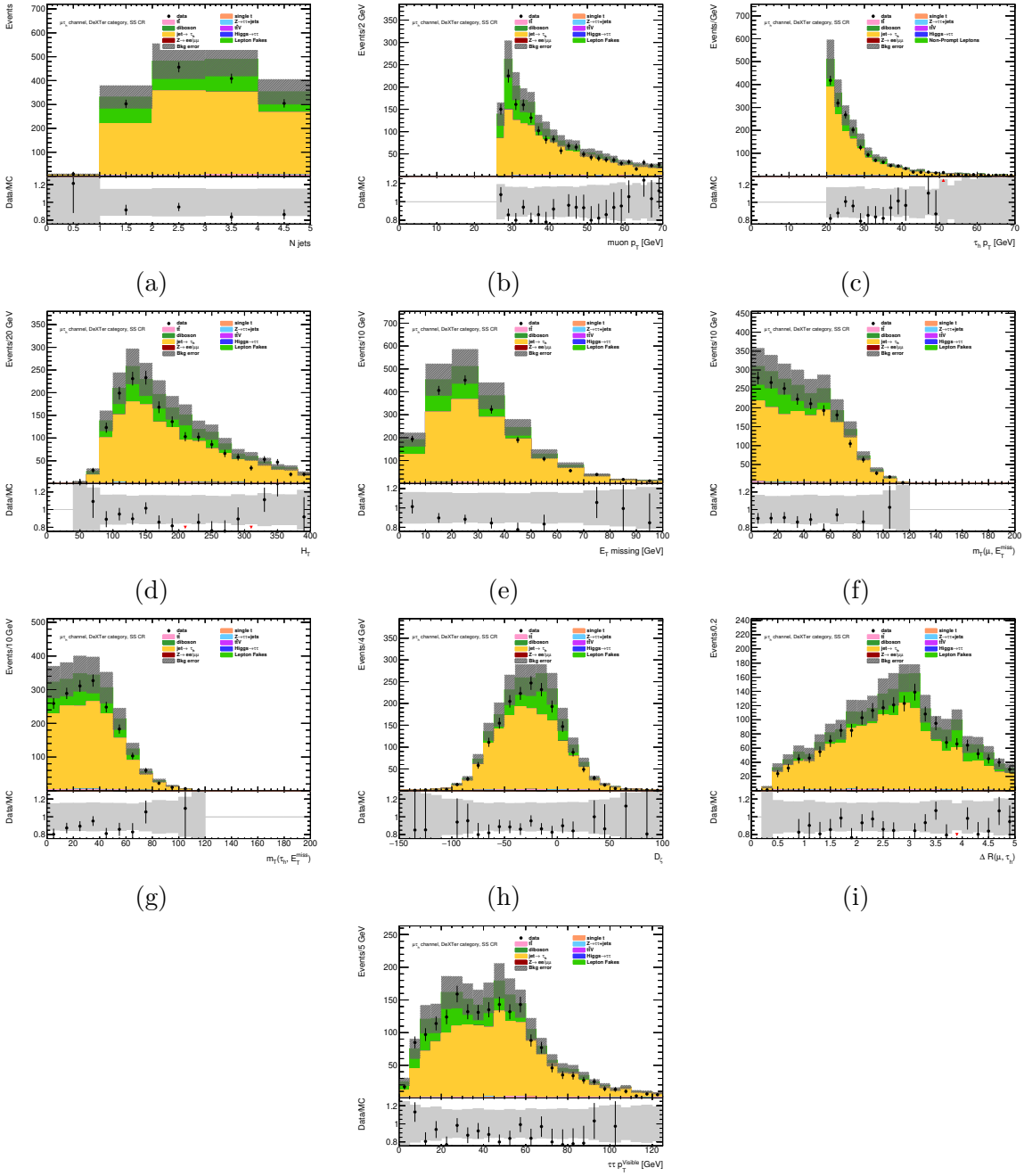


Figure A.23: Additional plots validating the kinematic modeling in the SS CR for the $\mu\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

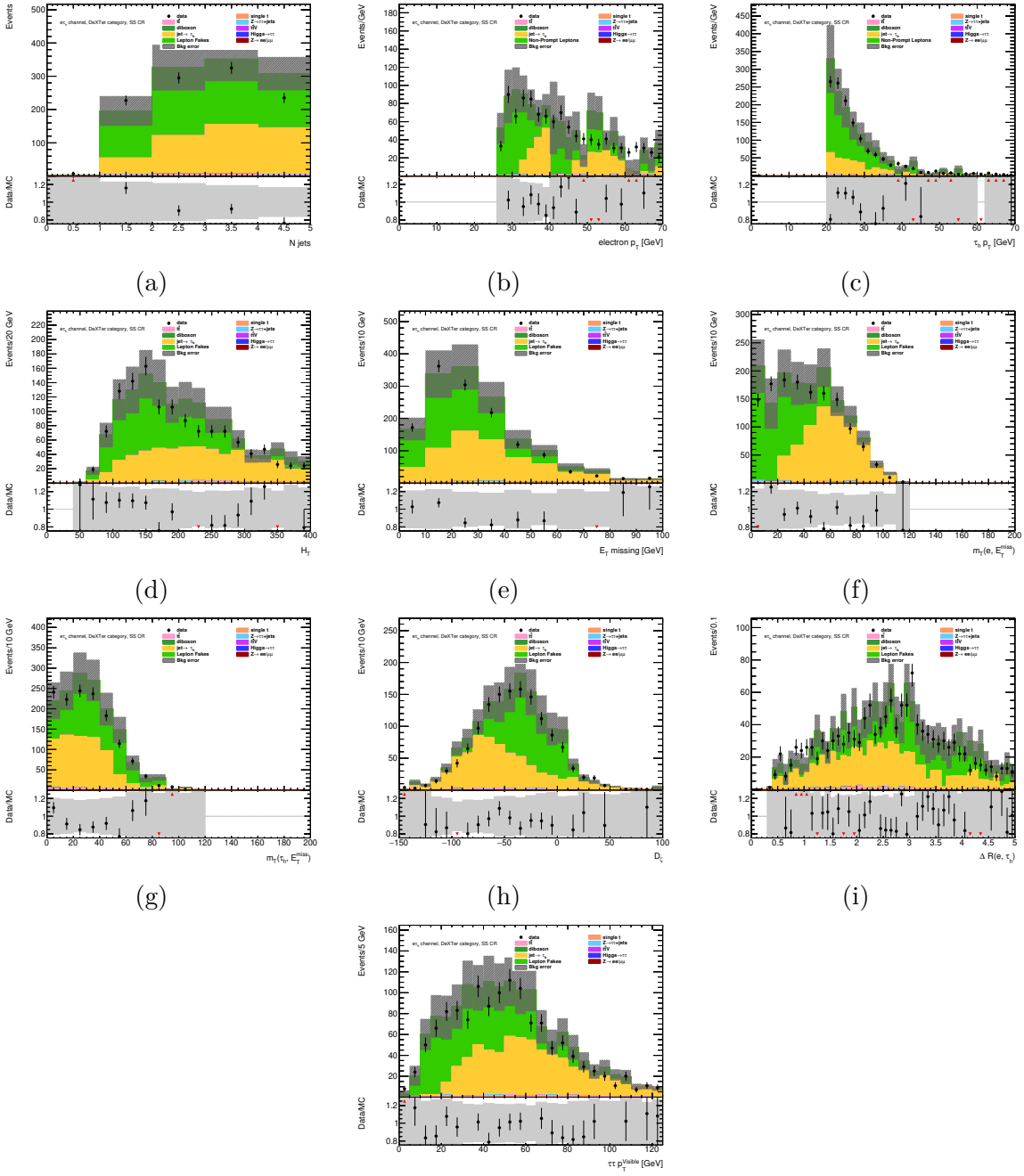


Figure A.24: Additional plots validating the kinematic modeling in the SS CR for the $e\tau_{\text{had}}$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

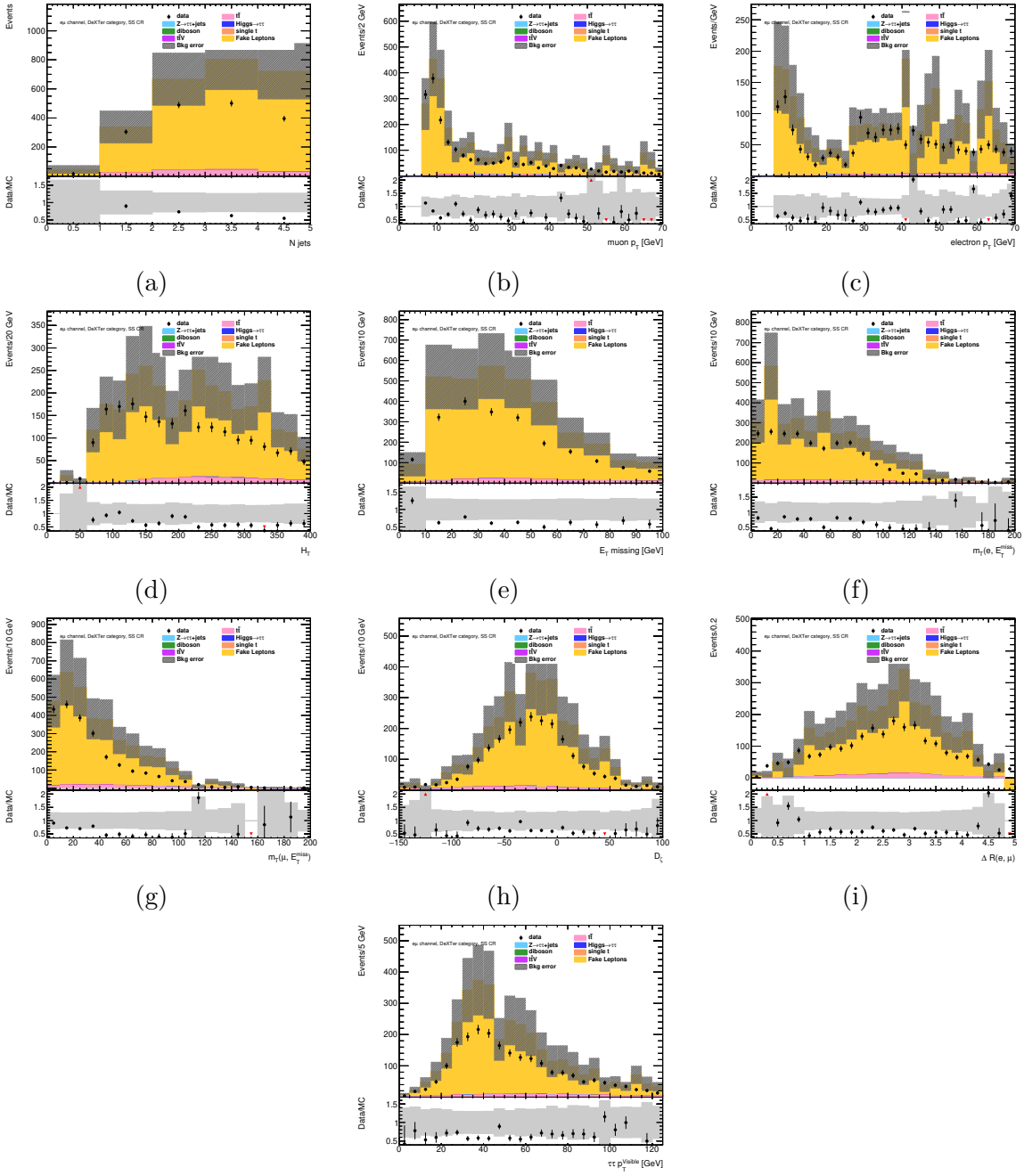
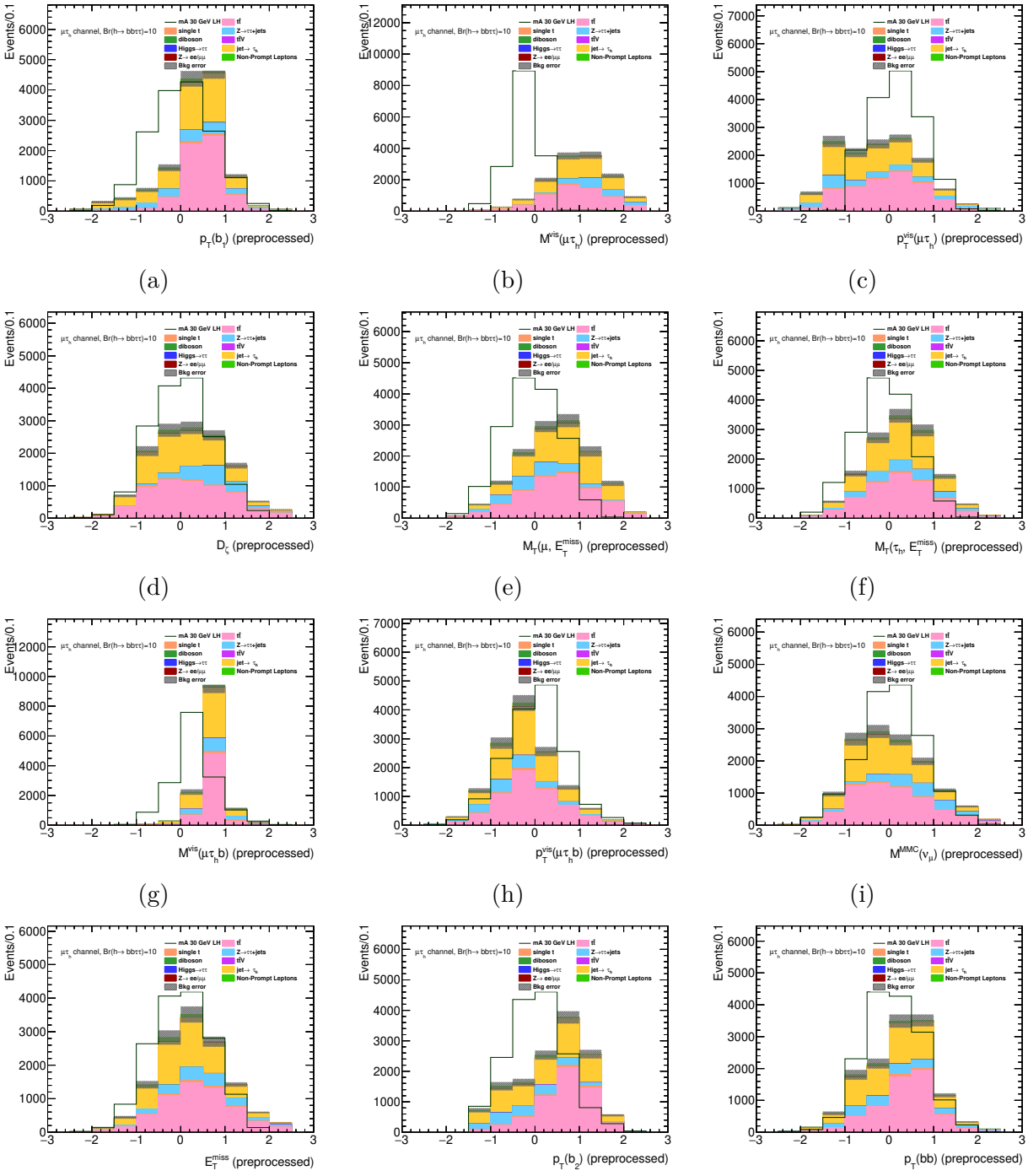


Figure A.25: Additional plots validating the kinematic modeling in the SS CR for the $e\mu$ channel, with the DeXTer selection applied. The uncertainty bands illustrate a combination of the statistical error on the MC samples and the systematic errors coming from the estimation of the fake lepton and fake tau backgrounds.

A.3 Neural Network

In this section additional PNN validation plots are presented. Refer to Section 5.5 for more information about the network architecture and training. Figures A.26–A.38 show the pre-processed neural network input variables for the background and signal prediction. All channels and categories are presented here, except the $\mu\tau_{\text{had}}$ 1 b-jet category, which is presented in Figure 5.34. Unblinded validation plots in the low sideband of $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}}))$ are presented in Figures A.40–A.52, again with the $\mu\tau_{\text{had}}$ 1 b-jet category appearing in the main text in Figure 5.36. Cross-entropy loss curves during NN training are shown in Figures A.54–A.56, with no evidence of over-fitting arising from a difference between the testing and training loss. The discriminating power of the NNs are shown in Figures A.57–A.59, which plots signal samples against the background prediction, with the high NN output score region blinded.



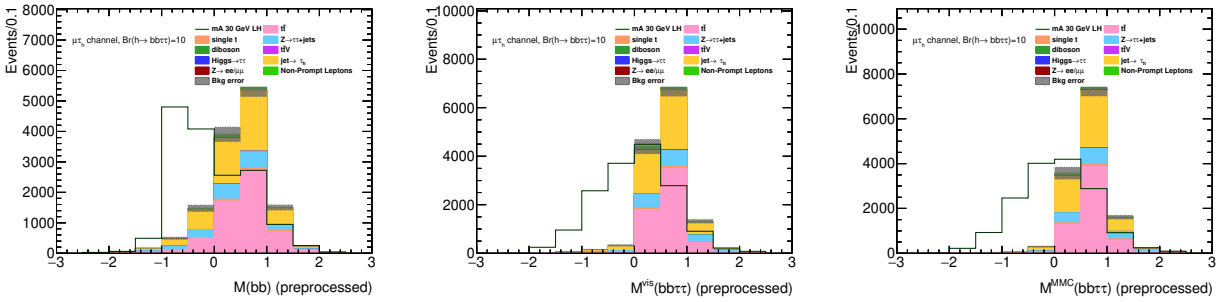
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

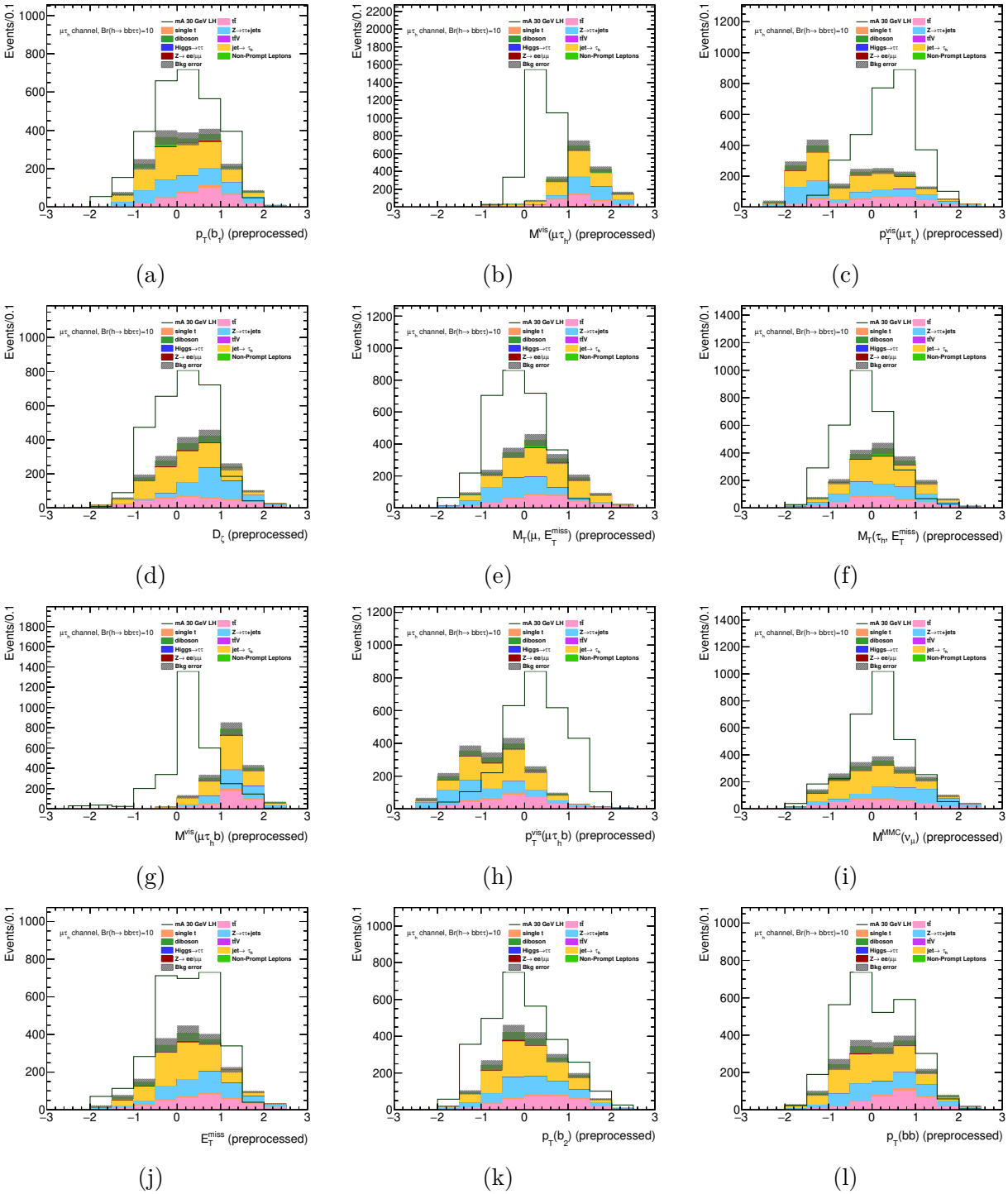
(j) (k) (l)

Figure A.26: Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, >1 b-jet category. Continued in Figure A.27.



(a) (b) (c)

Figure A.27: Pre-processed network input distributions for the $\mu\tau_{\text{had}}$ channel, >1 b-jet category. Continued from Figure A.26.



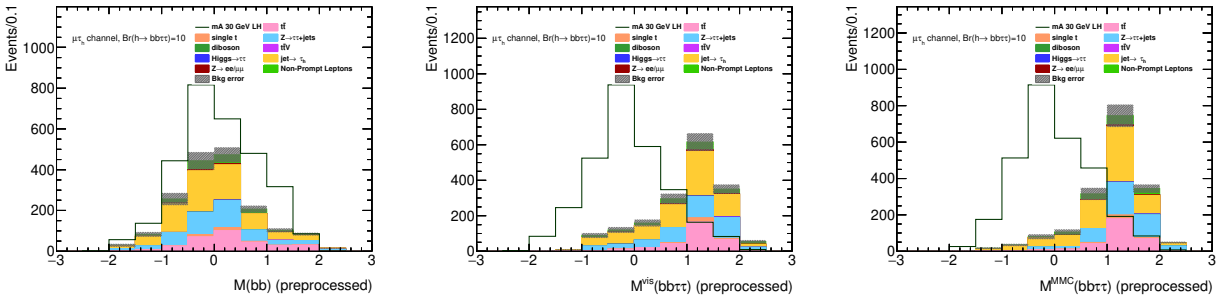
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.28: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. Continued in Figure A.29.



(a) (b) (c)

Figure A.29: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. Continued from Figure A.28.

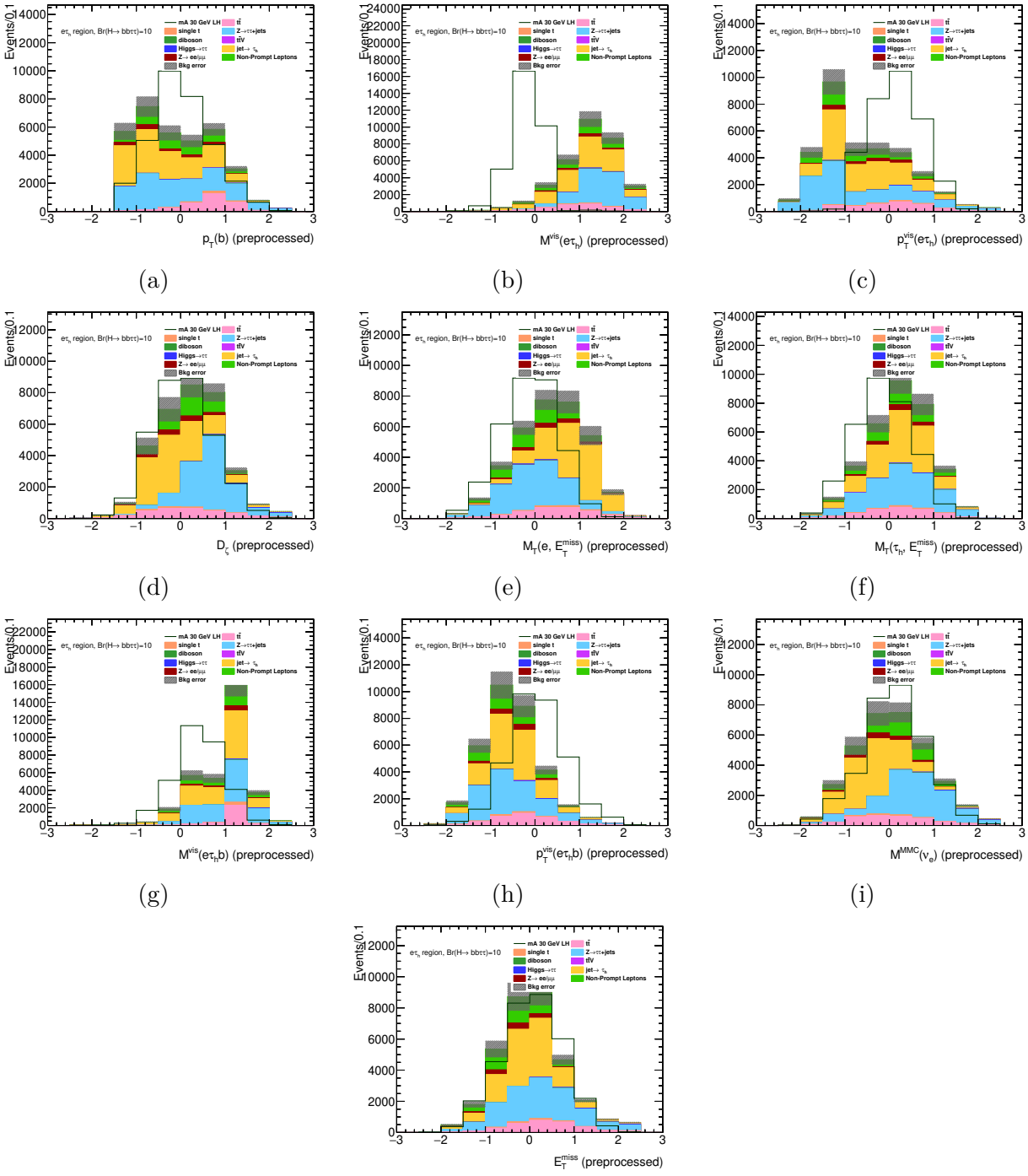


Figure A.30: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, 1 b-jet category.

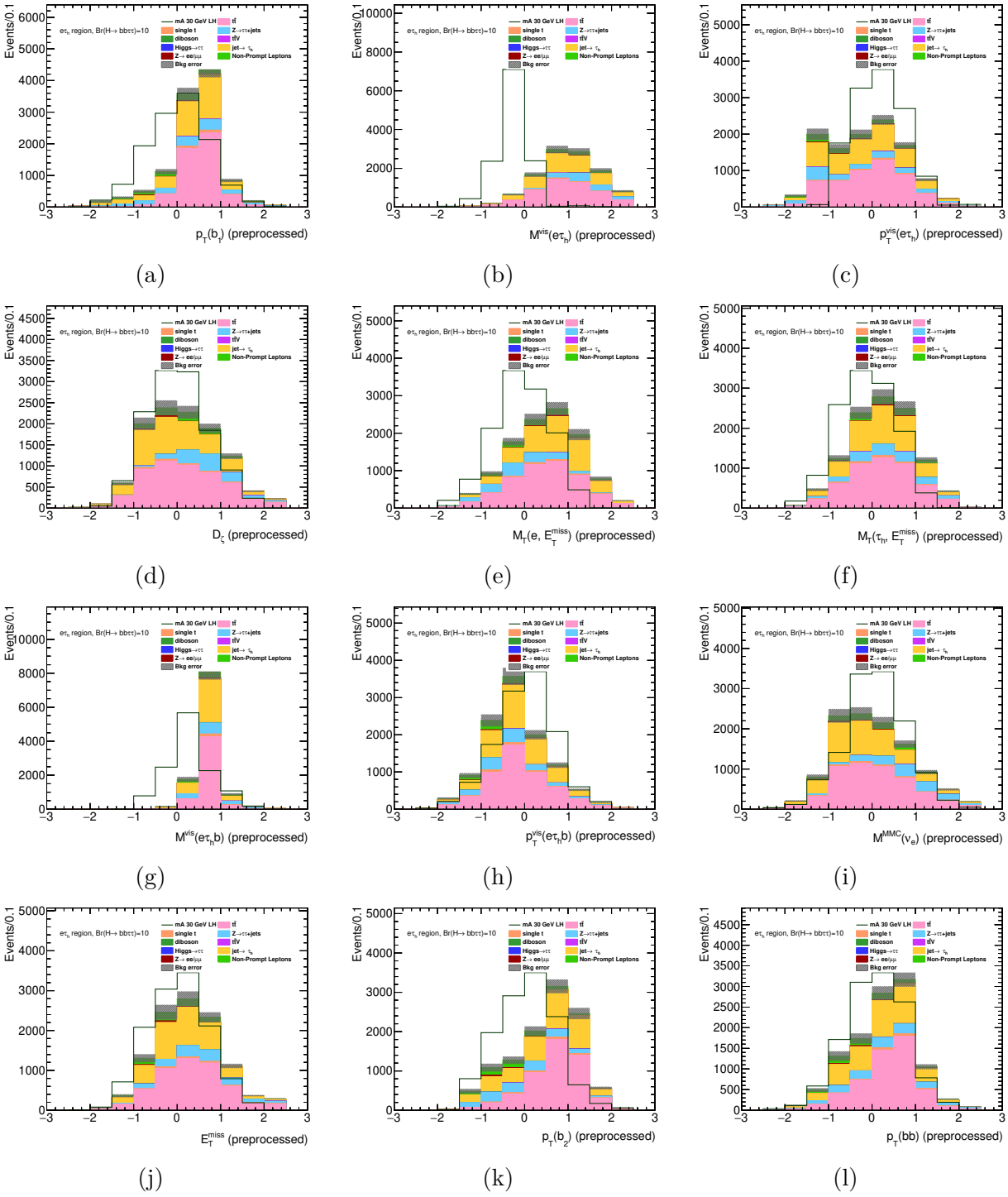
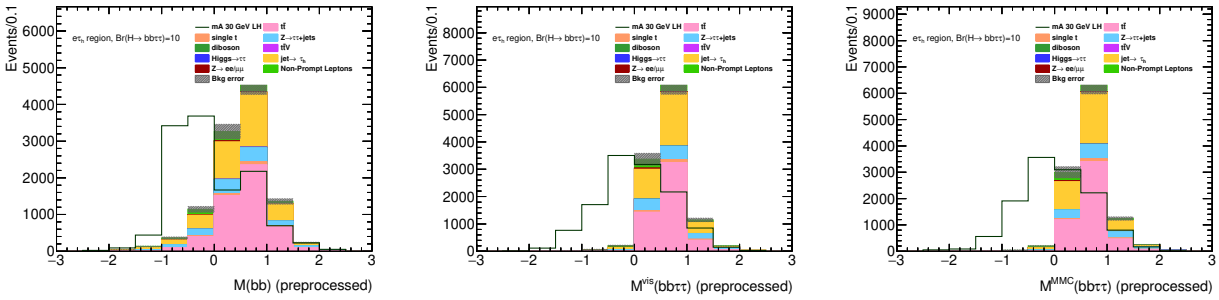


Figure A.31: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, >1 b-jet category. Continued in Figure A.32.



(a) (b) (c)

Figure A.32: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, $>$ 1 b-jet category. Continued from Figure A.31.

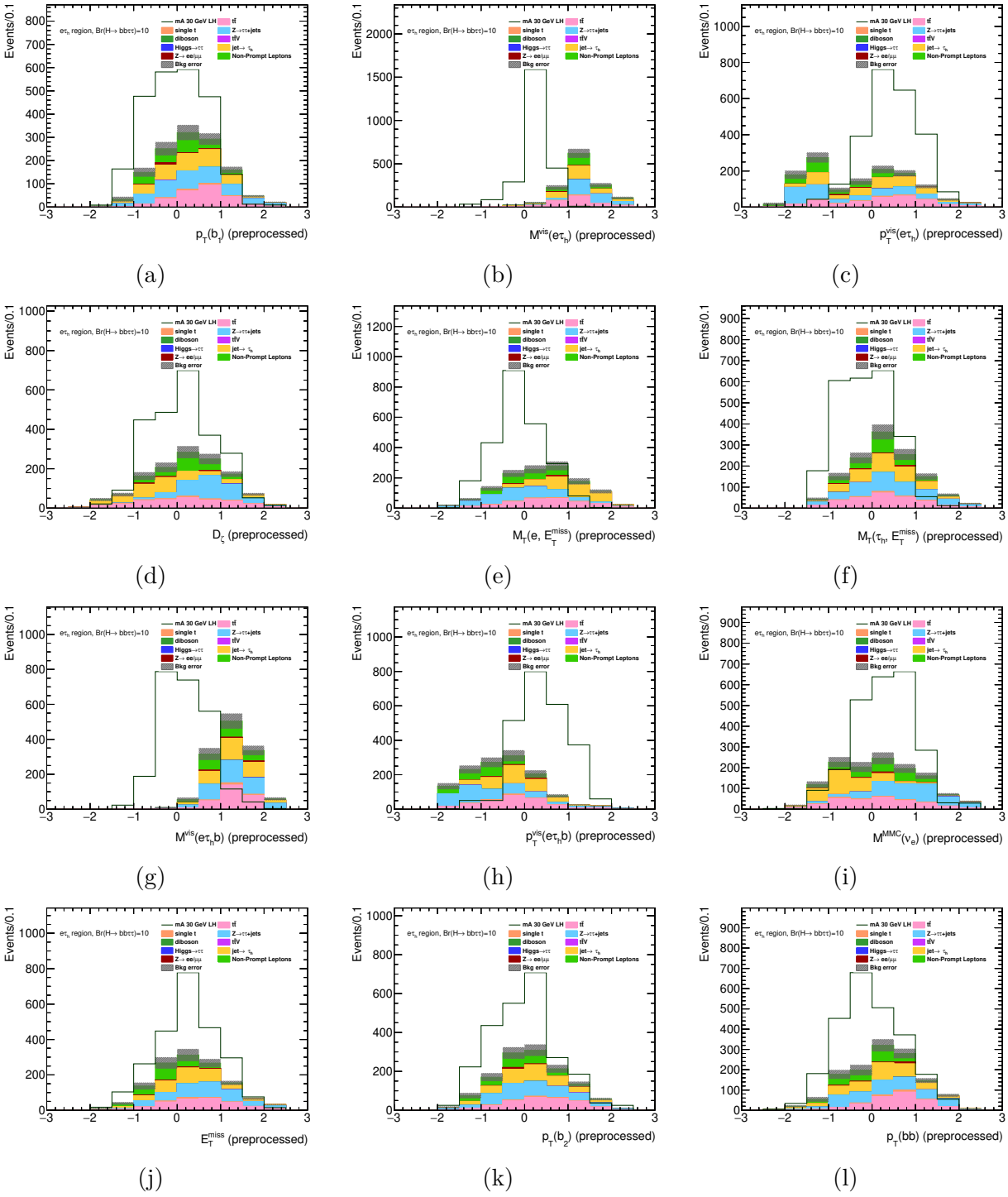
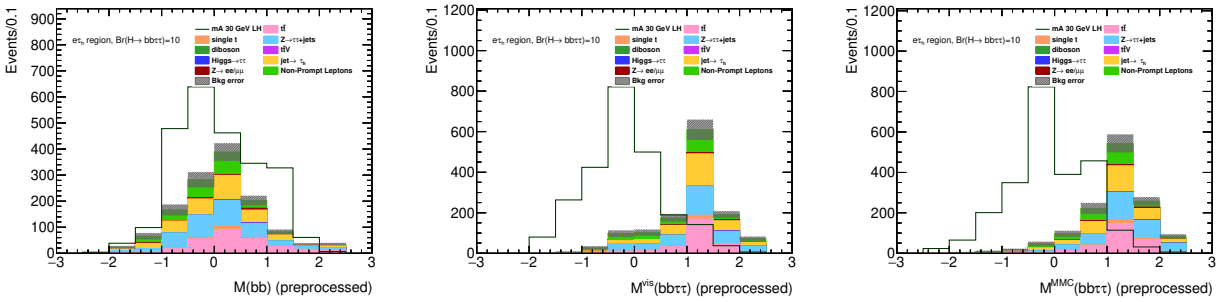


Figure A.33: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. Continued in Figure A.34.



(a) (b) (c)

Figure A.34: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. Continued from Figure A.33.

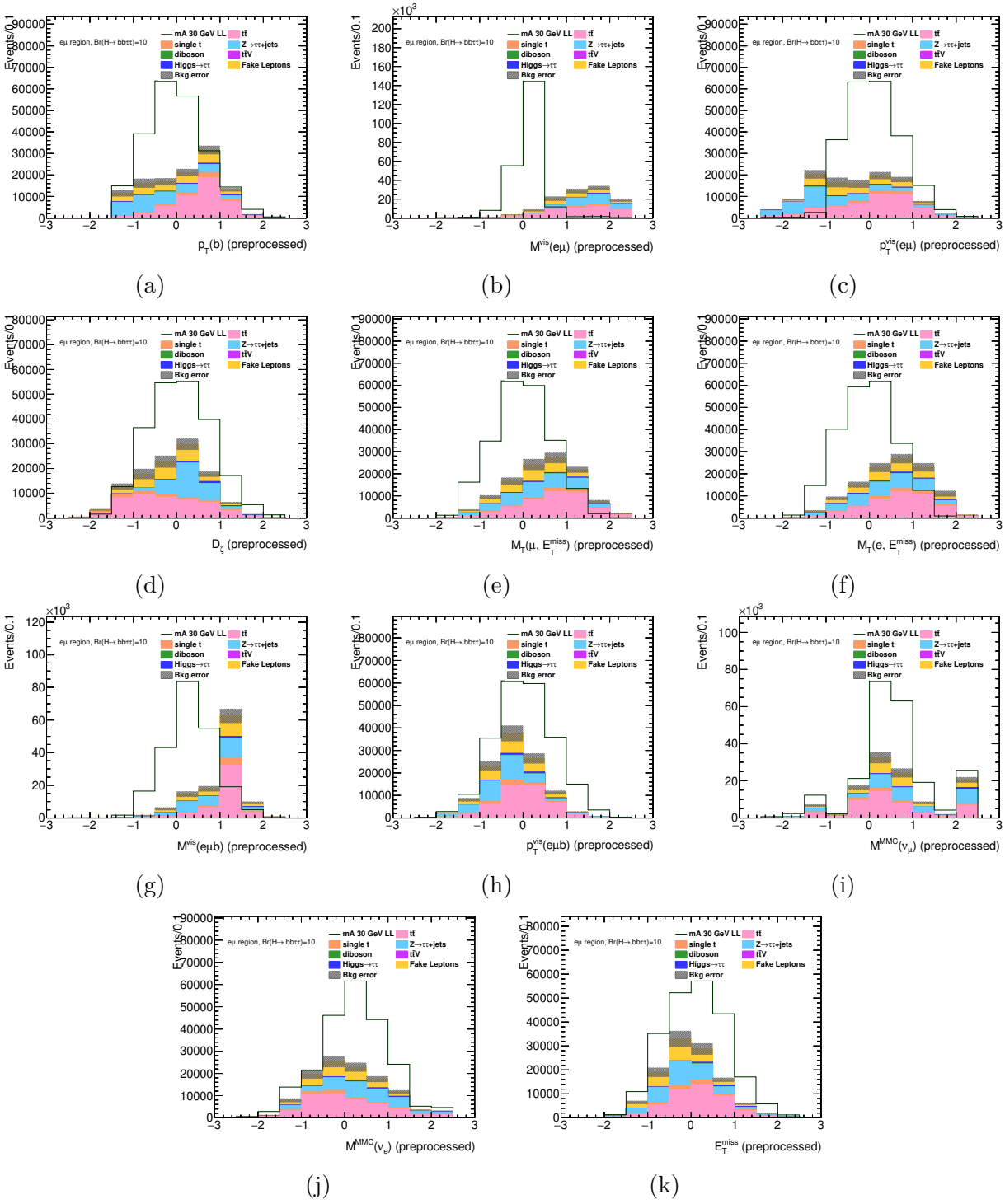


Figure A.35: Pre-Processed network input distributions for the $e\mu$ channel, 1 b-jet category.

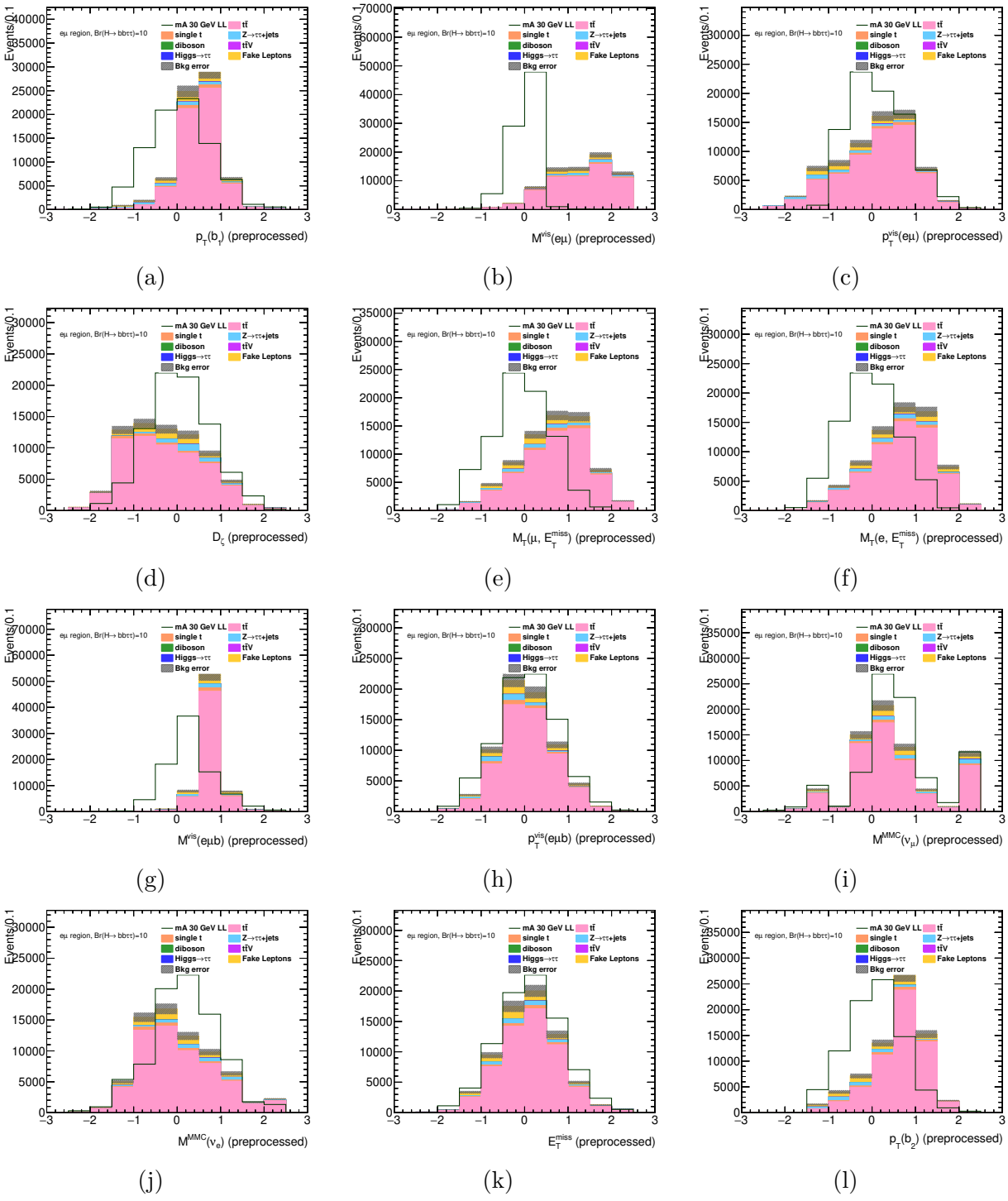
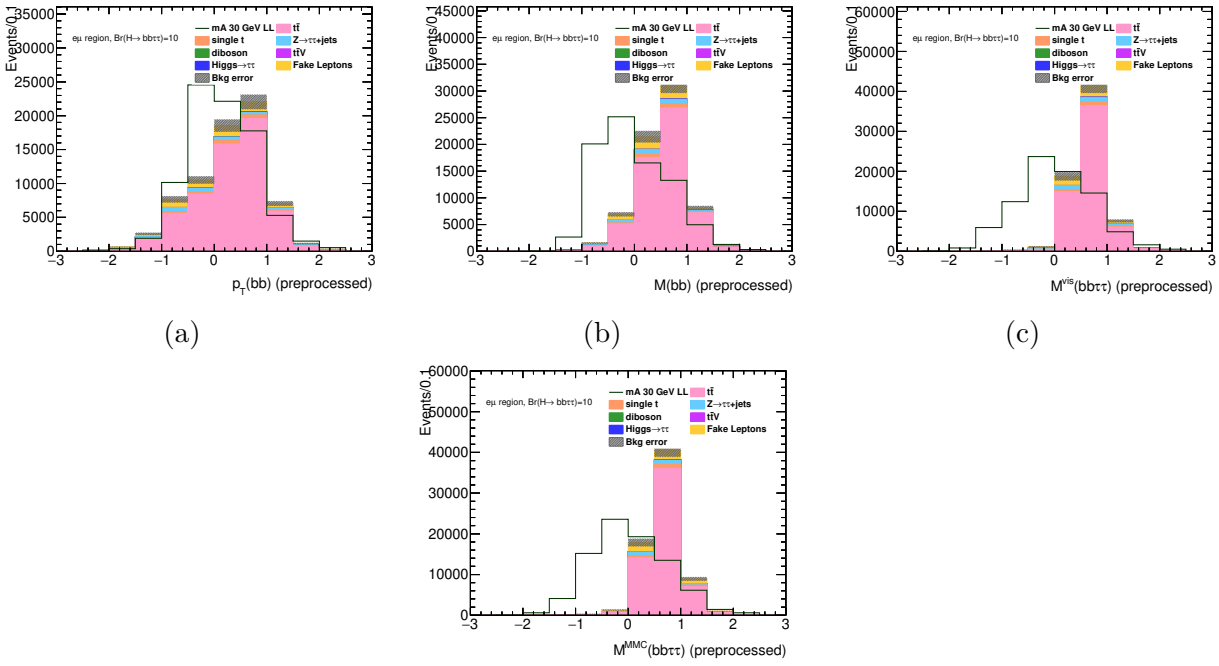


Figure A.36: Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. Continued in Figure A.37.



(a) (b) (c)

(d)

Figure A.37: Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. Continued from Figure A.36.

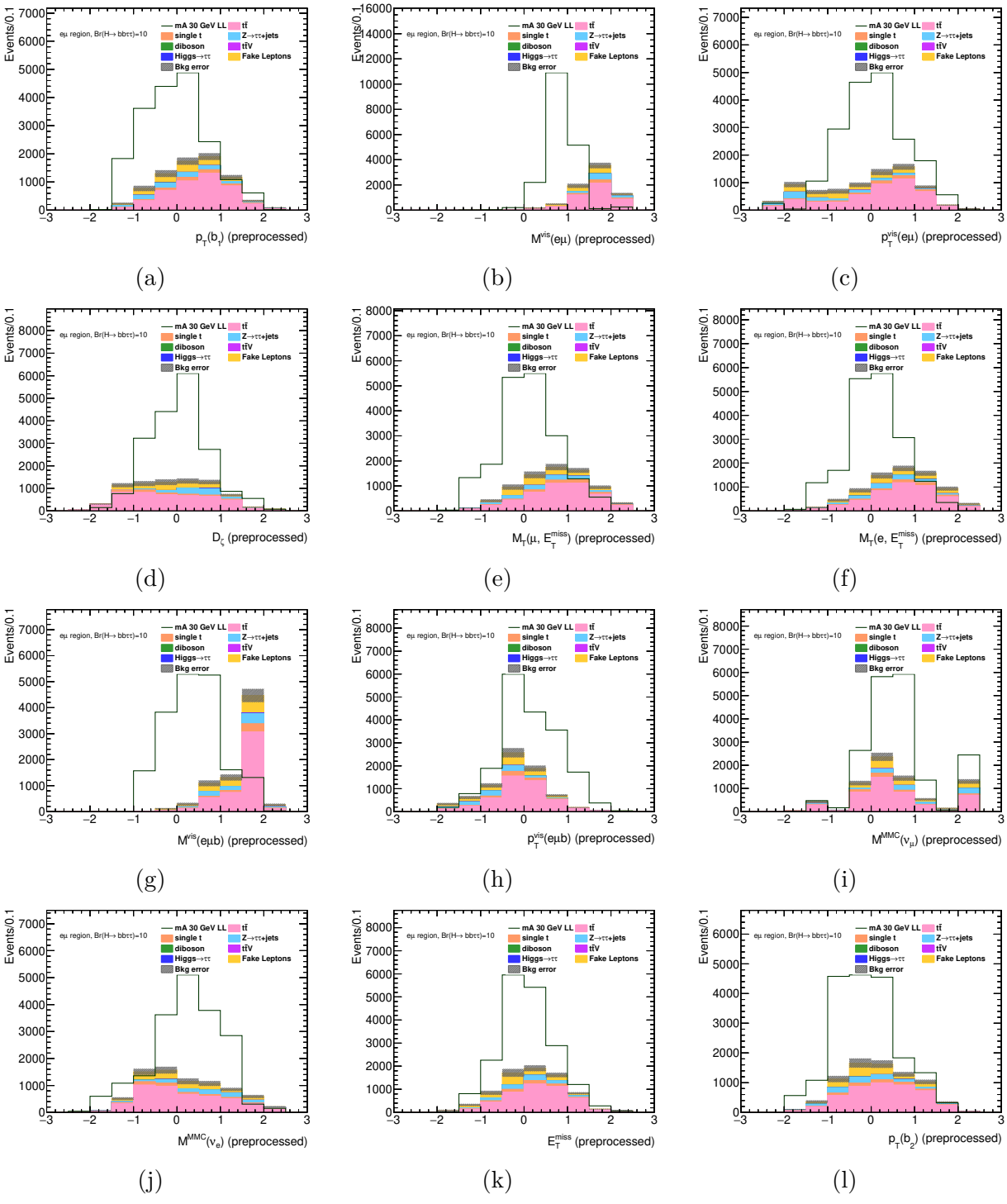
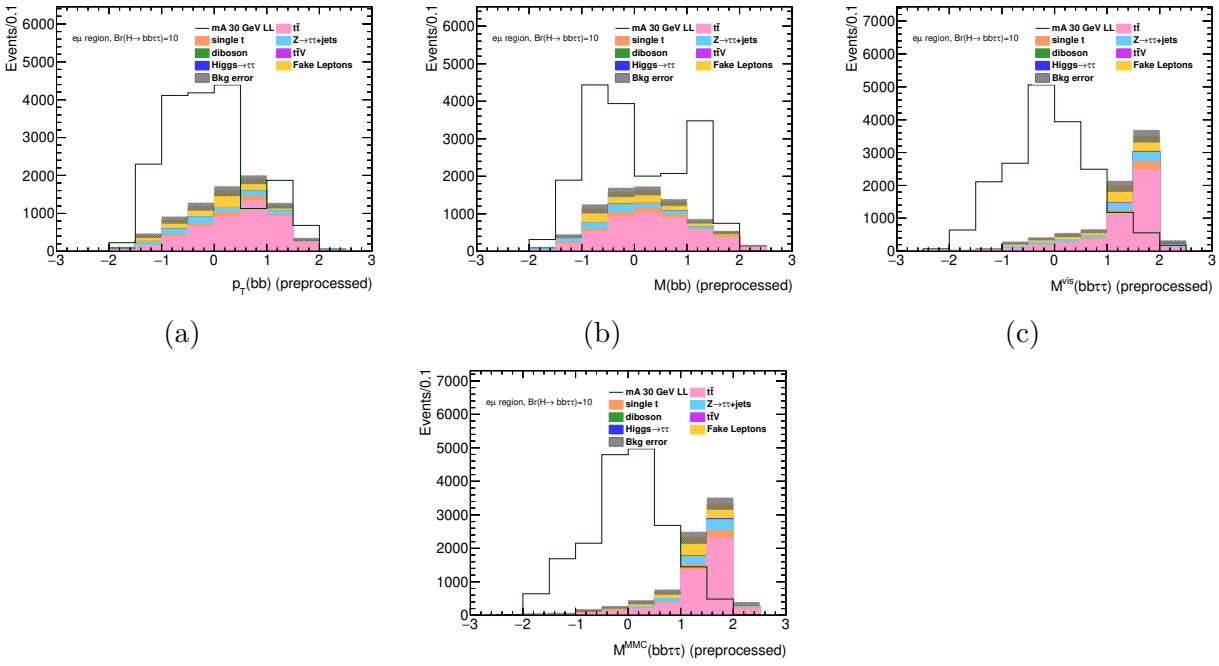


Figure A.38: Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. Continued in Figure A.39.



(a) (b) (c)

(d)

Figure A.39: Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. Continued from Figure A.38.

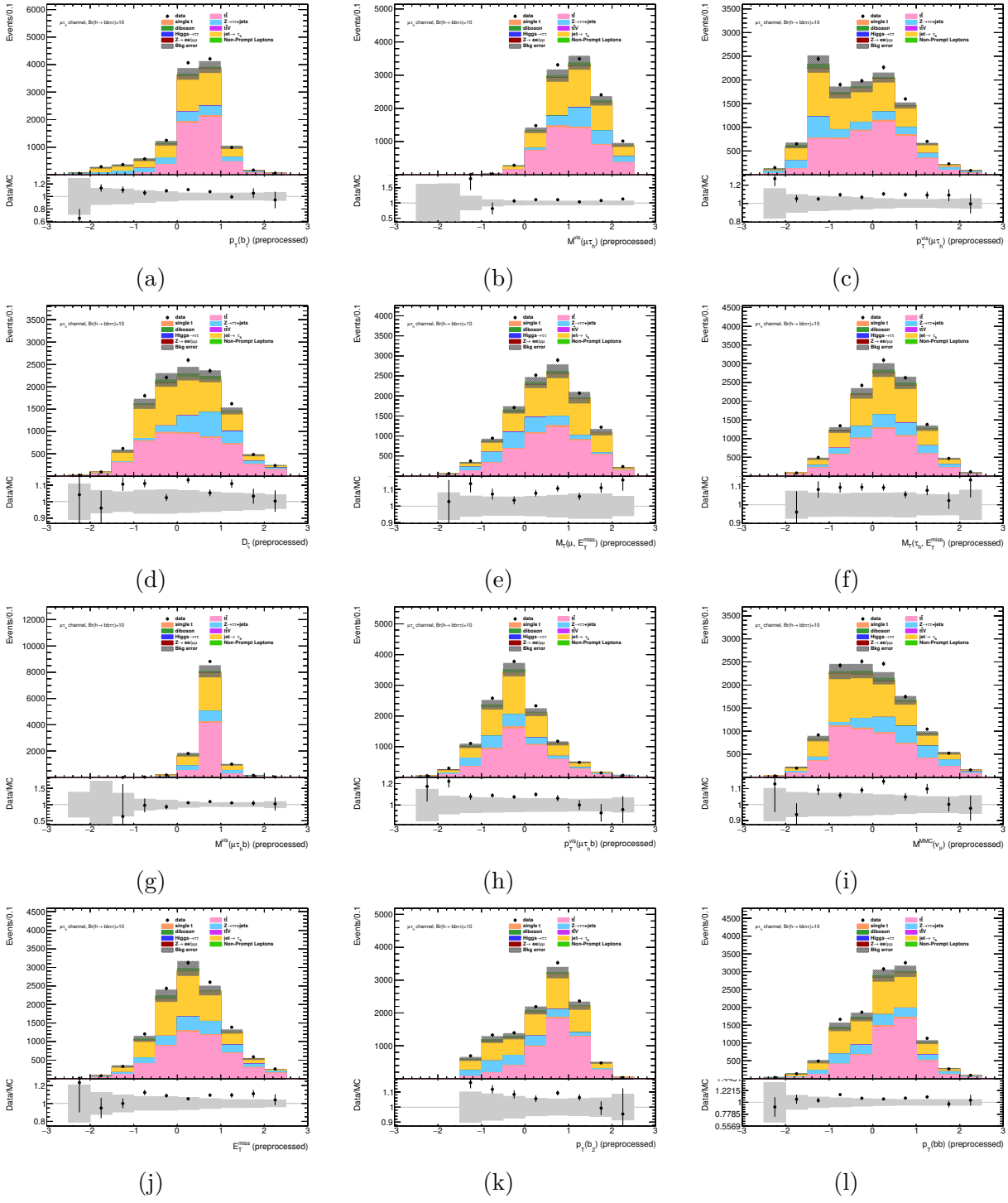
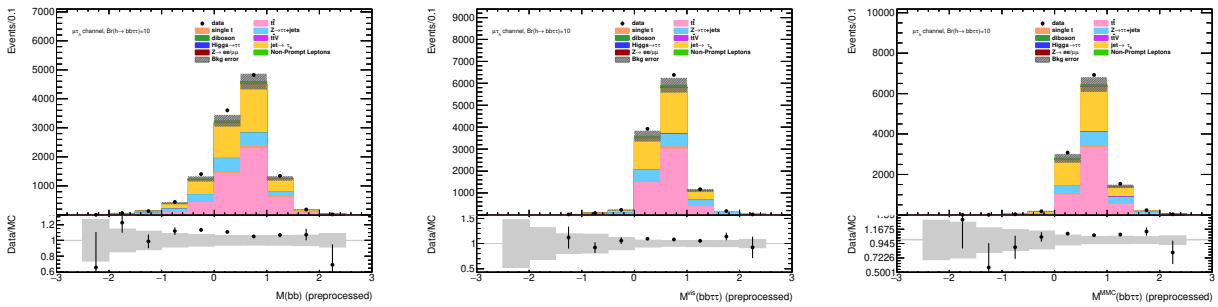


Figure A.40: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.41



(a) (b) (c)

Figure A.41: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.40.

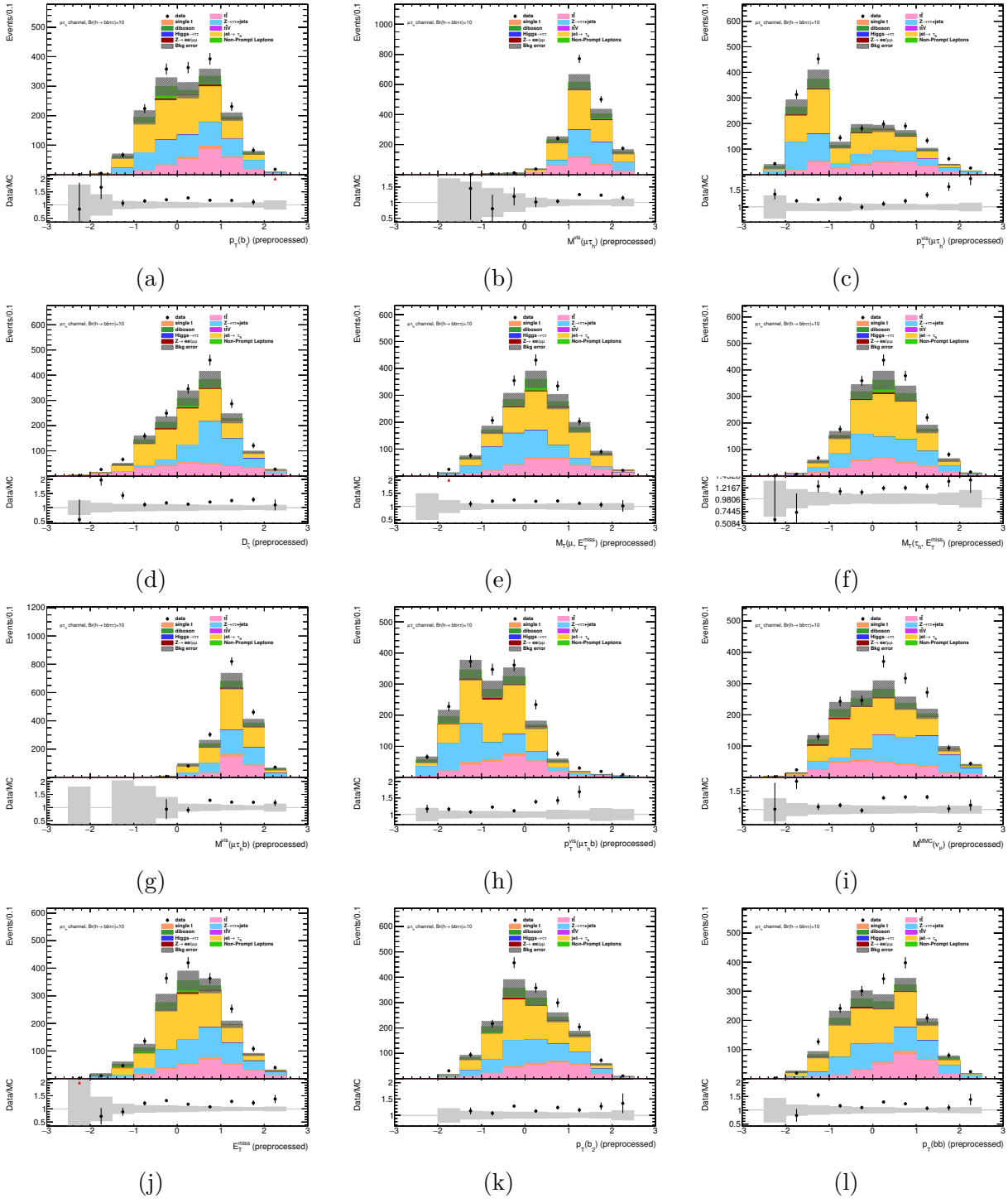
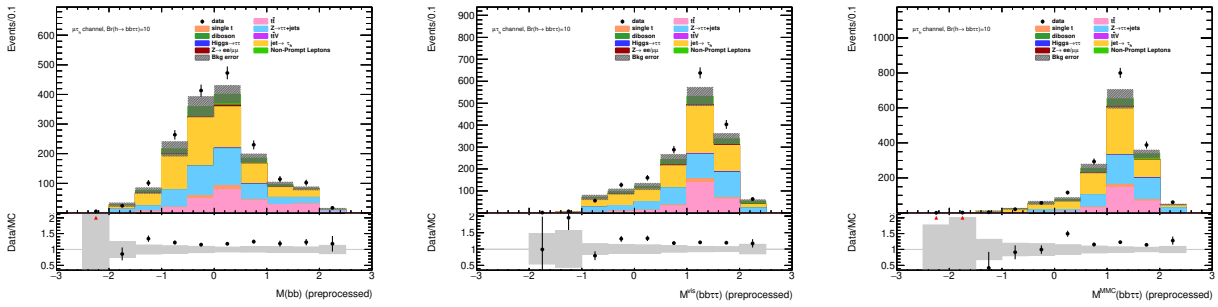


Figure A.42: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\,\text{pre-trained}})) < -6$ is applied. Continued in Figure A.43.



(a) (b) (c)

Figure A.43: Pre-Processed network input distributions for the $\mu\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.42.

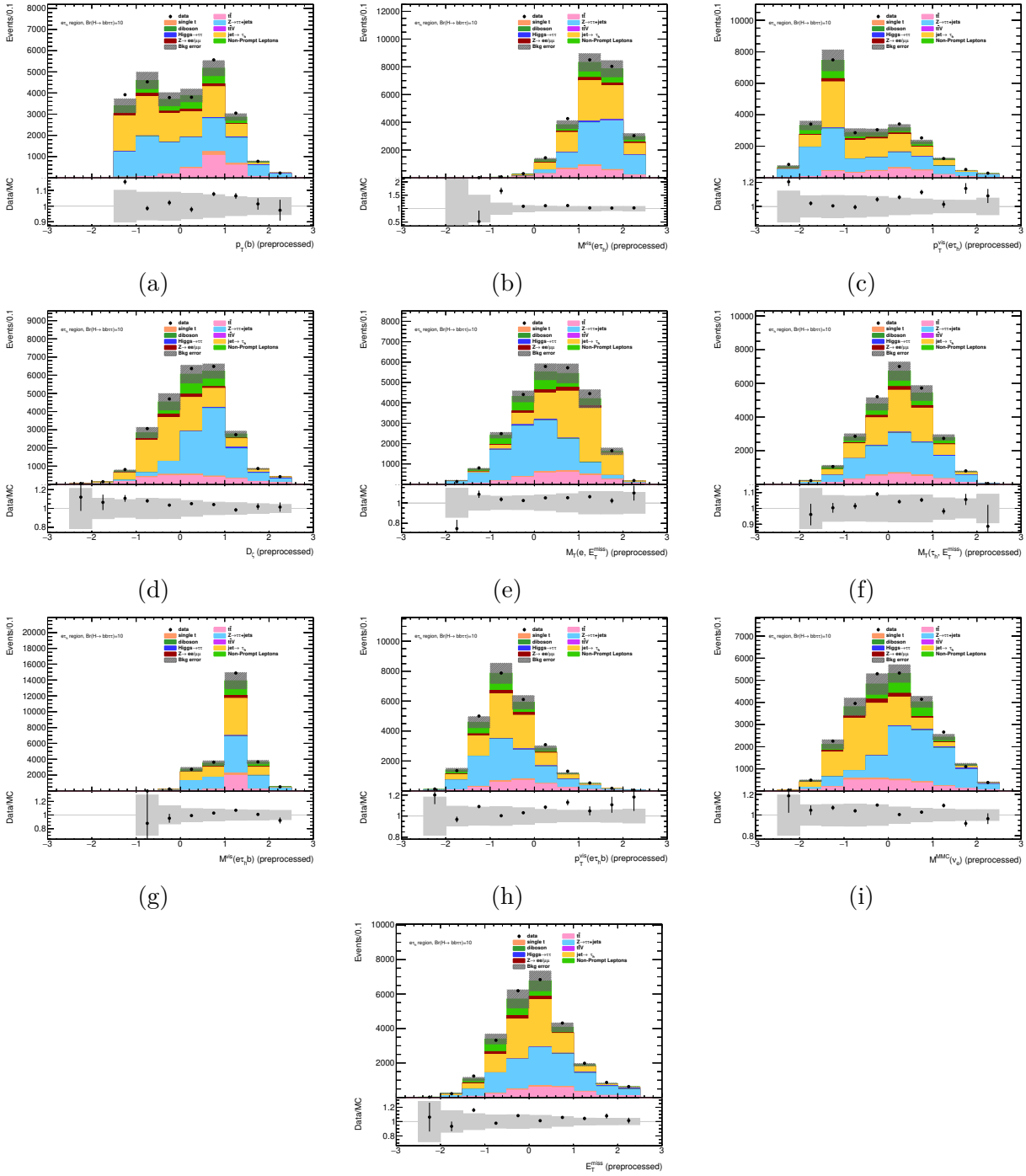


Figure A.44: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied.

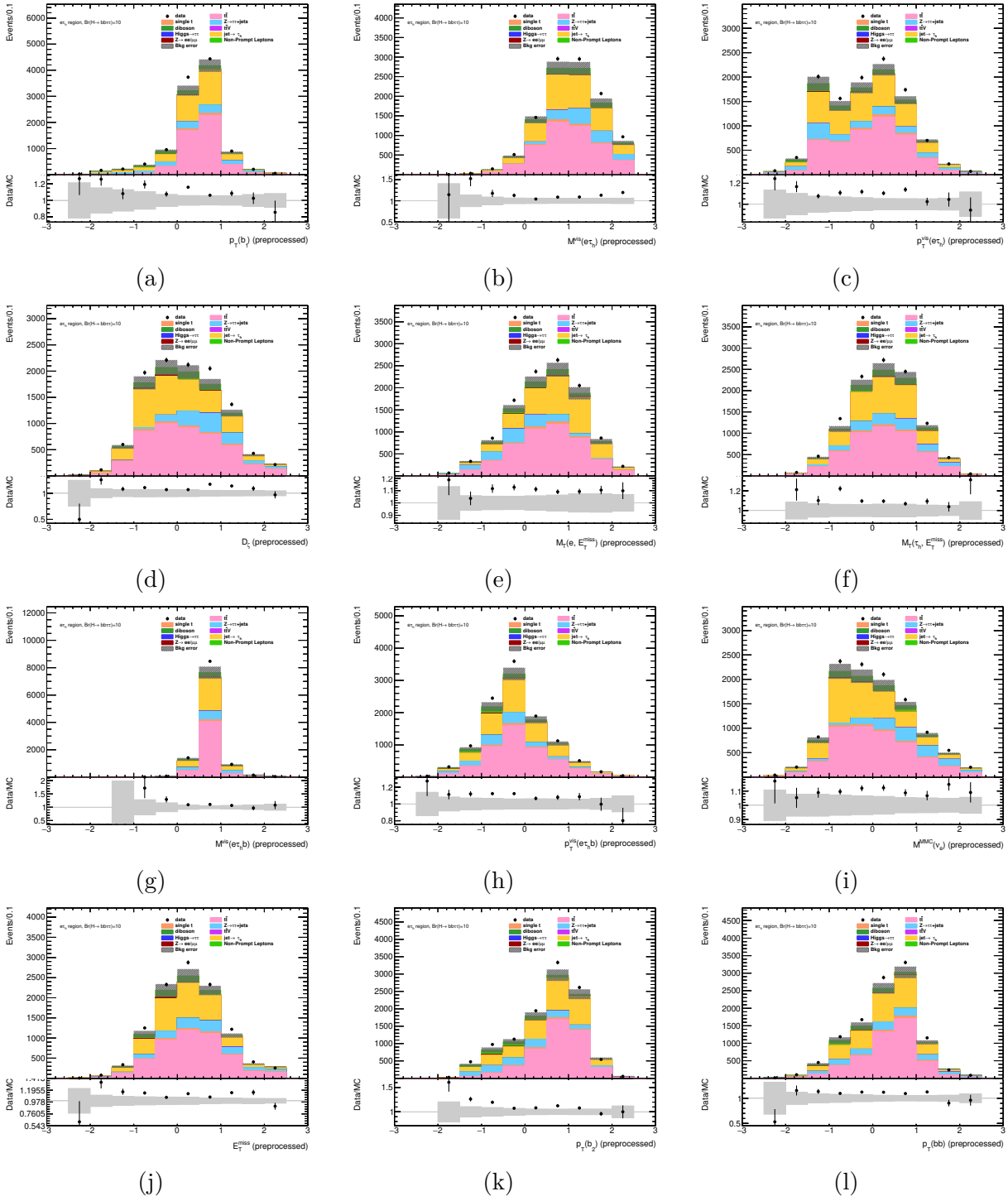
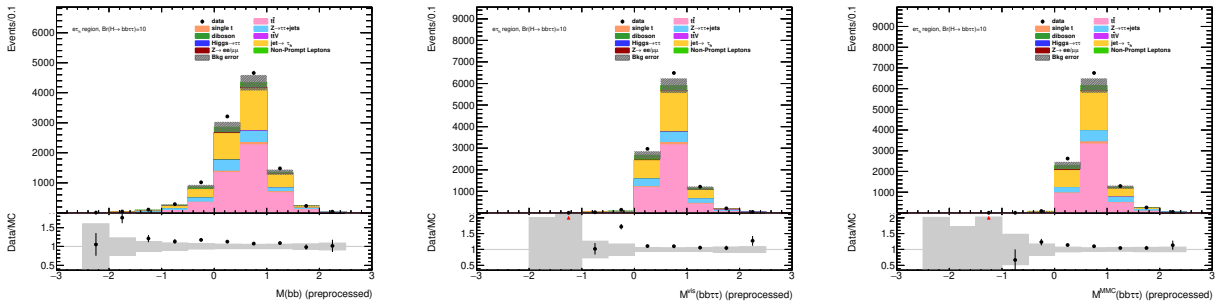


Figure A.45: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued in Figure A.46.



(a) (b) (c)

Figure A.46: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, > 1 b-jet category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued from Figure A.45.

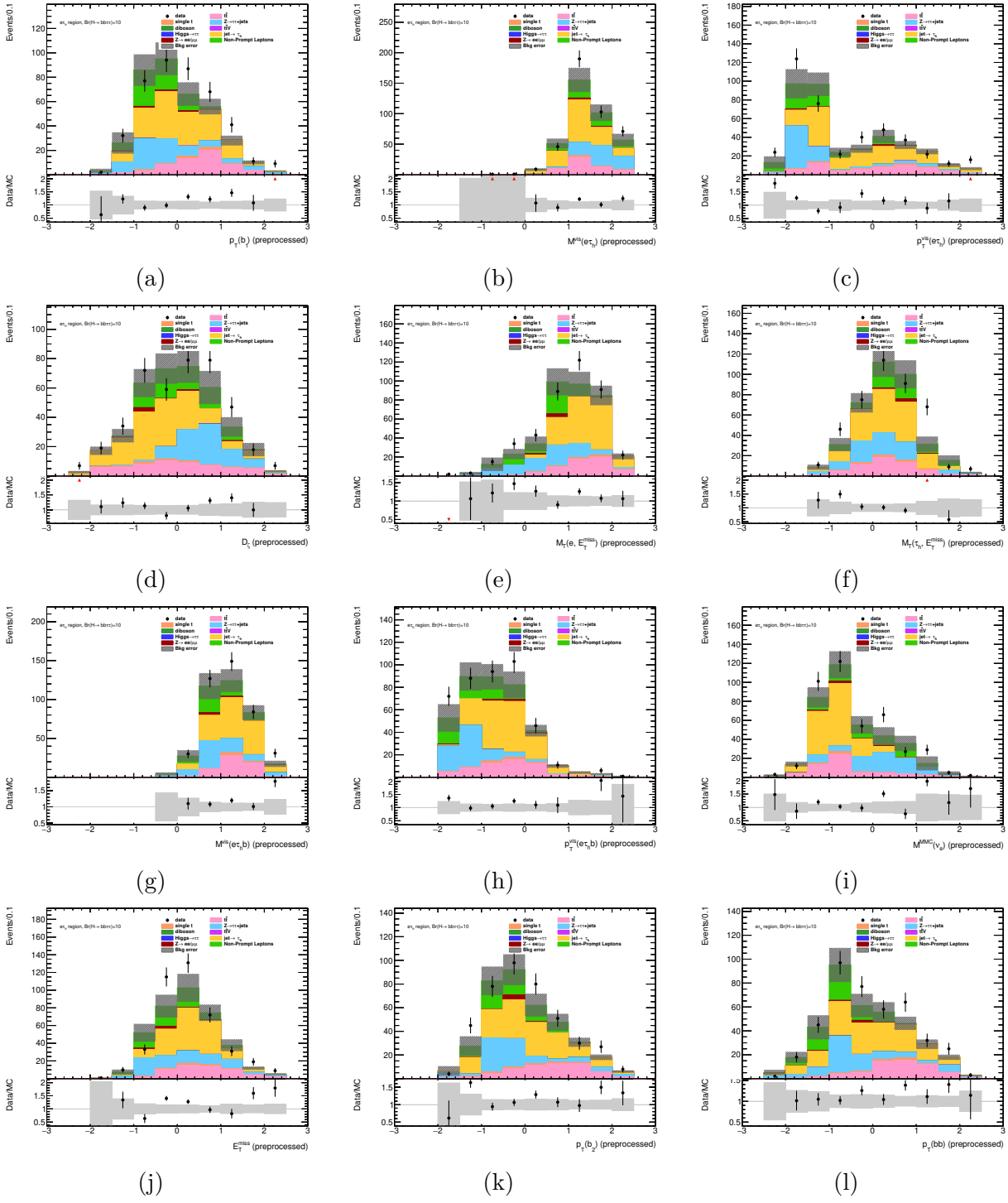
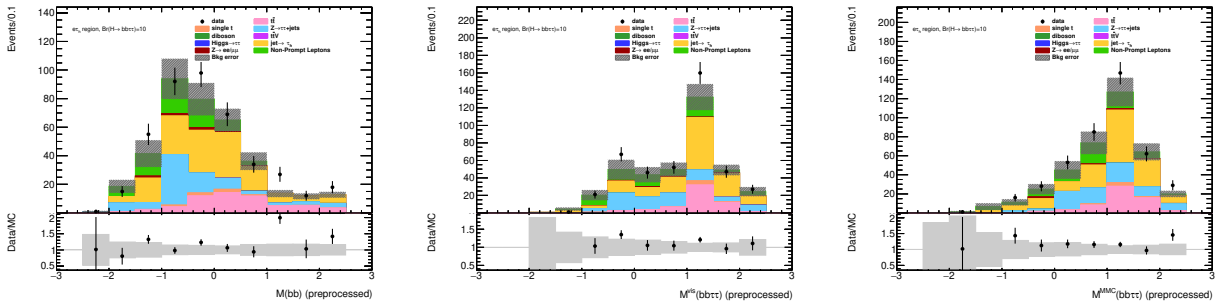


Figure A.47: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued in Figure A.48.



(a) (b) (c)

Figure A.48: Pre-Processed network input distributions for the $e\tau_{\text{had}}$ channel, DeXTer category. The cut $\ln(\text{PNN}(\vec{z}^{\text{pre-trained}})) < -12$ is applied. Continued from Figure A.47.

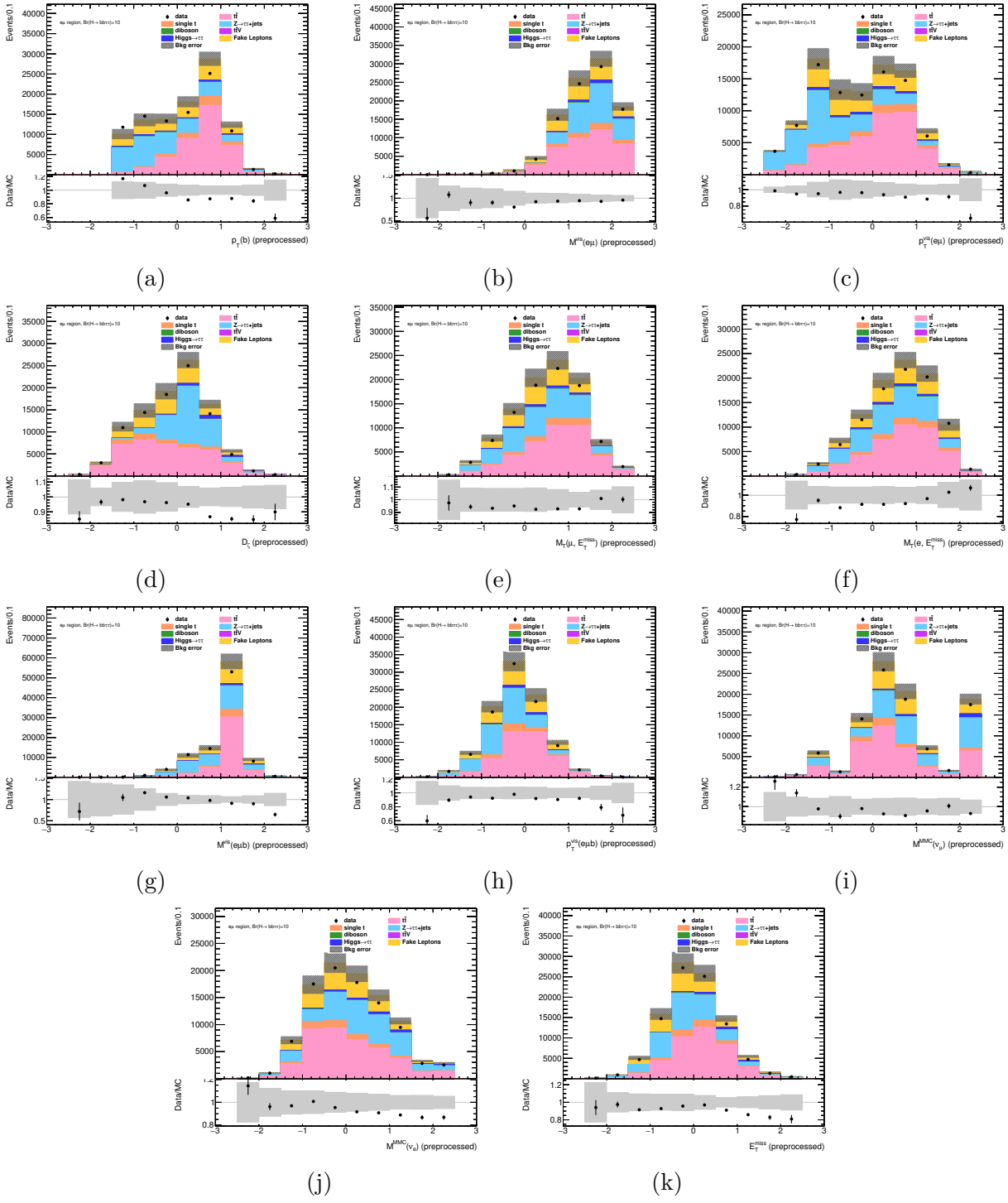


Figure A.49: Pre-Processed network input distributions for the $e\mu$ channel, 1 b-jet category. The cut $\ln(\mathrm{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied.

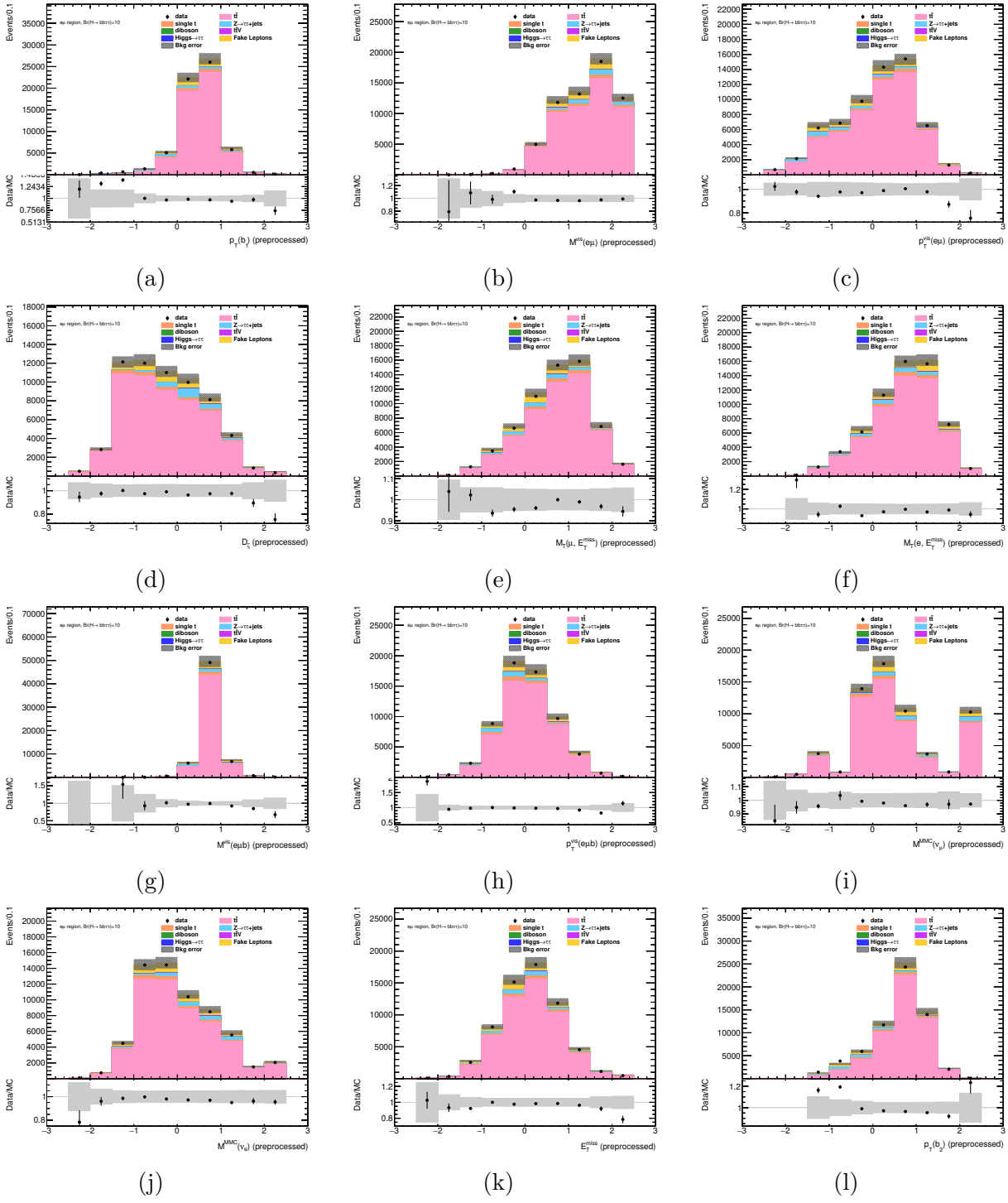
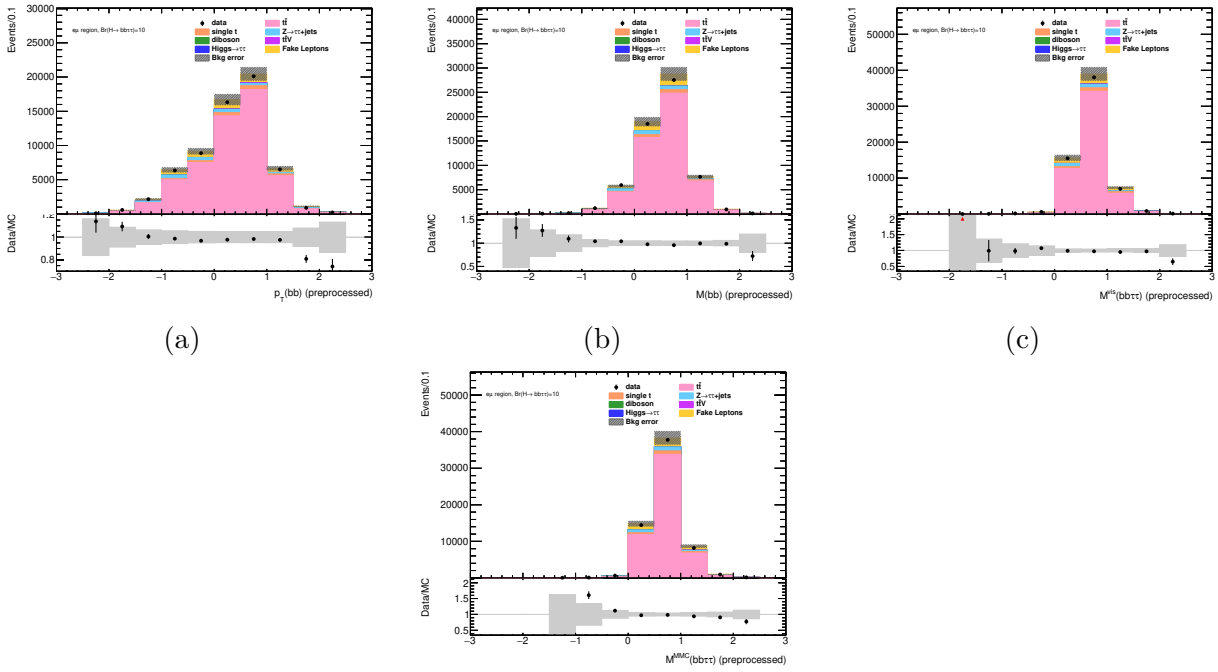


Figure A.50: Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. The cut $\ln(\mathrm{PNN}(\vec{z}^{\,\text{pre-trained}})) < -6$ is applied. Continued in Figure A.51.



(a) (b) (c)

(d)

Figure A.51: Pre-Processed network input distributions for the $e\mu$ channel, > 1 b-jet category. The cut $\ln(\mathrm{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.50.

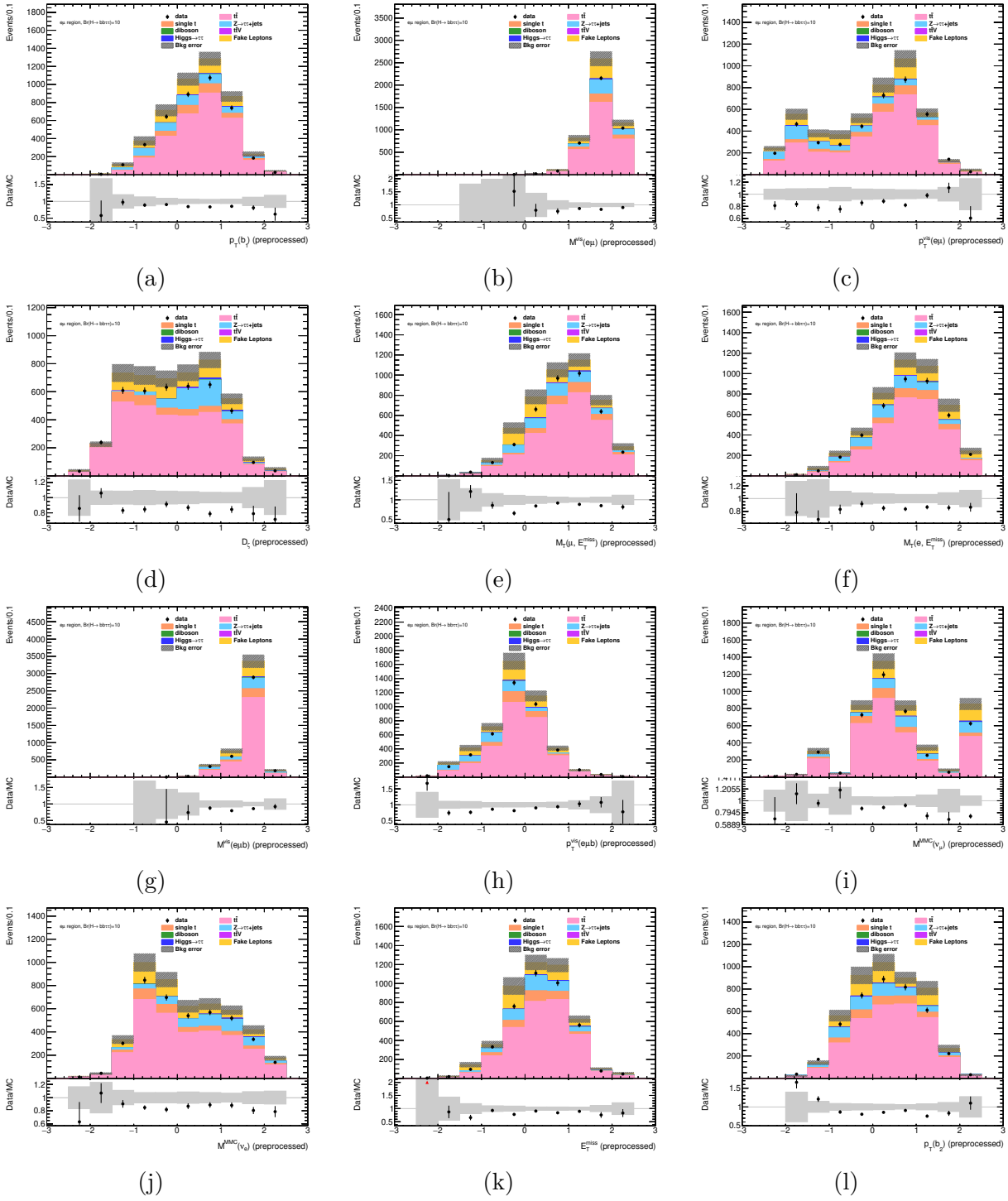
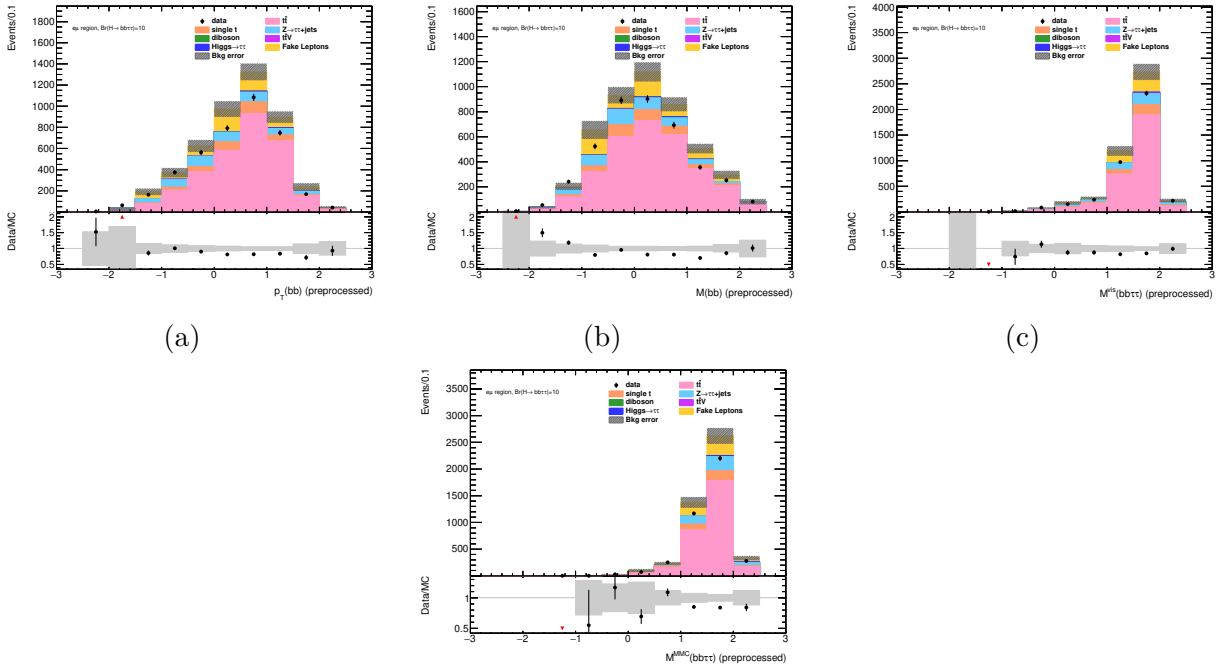


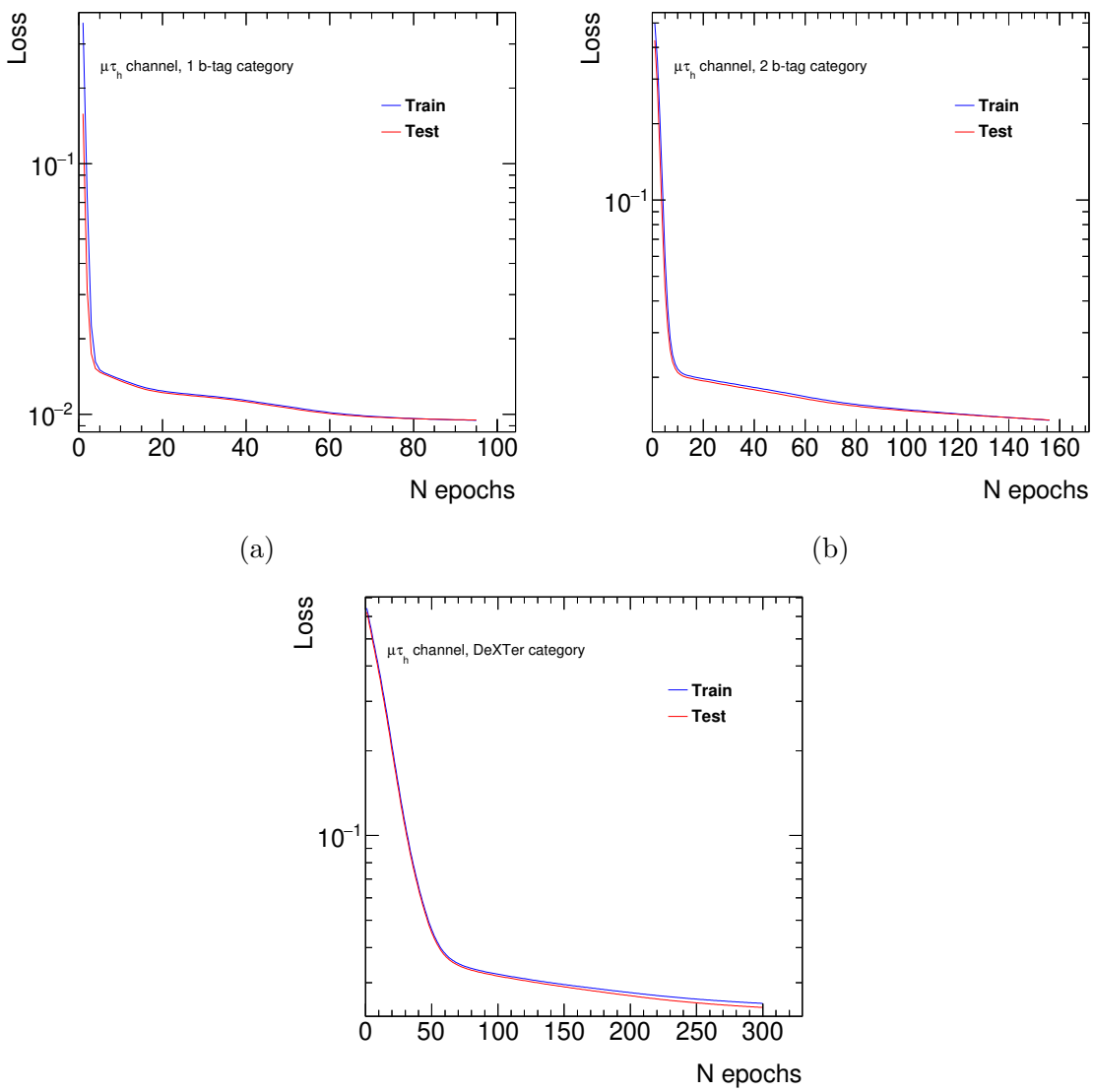
Figure A.52: Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. The cut $\ln(\mathrm{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued in Figure A.53.



(a) (b) (c)

(d)

Figure A.53: Pre-Processed network input distributions for the $e\mu$ channel, DeXTer category. The cut $\ln(\mathrm{PNN}(\vec{z}^{\text{pre-trained}})) < -6$ is applied. Continued from Figure A.52.



(a) (b)

(c)

Figure A.54: Cross-entropy loss function in the $\mu\tau_{\text{had}}$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories.

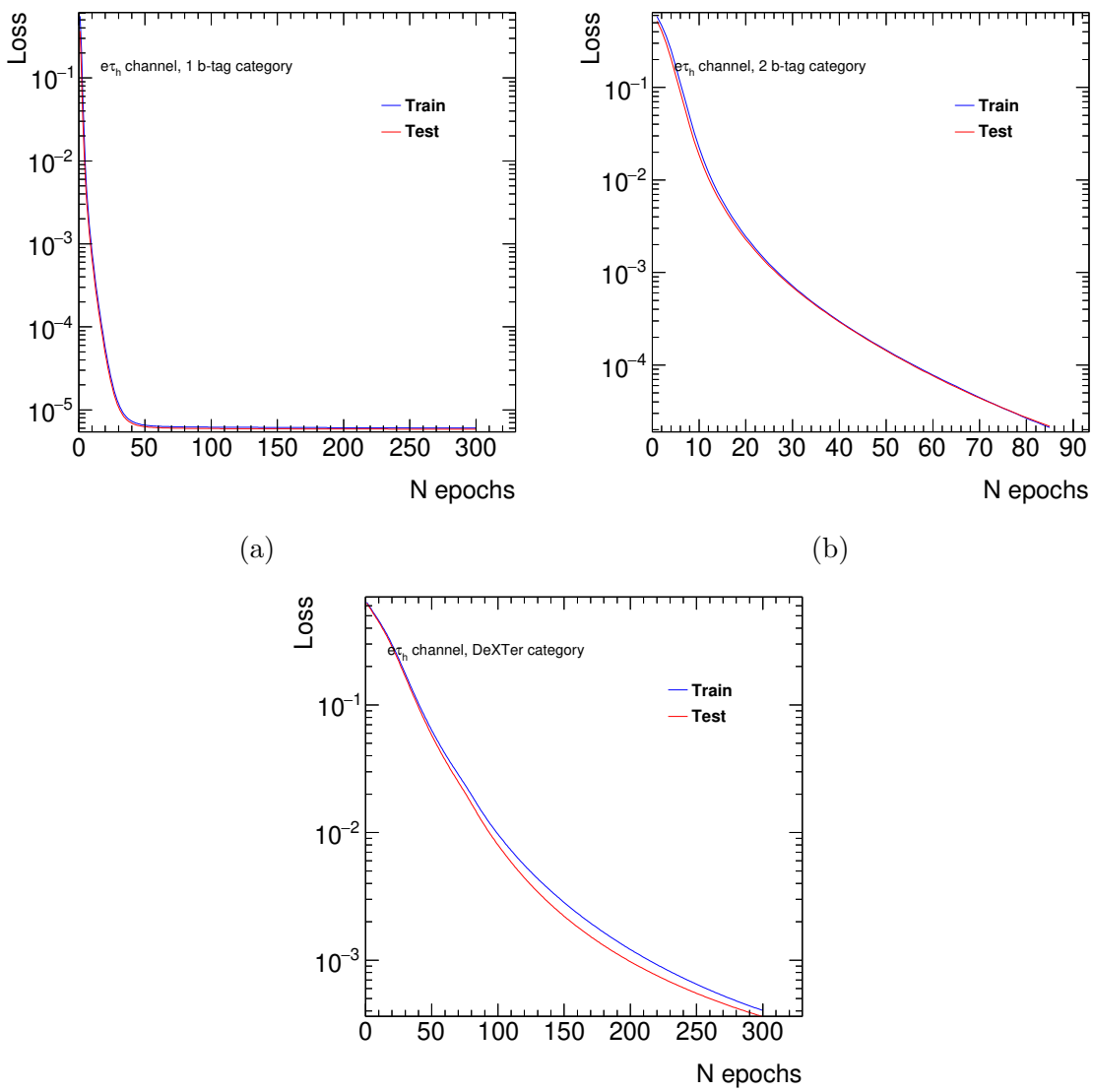


Figure A.55: Cross-entropy loss function in the $e\tau_{\text{had}}$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories.

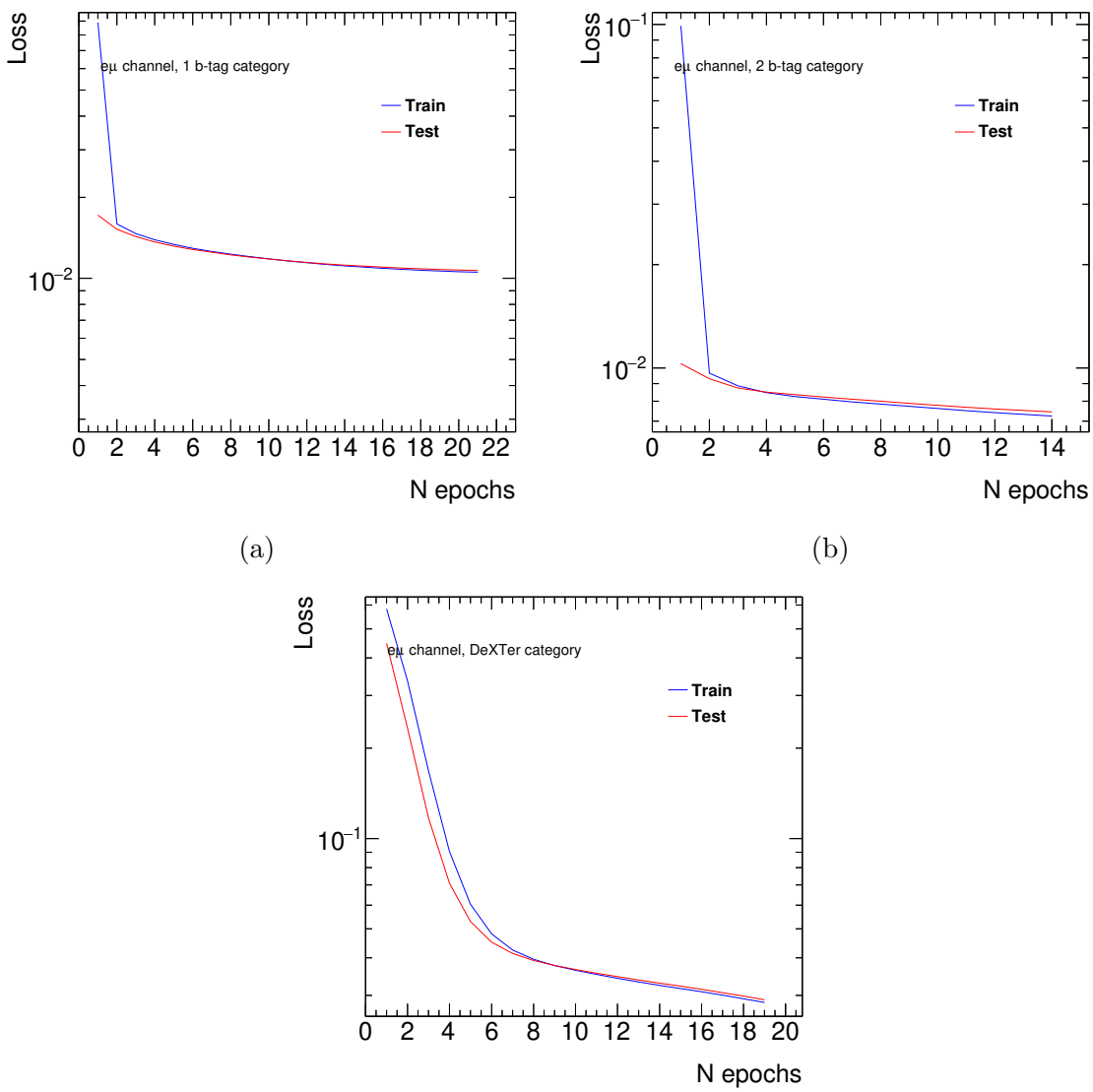
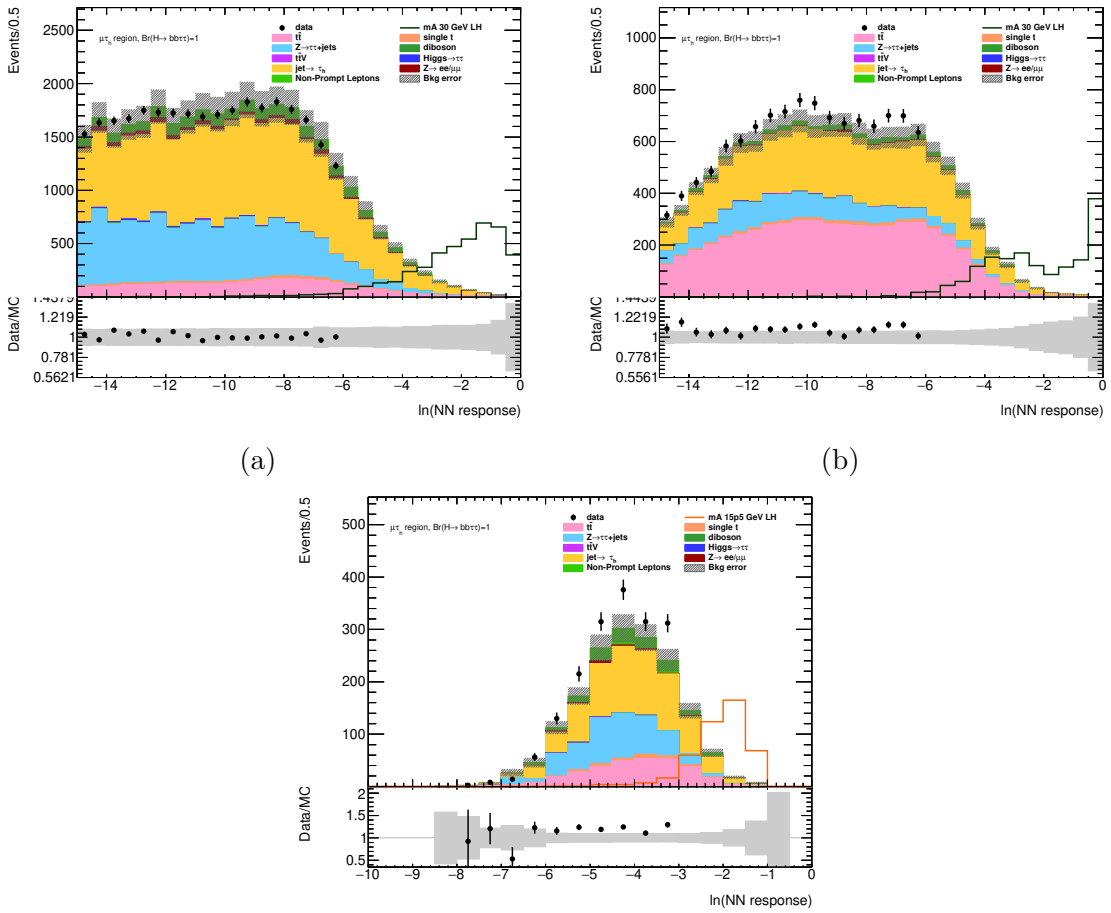


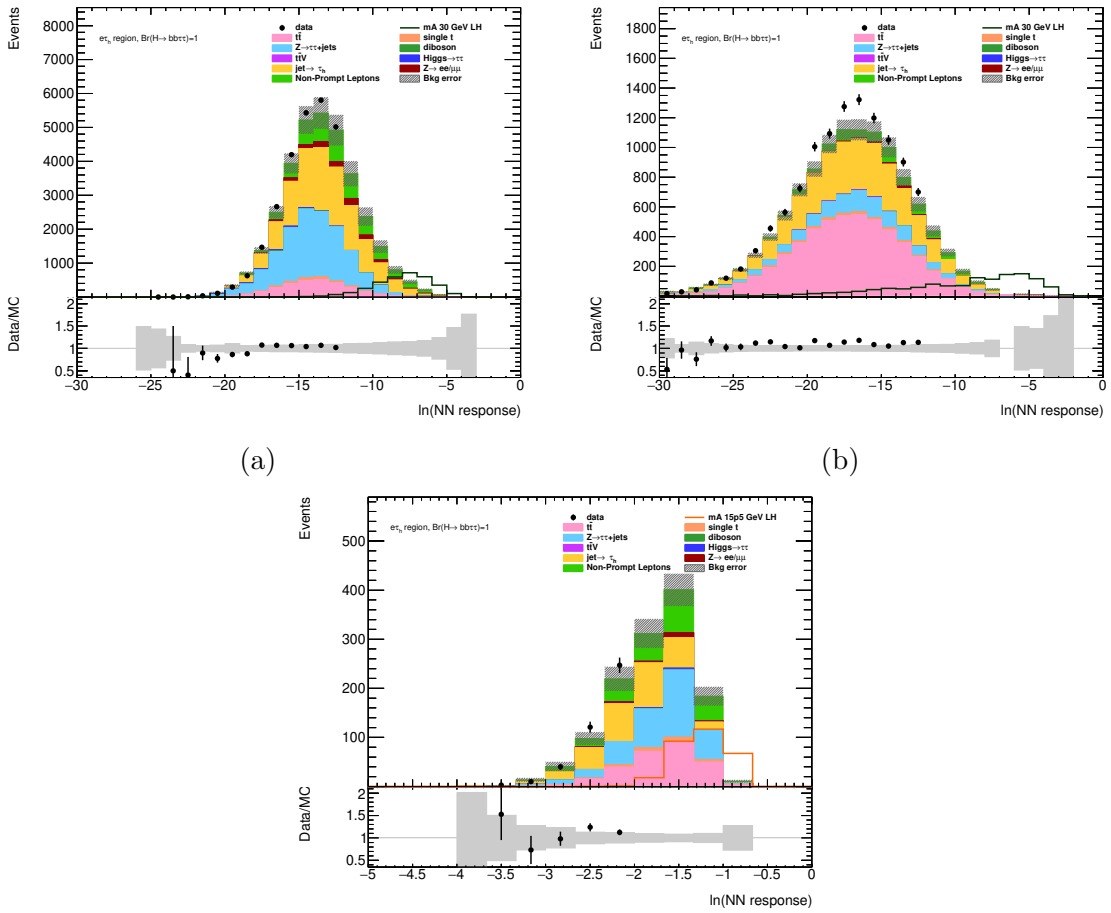
Figure A.56: Cross-entropy loss function in the $e\mu$ channel for (a) the 1 b-jet, (b) 2 b-jet and (c) DeXTer categories.



(a) (b)

(c)

Figure A.57: Network output distributions for the $\mu\tau_{\text{had}}$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded.



(a) (b)

(c)

Figure A.58: Network output distributions for the $e\tau_{\text{had}}$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded.

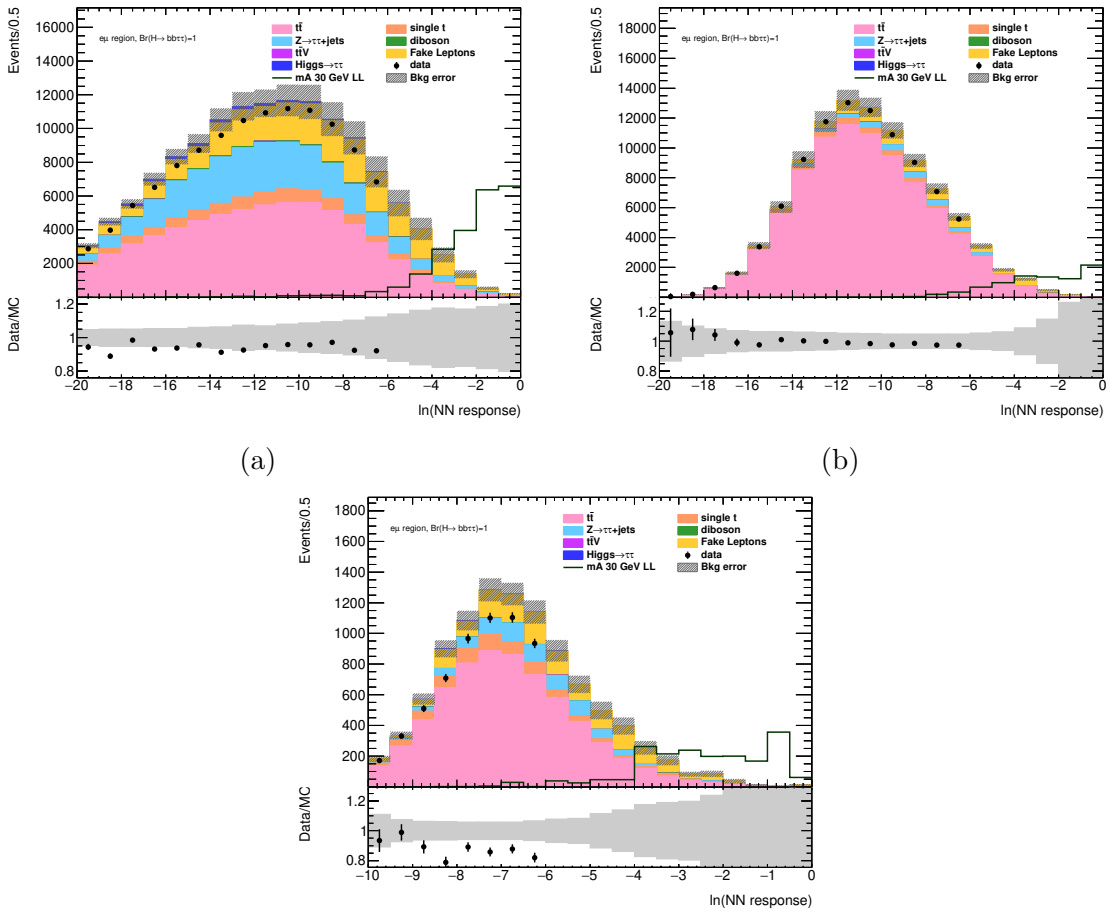
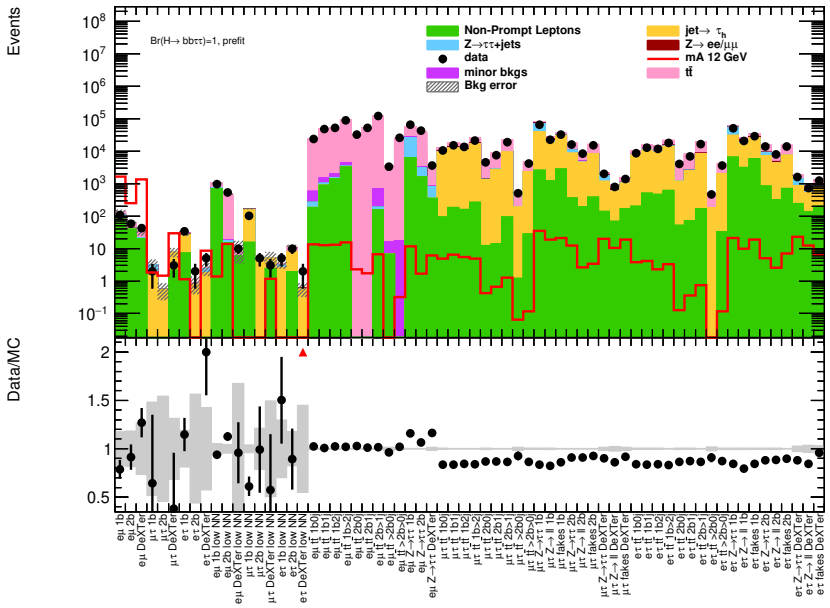


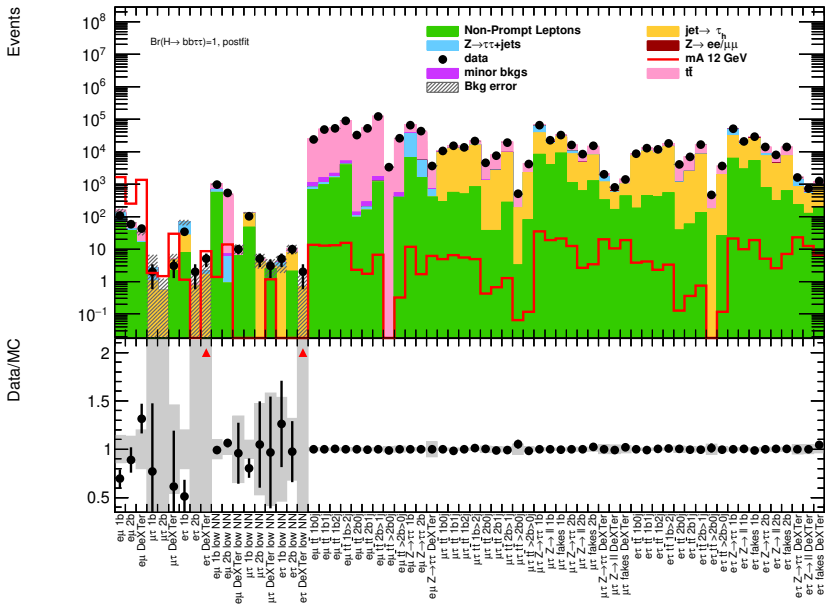
Figure A.59: Network output distributions for the $e\mu$ channel for (a) 1 b-jet, (b) > 1 b-jet and (c) DeXTer categories. The high output sideband, where the signal is expected, is blinded.

A.4 Model Validation

Additional pre-fit and post-fit model validation plots are presented for the $m_a = 12$, 15.5, 20, 25, 40, 50, 60 GeV mass points in Figures A.60–A.66. The NP pulls and correlations are shown in Figures A.67–A.73. The $m_a = 30$ GeV mass point is shown in the main body of the text in Figures 5.37 and 5.38.

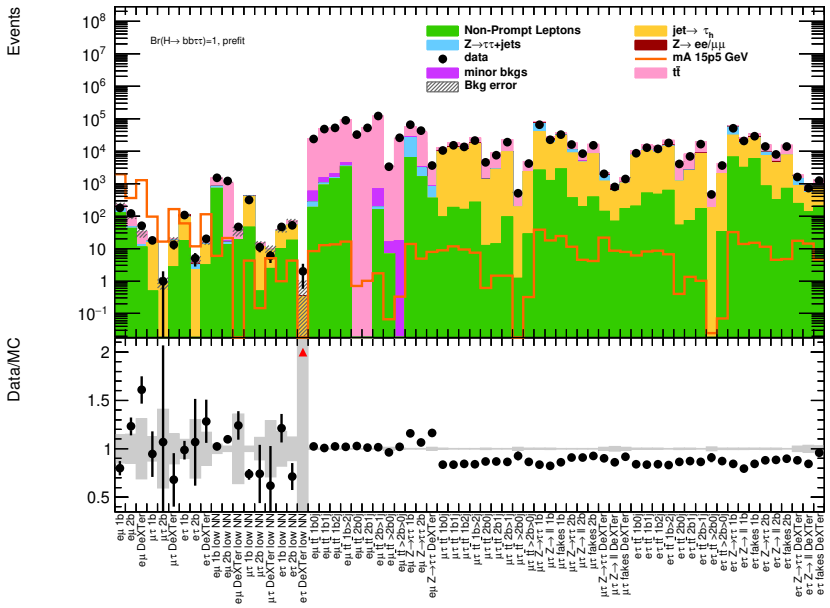


(a)

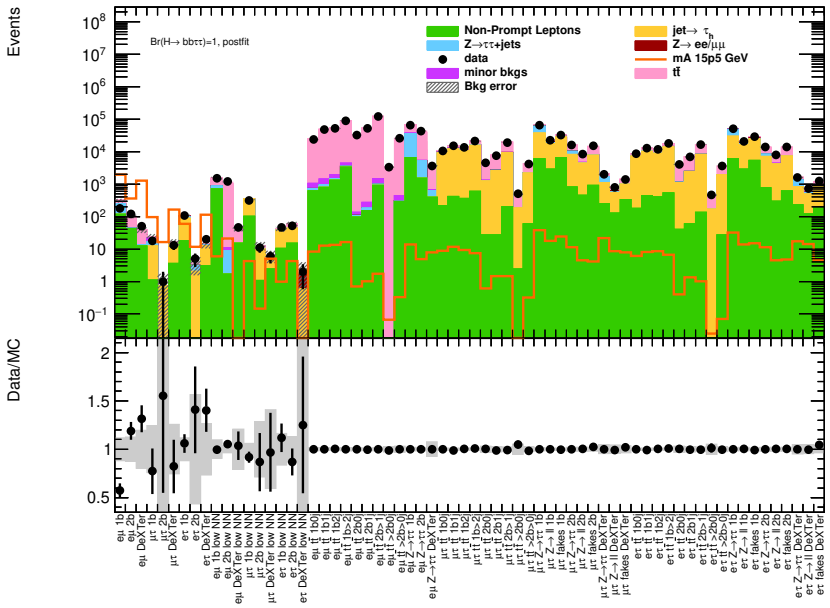


(b)

Figure A.60: (a) Prefit and (b) postfit plots for the $m_a = 12$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

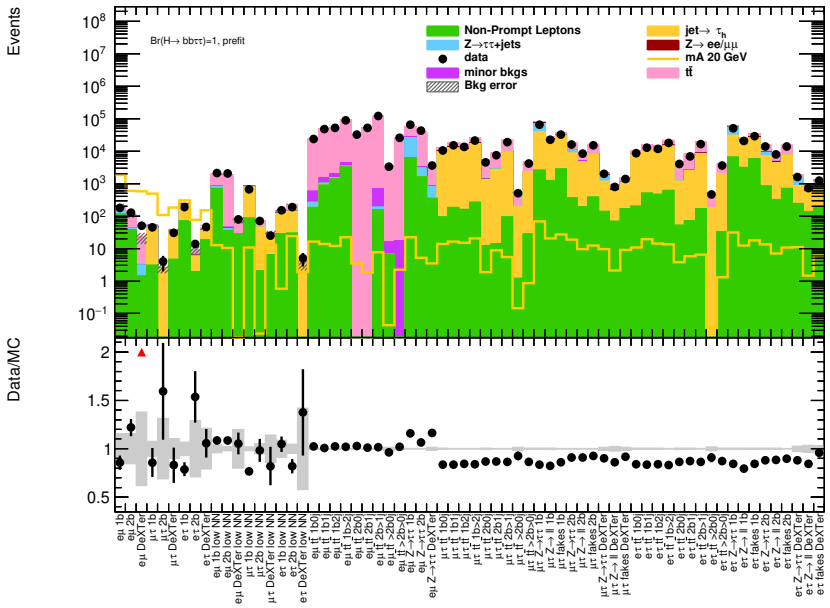


(a)

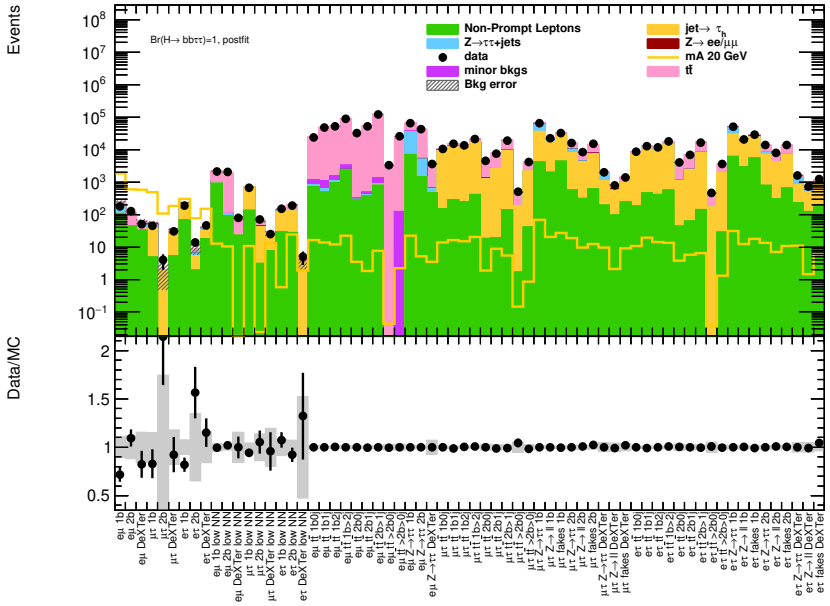


(b)

Figure A.61: (a) Prefit and (b) postfit plots for the $m_a = 15.5$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

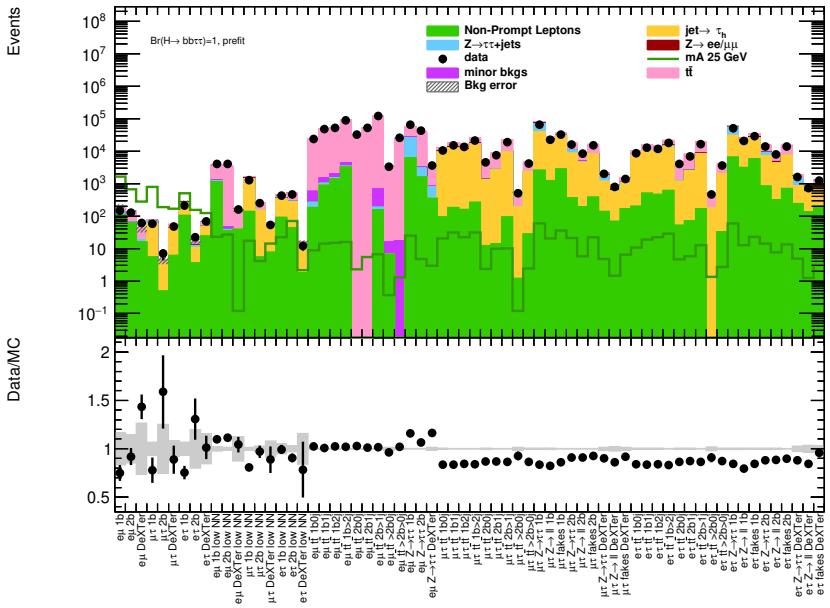


(a)

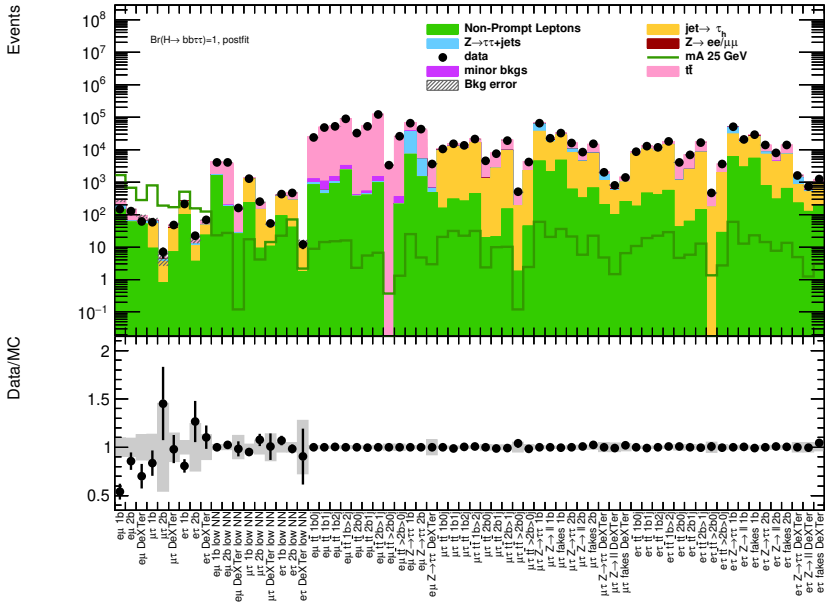


(b)

Figure A.62: (a) Prefit and (b) postfit plots for the $m_a = 20$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

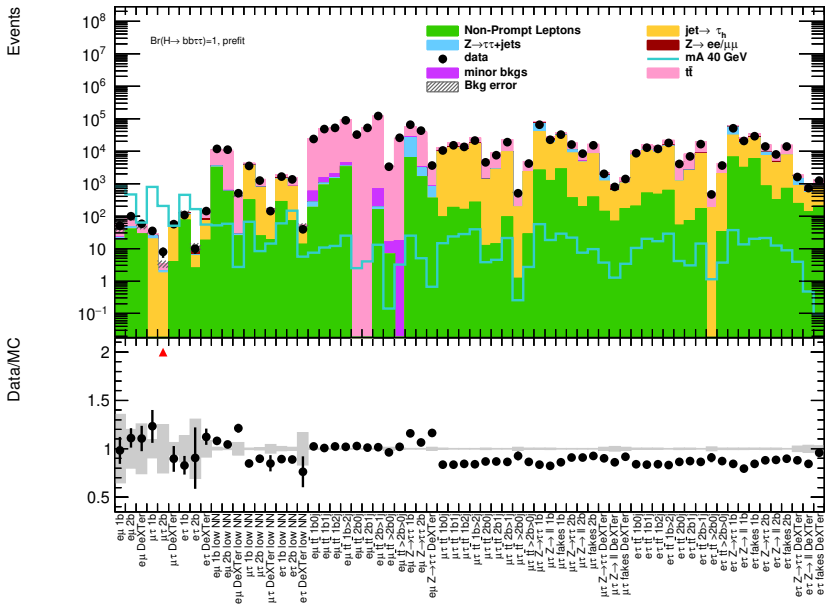


(a)

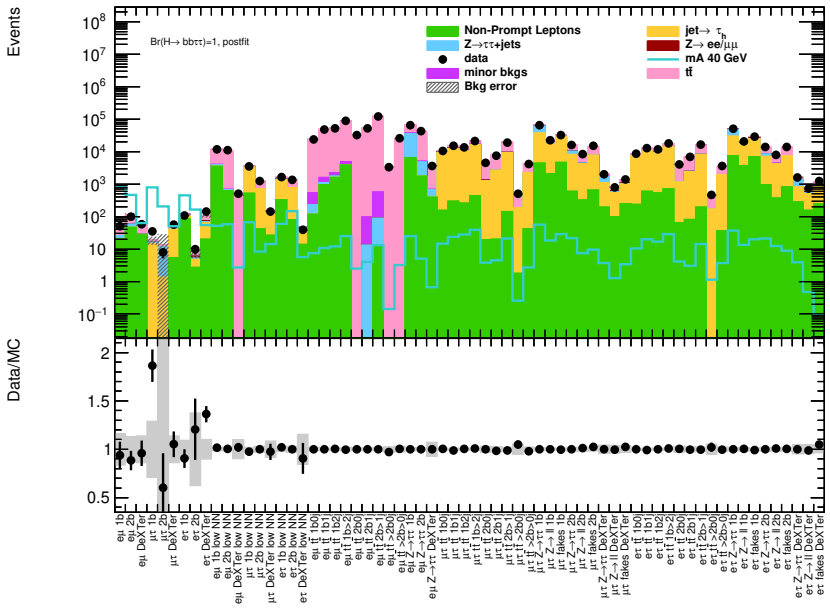


(b)

Figure A.63: (a) Prefit and (b) postfit plots for the $m_a = 25$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

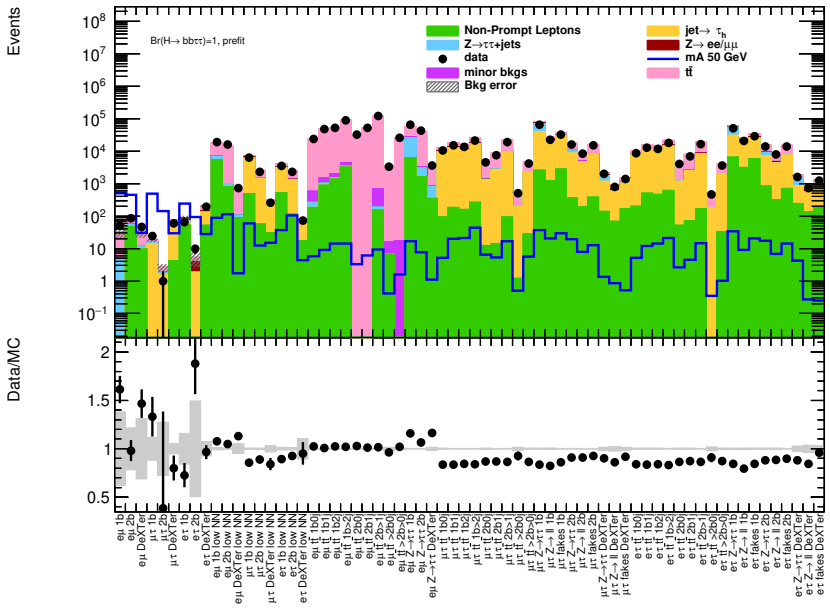


(a)

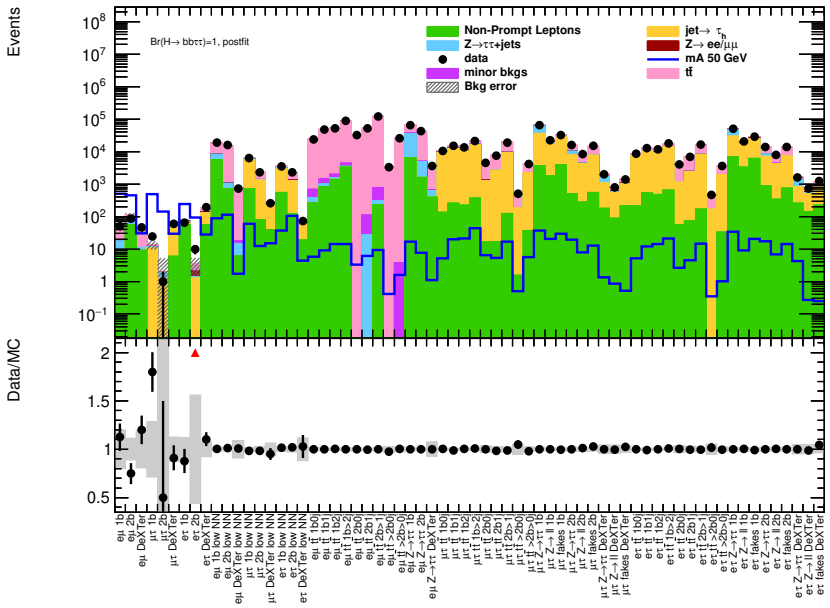


(b)

Figure A.64: (a) Prefit and (b) postfit plots for the $m_a = 40$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

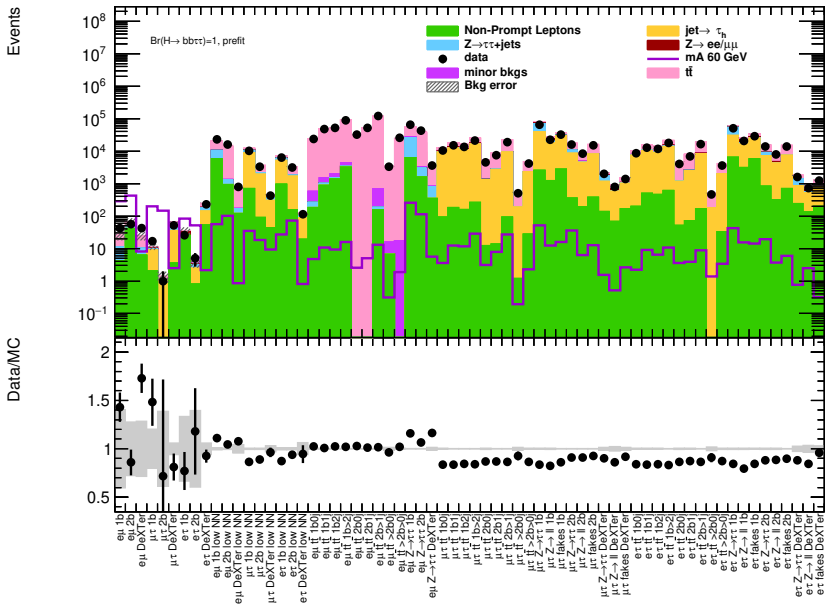


(a)

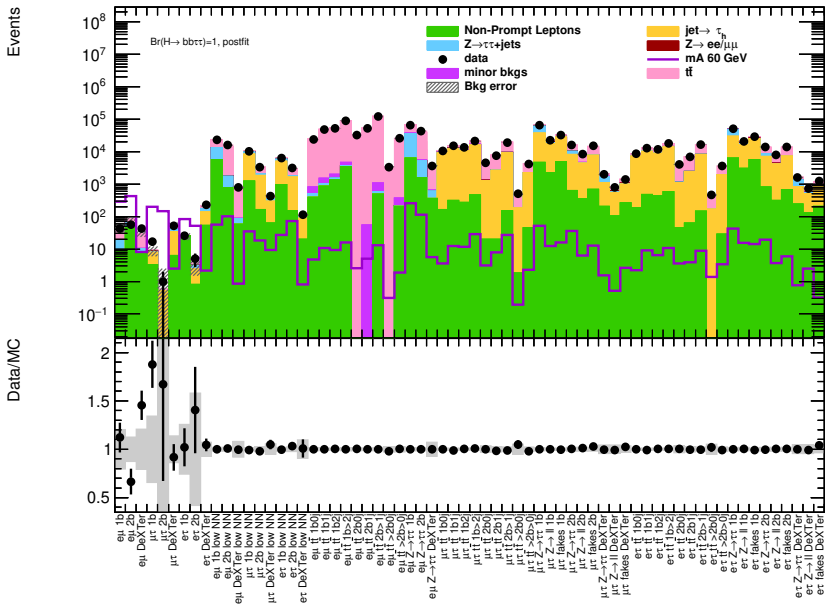


(b)

Figure A.65: (a) Prefit and (b) postfit plots for the $m_a = 50$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.

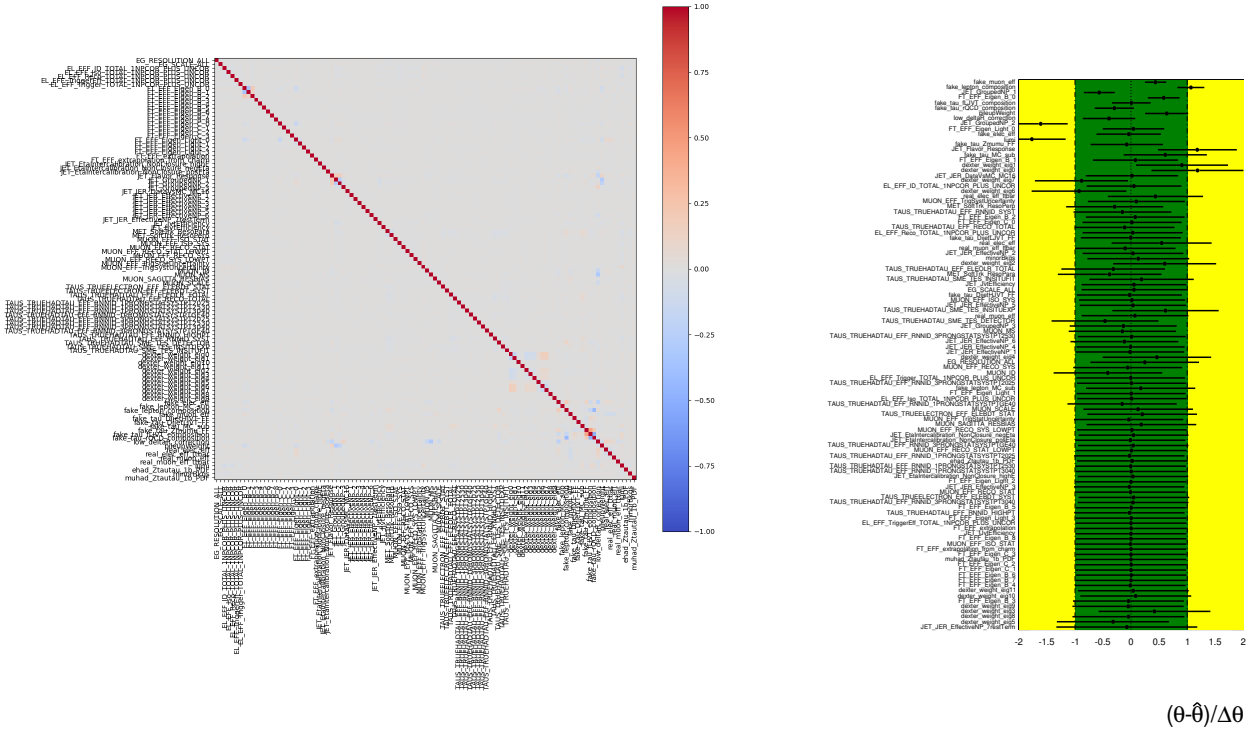


(a)



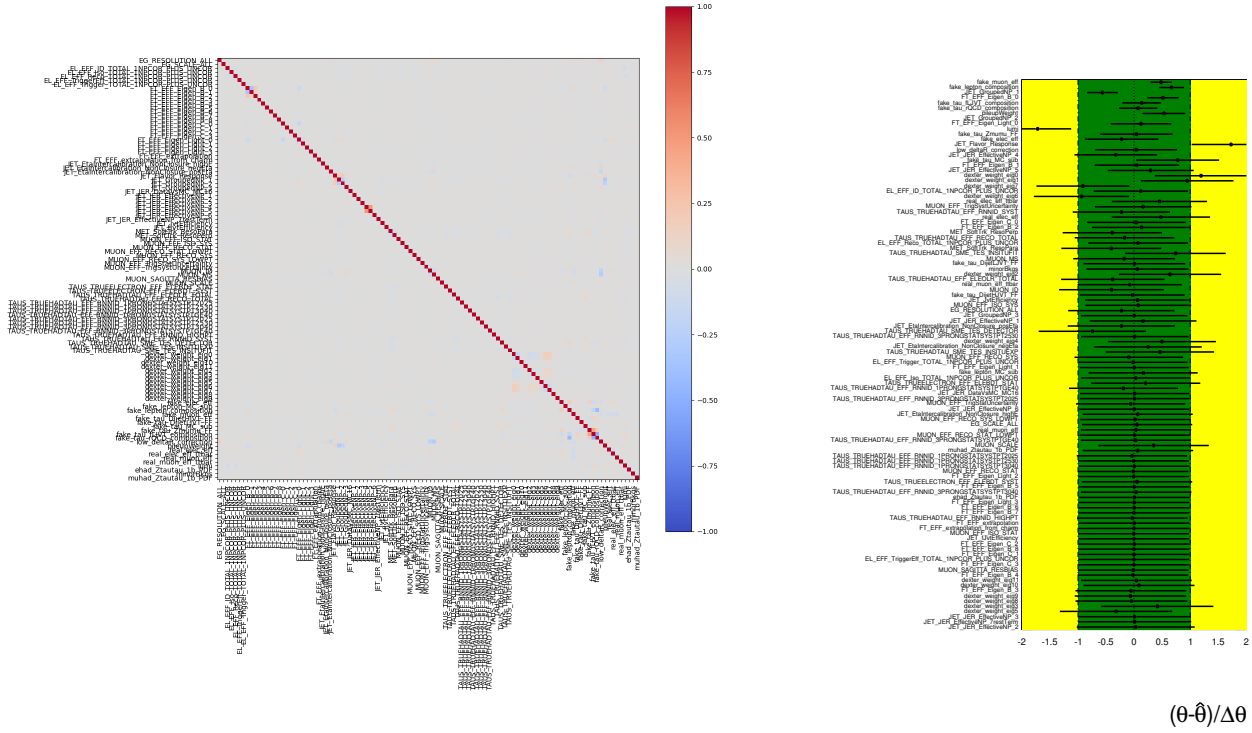
(b)

Figure A.66: (a) Prefit and (b) postfit plots for the $m_a = 60$ GeV mass point. The first 9 bins are the blinded SRs, the next 9 bins are the VRs in the low PNN discriminant sideband, and the rest of the bins are the CRs. The error bars on the prefit background prediction are the MC statistical error, a 25% error on fake backgrounds, and error on the unconstrained NPs fit in CRs for the $t\bar{t}$ and Z+jets reweighting. The error bars on the postfit background prediction are from the postfit NP covariance matrix.



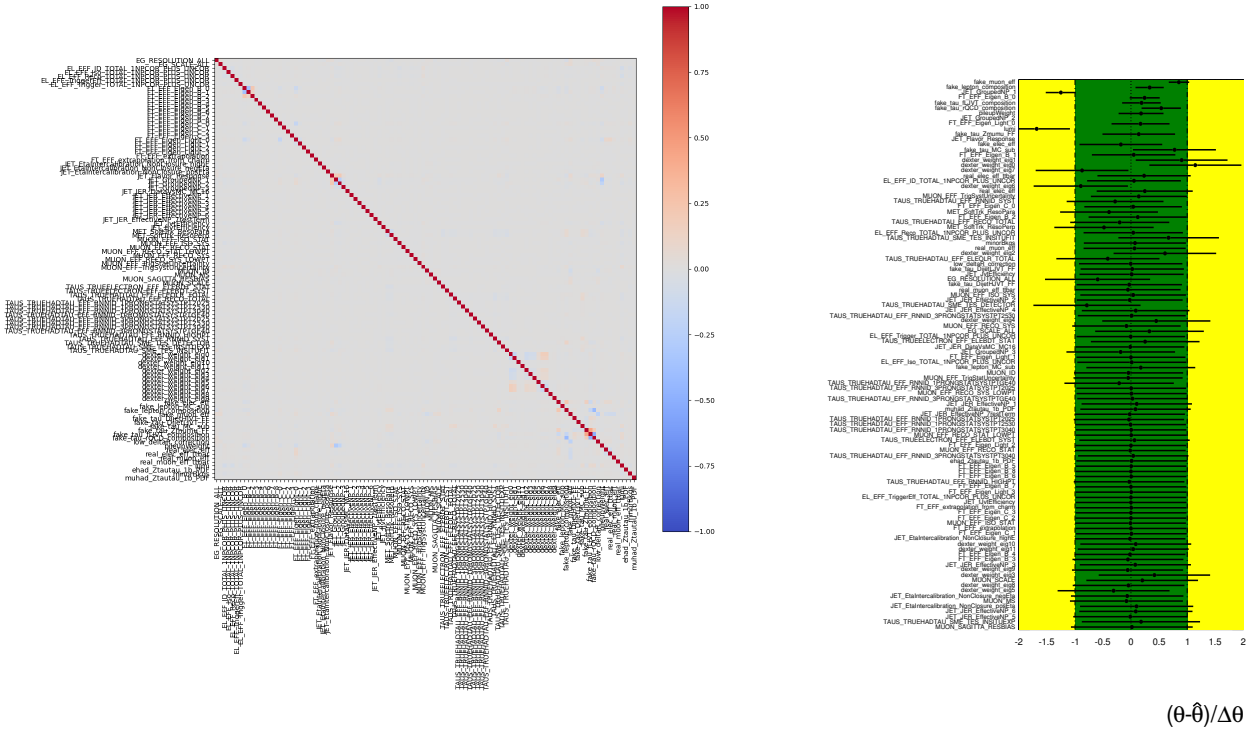
(a) (b)

Figure A.67: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 12$ GeV mass point.



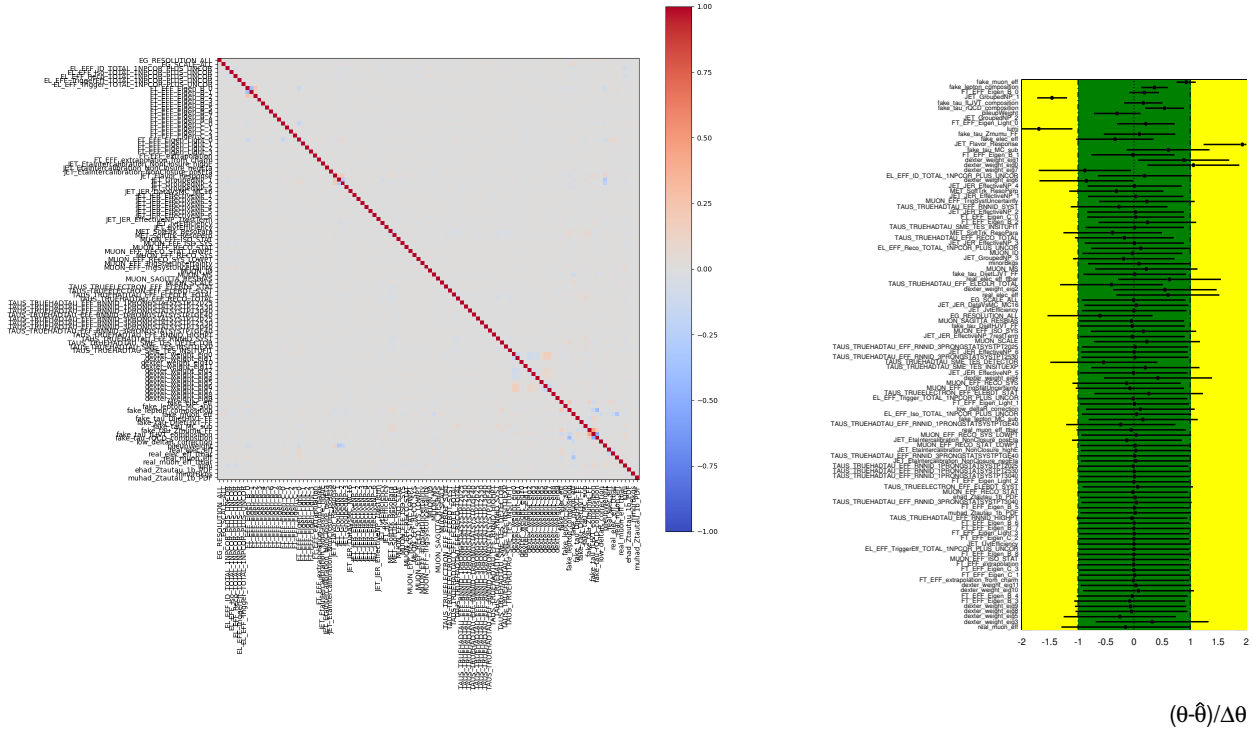
(a) (b)

Figure A.68: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 15.5$ GeV mass point.



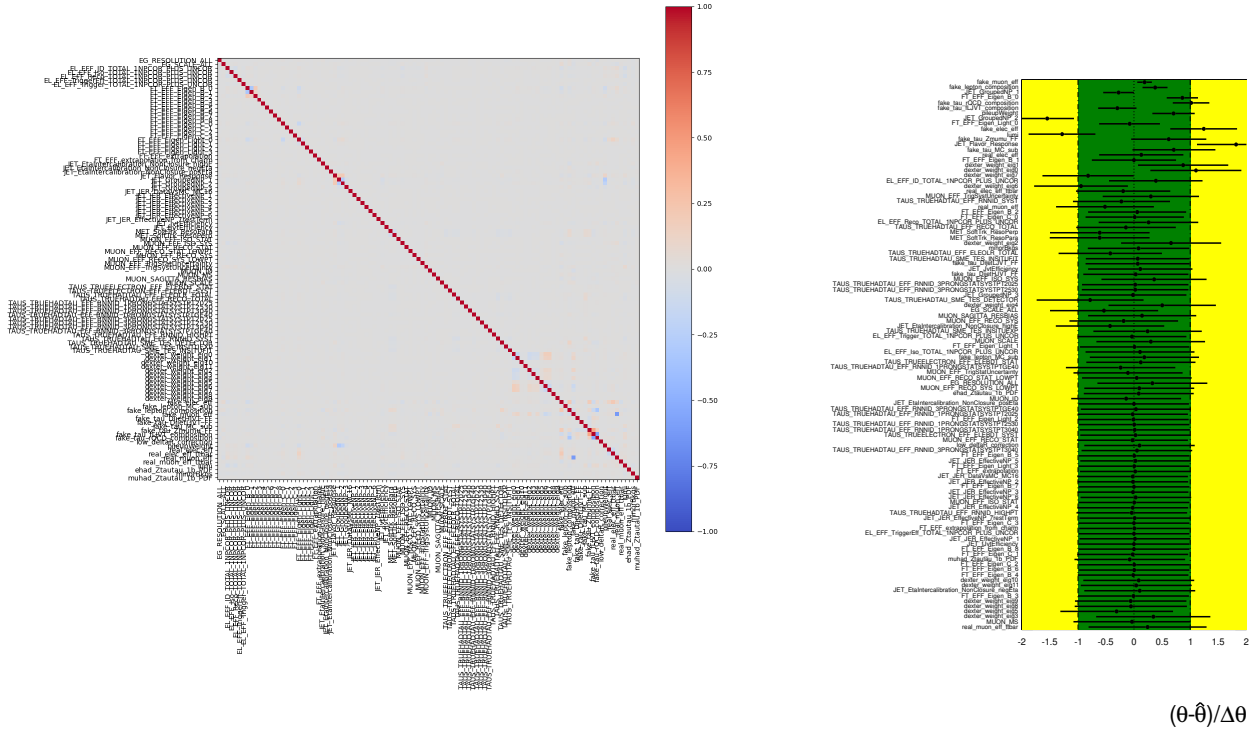
(a) (b)

Figure A.69: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 20$ GeV mass point.



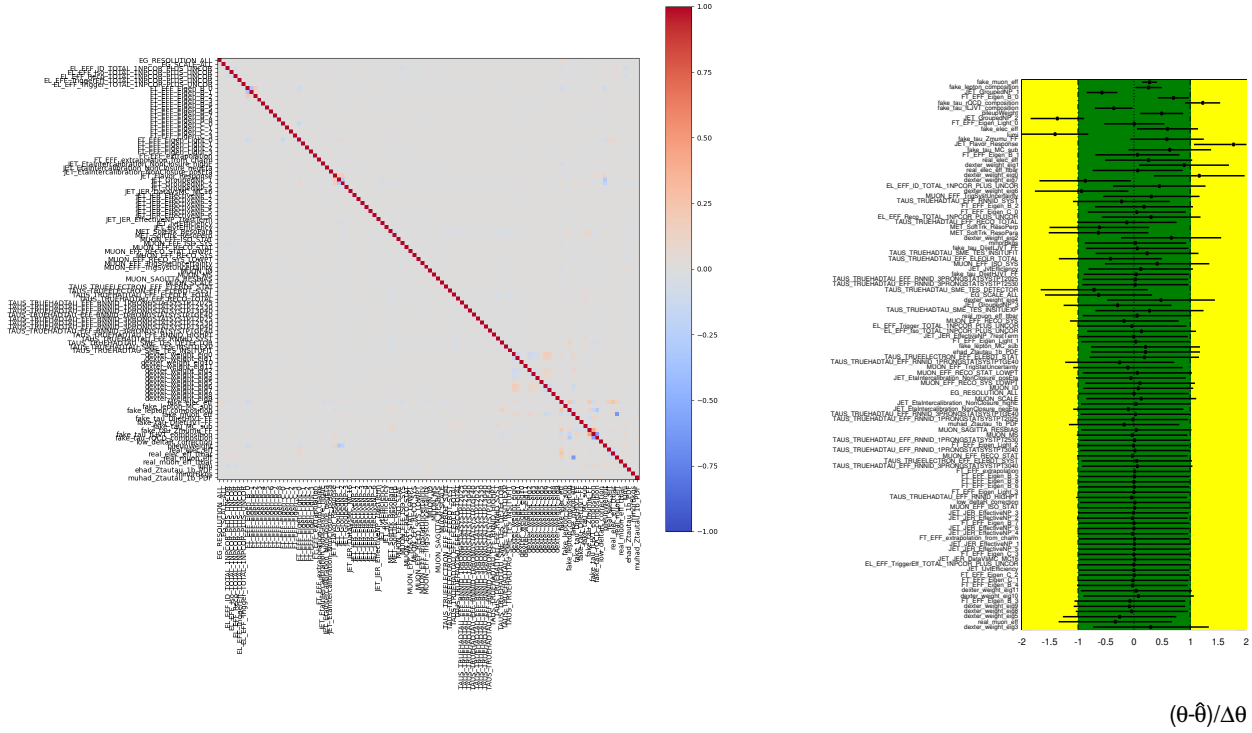
(a) (b)

Figure A.70: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 25$ GeV mass point.



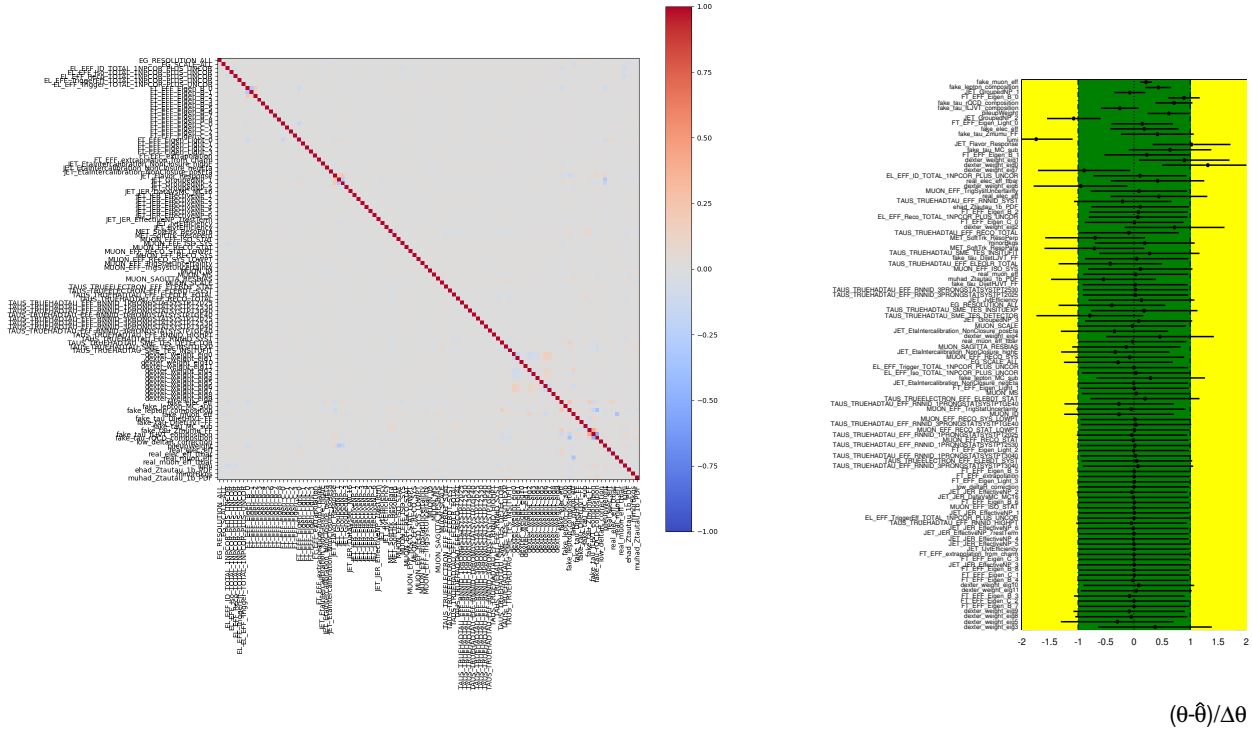
(a) (b)

Figure A.71: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 40$ GeV mass point.



(a) (b)

Figure A.72: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 50$ GeV mass point.



(a) (b)

Figure A.73: (a) Correlation matrix and (b) pull plot for the constrained nuisance parameters at the $m_a = 60$ GeV mass point.

APPENDIX B

Cutflow tables

This appendix presents cutflow tables for simulated signal and background samples. All cutflow tables present weighted event counts for each sample.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.21793e+07	1	—
Baseline selection	850681	0.0698466	0.0698466
Opposite sign	843565	0.0692624	0.991636
Both leptons signal	783927	0.0643657	0.929302
Any heavy flavor	709588	0.058262	0.905172
Veto ttbar CR	326573	0.0268139	0.460229
$m_{\tau\tau} < 60$GeV	106782	0.00876751	0.326977

Table B.1: Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.04988e+06	1	—
Baseline selection	80795.1	0.0769564	0.0769564
Opposite sign	80122.6	0.0763159	0.991677
Both leptons signal	74760.4	0.0712084	0.933075
Any heavy flavor	58932.7	0.0561328	0.788288
Veto ttbar CR	25804	0.024578	0.437855
$m_{\tau\tau} < 60$GeV	7709.44	0.00734316	0.298769

Table B.2: Weighted cutflow table for t full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	8.76096e+06	1	—
Baseline selection	54041.5	0.00616844	0.00616844
Opposite sign	49569.6	0.00565801	0.917251
Both leptons signal	46270.5	0.00528144	0.933445
Any heavy flavor	4546.65	0.000518967	0.0982624
Veto ttbar CR	2490.41	0.000284263	0.547747
$m_{\tau\tau} < 60$GeV	963.027	0.000109923	0.386693

Table B.3: Weighted cutflow table for diboson full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	206054	1	—
Baseline selection	2652.39	0.0128723	0.0128723
Opposite sign	1973.66	0.00957834	0.744106
Both leptons signal	1780.04	0.00863872	0.901901
Any heavy flavor	1623	0.00787657	0.911775
Veto ttbar CR	691.039	0.00335368	0.425779
$m_{\tau\tau} < 60$GeV	225.644	0.00109507	0.326528

Table B.4: Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	7.45203e+08	1	—
Baseline selection	680924	0.000913743	0.000913743
Opposite sign	677312	0.000908895	0.994695
Both leptons signal	579996	0.000778306	0.856321
Any heavy flavor	57587.9	7.72782e-05	0.0992901
Veto ttbar CR	56928.1	7.63928e-05	0.988543
$m_{\tau\tau} < 60$GeV	46140.2	6.19162e-05	0.810498

Table B.5: Weighted cutflow table for $Z \to \tau\tau$ full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	265270	1	—
Baseline selection	44848.4	0.169067	0.169067
Opposite sign	44606.2	0.168154	0.9946
Both leptons signal	39268.8	0.148033	0.880344
Any heavy flavor	5058.3	0.0190685	0.128812
Veto ttbar CR	4938.25	0.0186159	0.976268
$m_{\tau\tau} < 60$GeV	2993.27	0.0112838	0.606139

Table B.6: Weighted cutflow table for SM $H \to \tau\tau$ full Run 2 MC for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.8341e+06	1	—
Baseline selection	49416.7	0.0269433	0.0269433
Opposite sign	45336.1	0.0247184	0.917424
Both leptons signal	34813.3	0.0189811	0.767893
Any heavy flavor	23898.8	0.0130303	0.686486
Veto ttbar CR	23565.8	0.0128487	0.986064
$m_{\tau\tau} < 60$GeV	23332.5	0.0127215	0.9901

Table B.7: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 12$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.83133e+06	1	—
Baseline selection	64842.9	0.0354075	0.0354075
Opposite sign	59174.2	0.032312	0.912577
Both leptons signal	46364	0.025317	0.783517
Any heavy flavor	28865.1	0.0157618	0.622577
Veto ttbar CR	28544.6	0.0155868	0.988895
$m_{\tau\tau} < 60$GeV	28373.6	0.0154934	0.994011

Table B.8: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.81321e+06	1	—
Baseline selection	76493.7	0.0421868	0.0421868
Opposite sign	69367	0.0382564	0.906833
Both leptons signal	53857.4	0.0297027	0.776412
Any heavy flavor	34055.8	0.018782	0.632332
Veto ttbar CR	33485.1	0.0184673	0.983243
$m_{\tau\tau} < 60$GeV	33204.3	0.0183124	0.991615

Table B.9: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 20$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.79771e+06	1	—
Baseline selection	80515.9	0.0447881	0.0447881
Opposite sign	72806	0.0404994	0.904244
Both leptons signal	56981.5	0.0316968	0.782648
Any heavy flavor	36517.5	0.0203134	0.640866
Veto ttbar CR	35981.5	0.0200152	0.985323
$m_{\tau\tau} < 60$GeV	35745.3	0.0198838	0.993434

Table B.10: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 25$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.77763e+06	1	—
Baseline selection	79211.7	0.0445603	0.0445603
Opposite sign	71775.6	0.0403772	0.906124
Both leptons signal	56039.6	0.031525	0.780762
Any heavy flavor	36647	0.0206157	0.653948
Veto ttbar CR	36178.3	0.020352	0.987211
$m_{\tau\tau} < 60$GeV	35899.9	0.0201954	0.992304

Table B.11: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 30$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.70649e+06	1	—
Baseline selection	76265.6	0.0446916	0.0446916
Opposite sign	68351.6	0.040054	0.896232
Both leptons signal	50509.9	0.0295988	0.73897
Any heavy flavor	33171.5	0.0194385	0.656733
Veto ttbar CR	32548.7	0.0190735	0.981225
$m_{\tau\tau} < 60$GeV	32248.7	0.0188977	0.990784

Table B.12: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.65552e+06	1	—
Baseline selection	68430.8	0.0413349	0.0413349
Opposite sign	61216.8	0.0369773	0.894579
Both leptons signal	42857.8	0.0258878	0.700099
Any heavy flavor	29441.8	0.017784	0.686964
Veto ttbar CR	28907	0.017461	0.981838
$m_{\tau\tau} < 60$GeV	28547	0.0172435	0.987544

Table B.13: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.62147e+06	1	—
Baseline selection	57235.5	0.0352985	0.0352985
Opposite sign	51694.5	0.0318813	0.903191
Both leptons signal	35627.8	0.0219726	0.689199
Any heavy flavor	25838.9	0.0159355	0.725243
Veto ttbar CR	25296.7	0.0156011	0.979019
$m_{\tau\tau} < 60$GeV	25004.6	0.015421	0.988451

Table B.14: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $e\mu$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.25711e+08	1	—
Baseline selection and OS	147786	0.0011756	0.0011756
Tau signal	115282	0.000917039	0.780059
Any heavy flavor	103985	0.000827177	0.902009
Veto ttbar CR	2578.52	2.05115e-05	0.024797
$m_{\tau\tau} < 60$GeV	522.641	4.15748e-06	0.20269

Table B.15: Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	9.50629e+08	1	—
Baseline selection and OS	737517	0.00077582	0.00077582
Tau signal	605257	0.000636692	0.820669
Any heavy flavor	63018.4	6.62913e-05	0.104118
Veto ttbar CR	747.419	7.86237e-07	0.0118603
$m_{\tau\tau} < 60$GeV	249.536	2.62495e-07	0.333863

Table B.16: Weighted cutflow table for $Z \to \tau\tau$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.23572e+09	1	—
Baseline selection and OS	509965	0.000412688	0.000412688
Tau signal	139441	0.000112843	0.273433
Any heavy flavor	13661.4	1.10554e-05	0.0979723
Veto ttbar CR	130.662	1.05738e-07	0.00956433
$m_{\tau\tau} < 60$GeV	15.3288	1.24048e-08	0.117316

Table B.17: Weighted cutflow table for $Z \to \ell\ell$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	495492	1	—
Baseline selection and OS	23751.3	0.0479349	0.0479349
Tau signal	20121.3	0.0406087	0.847163
Any heavy flavor	2526.9	0.00509977	0.125583
Veto ttbar CR	21.2506	4.28879e-05	0.00840976
$m_{\tau\tau} < 60$GeV	3.55927	7.1833e-06	0.16749

Table B.18: Weighted cutflow table for SM $H \to \tau\tau$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	4.04313e+07	1	—
Baseline selection and OS	15782.1	0.000390344	0.000390344
Tau signal	12172.9	0.000301075	0.771308
Any heavy flavor	9525.77	0.000235604	0.782541
Veto ttbar CR	126.224	3.12193e-06	0.0132508
$m_{\tau\tau} < 60$GeV	26.299	6.5046e-07	0.208352

Table B.19: Weighted cutflow table for t full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.75219e+07	1	—
Baseline selection and OS	10032.9	0.00057259	0.00057259
Tau signal	8317.97	0.000474718	0.829072
Any heavy flavor	930.561	5.31084e-05	0.111874
Veto ttbar CR	10.5849	6.04096e-07	0.0113748
$m_{\tau\tau} < 60$GeV	3.51279	2.0048e-07	0.331868

Table B.20: Weighted cutflow table for diboson full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	412108	1	—
Baseline selection and OS	355.593	0.000862864	0.000862864
Tau signal	273.132	0.000662767	0.768101
Any heavy flavor	251.482	0.000610232	0.920734
Veto ttbar CR	8.68314	2.107e-05	0.0345279
$m_{\tau\tau} < 60$GeV	1.88179	4.56625e-06	0.216718

Table B.21: Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.70139e+06	1	—
Baseline selection and OS	2519.36	0.00148077	0.00148077
Tau signal	488.123	0.000286897	0.193749
Any heavy flavor	345.074	0.000202819	0.70694
Veto ttbar CR	9.21538	5.41638e-06	0.0267055
$m_{\tau\tau} < 60$GeV	2.85591	1.67857e-06	0.309907

Table B.22: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 12$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.70202e+06	1	—
Baseline selection and OS	5184.22	0.00304592	0.00304592
Tau signal	1872.28	0.00110003	0.361149
Any heavy flavor	1290.47	0.000758196	0.689249
Veto ttbar CR	27.9564	1.64254e-05	0.0216638
$m_{\tau\tau} < 60$GeV	19.7632	1.16116e-05	0.706929

Table B.23: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.69283e+06	1	—
Baseline selection and OS	10851.9	0.00641051	0.00641051
Tau signal	5552.72	0.00328013	0.511681
Any heavy flavor	3693.3	0.00218172	0.665133
Veto ttbar CR	102.027	6.02701e-05	0.027625
$m_{\tau\tau} < 60$GeV	92.4759	5.46278e-05	0.906383

Table B.24: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 20$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.67372e+06	1	—
Baseline selection and OS	15386.9	0.00919321	0.00919321
Tau signal	9105	0.00543998	0.591739
Any heavy flavor	5876.4	0.00351098	0.645404
Veto ttbar CR	173.588	0.000103714	0.0295399
$m_{\tau\tau} < 60$GeV	161.736	9.66328e-05	0.931723

Table B.25: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 25$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.64556e+06	1	—
Baseline selection and OS	17766.2	0.0107965	0.0107965
Tau signal	11098	0.00674424	0.624671
Any heavy flavor	7289.13	0.00442959	0.656795
Veto ttbar CR	244.546	0.00014861	0.0335494
$m_{\tau\tau} < 60$GeV	231.484	0.000140672	0.946586

Table B.26: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 30$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.56637e+06	1	—
Baseline selection and OS	17508.9	0.011178	0.011178
Tau signal	10895.3	0.00695579	0.622273
Any heavy flavor	7235.52	0.0046193	0.664094
Veto ttbar CR	247.414	0.000157954	0.0341944
$m_{\tau\tau} < 60$GeV	226.205	0.000144414	0.914278

Table B.27: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.44953e+06	1	—
Baseline selection and OS	13691.3	0.00944528	0.00944528
Tau signal	7924.3	0.00546679	0.578785
Any heavy flavor	5591.31	0.00385731	0.70559
Veto ttbar CR	218.781	0.000150932	0.0391288
$m_{\tau\tau} < 60$GeV	200.271	0.000138162	0.915395

Table B.28: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.28269e+06	1	—
Baseline selection and OS	10846.5	0.00845603	0.00845603
Tau signal	6024.22	0.00469655	0.555409
Any heavy flavor	4674.32	0.00364415	0.775922
Veto ttbar CR	169.411	0.000132074	0.0362428
$m_{\tau\tau} < 60$GeV	156.575	0.000122067	0.924232

Table B.29: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $\mu\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.25711e+08	1	—
Baseline selection and OS	125666	0.000999642	0.000999642
Tau signal	98030.7	0.00077981	0.780089
Any heavy flavor	88576.2	0.000704602	0.903556
Veto ttbar CR	42023.5	0.000334286	0.474433
$m_{\tau\tau} < 60$GeV	8565.37	6.81354e-05	0.203823

Table B.30: Weighted cutflow table for $t\bar{t}$ full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	9.50629e+08	1	—
Baseline selection and OS	524304	0.000551534	0.000551534
Tau signal	429845	0.000452169	0.819838
Any heavy flavor	46879	4.93137e-05	0.10906
Veto ttbar CR	43321	4.55709e-05	0.924101
$m_{\tau\tau} < 60$GeV	13713.9	1.44261e-05	0.316565

Table B.31: Weighted cutflow table for $Z \to \tau\tau$ full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.23572e+09	1	—
Baseline selection and OS	777155	0.000628911	0.000628911
Tau signal	166968	0.000135119	0.214845
Any heavy flavor	18732.1	1.51589e-05	0.11219
Veto ttbar CR	14208.4	1.14982e-05	0.758507
$m_{\tau\tau} < 60$GeV	2289.1	1.85245e-06	0.161108

Table B.32: Weighted cutflow table for $Z \to \ell\ell$.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	495492	1	—
Baseline selection and OS	19804.7	0.0399699	0.0399699
Tau signal	16799.9	0.0339054	0.848275
Any heavy flavor	2142.29	0.00432356	0.127518
Veto ttbar CR	1908.85	0.00385244	0.891033
$m_{\tau\tau} < 60$GeV	289.716	0.000584703	0.151775

Table B.33: Weighted cutflow table for SM $H \to \tau\tau$ full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	4.04313e+07	1	—
Baseline selection and OS	13690.3	0.000338606	0.000338606
Tau signal	10587.7	0.00026187	0.773375
Any heavy flavor	8311.94	0.000205582	0.785053
Veto ttbar CR	3702.64	9.15784e-05	0.44546
$m_{\tau\tau} < 60$GeV	738.503	1.82656e-05	0.199453

Table B.34: Weighted cutflow table for t full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.75219e+07	1	—
Baseline selection and OS	8396.71	0.000479212	0.000479212
Tau signal	6979.13	0.000398308	0.831174
Any heavy flavor	801.963	4.57692e-05	0.114909
Veto ttbar CR	415.981	2.37406e-05	0.518703
$m_{\tau\tau} < 60$GeV	119.775	6.83574e-06	0.287934

Table B.35: Weighted cutflow table for diboson full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	412108	1	—
Baseline selection and OS	298.106	0.00072337	0.00072337
Tau signal	229.12	0.00055597	0.768584
Any heavy flavor	210.341	0.000510401	0.918038
Veto ttbar CR	107.565	0.000261012	0.511386
$m_{\tau\tau} < 60$GeV	27.3584	6.63866e-05	0.254343

Table B.36: Weighted cutflow table for $t\bar{t}V$ full Run 2 MC for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.70139e+06	1	—
Baseline selection and OS	1485.67	0.000873209	0.000873209
Tau signal	261.638	0.000153779	0.176108
Any heavy flavor	185.02	0.000108746	0.70716
Veto ttbar CR	157.596	9.26278e-05	0.851778
$m_{\tau\tau} < 60$GeV	70.1547	4.12338e-05	0.445155

Table B.37: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 12$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.70202e+06	1	—
Baseline selection and OS	3447.43	0.00202549	0.00202549
Tau signal	1286.83	0.000756061	0.373273
Any heavy flavor	894.938	0.000525809	0.695458
Veto ttbar CR	811.507	0.00047679	0.906775
$m_{\tau\tau} < 60$GeV	669.433	0.000393316	0.824926

Table B.38: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 15.5$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.69283e+06	1	—
Baseline selection and OS	7142.05	0.00421899	0.00421899
Tau signal	3806.9	0.00224883	0.533026
Any heavy flavor	2616.45	0.0015456	0.687291
Veto ttbar CR	2429.25	0.00143502	0.928454
$m_{\tau\tau} < 60$GeV	2266.78	0.00133905	0.93312

Table B.39: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 20$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.67372e+06	1	—
Baseline selection and OS	10851	0.00648319	0.00648319
Tau signal	6731.82	0.00402207	0.620385
Any heavy flavor	4417.79	0.0026395	0.656254
Veto ttbar CR	4112.94	0.00245737	0.930996
$m_{\tau\tau} < 60$GeV	3934.5	0.00235075	0.956614

Table B.40: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 25$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.64556e+06	1	—
Baseline selection and OS	13418.4	0.00815433	0.00815433
Tau signal	8603.64	0.00522841	0.641182
Any heavy flavor	5612.44	0.00341067	0.652334
Veto ttbar CR	5295.65	0.00321815	0.943554
$m_{\tau\tau} < 60$GeV	5085.97	0.00309073	0.960406

Table B.41: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 30$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.56637e+06	1	—
Baseline selection and OS	13566.3	0.008661	0.008661
Tau signal	8543.42	0.00545429	0.629752
Any heavy flavor	5745.29	0.0036679	0.672481
Veto ttbar CR	5411.94	0.00345509	0.941979
$m_{\tau\tau} < 60$GeV	5237.75	0.00334389	0.967815

Table B.42: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 40$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.44953e+06	1	—
Baseline selection and OS	10122.4	0.00698317	0.00698317
Tau signal	5841.86	0.00403017	0.577125
Any heavy flavor	4025.89	0.00277737	0.689145
Veto ttbar CR	3774.43	0.00260389	0.937538
$m_{\tau\tau} < 60$GeV	3612.98	0.00249251	0.957226

Table B.43: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 50$ GeV for the $e\tau_{\text{had}}$ final state.

Cut	Sum Weights	Total Eff.	Relative Eff.
AOD Filtered	1.28269e+06	1	—
Baseline selection and OS	7297.9	0.00568952	0.00568952
Tau signal	3969.54	0.0030947	0.543929
Any heavy flavor	3002.64	0.00234089	0.756419
Veto ttbar CR	2770.45	0.00215988	0.922674
$m_{\tau\tau} < 60$GeV	2632.51	0.00205233	0.950209

Table B.44: Weighted cutflow table for $H \to aa \to bb\tau\tau$ MC at $m_a = 60$ GeV for the $e\tau_{\text{had}}$ final state.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. Curtin, R. Essig, S. Gori, P. Jaiswal, A. Katz, T. Liu et al., *Exotic decays of the 125 GeV Higgs boson*, *Physical Review D* **90** (oct, 2014) .

[2] H. Wiedemann, *Particle Accelerator Physics*. Springer, 1993.

[3] E. Mobs, *The CERN accelerator complex in 2019*, CERN-GRAPHICS-2019-002-1, 2020.

[4] ATLAS collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[5] ALICE collaboration, *The ALICE Experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[6] ATLAS collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[7] A. La Rosa, *The atlas insertable b-layer: from construction to operation*, *Journal of Instrumentation* **11** (2016) C12036.

[8] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) 317, [1611.09661].

[9] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) T05008, [1803.00844].

[10] *Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at* $\sqrt{s} = 13$ *TeV*, *JINST* **16** (2021) P07029.

[11] ATLAS collaboration, *Performance of the muon spectrometer alignment in 2017 and 2018 data*, ATL-MUON-PUB-2021-002, 2021.

[12] ATLAS collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *The European Physical Journal C* **80** (2020) .

[13] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **79** (2019) 639, [1902.04655].

[14] ATLAS collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J., C* **81** (2021) 578, [2012.00578].

[15] ATLAS Collaboration, "Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment." ATL-PHYS-PUB-2019-033, 2019.

[16] C. Wei, A. Chen, D. Amidei, N. Anderson, E. Carpenter, L. Cooperrider et al., *Construction and testing of smdt tubes at the university of michigan for the atlas muon spectrometer upgrade*, *Journal of Instrumentation* **17** (oct, 2022) P10010.

[17] Y. Arai, B. Ball, M. Beretta, H. Boterenbrood, G. W. Brandenburg, F. Ceradini et al., *Atlas muon drift tube electronics*, *JINST* **3** (sep, 2008) P09001.

[18] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98** (2018) 030001.

[19] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, 1999.

[20] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, 1996.

[21] ATLAS collaboration, *Search for Higgs boson decays into a pair of pseudoscalar particles in the bbμμ final state with the atlas detector in pp collisions at* $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **105** (2022) 012006.

[22] S. L. Glashow, *Partial-symmetries of weak interactions*, *Nuclear Physics* **22** (1961) 579–588.

[23] S. Weinberg, *A model of leptons*, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266.

[24] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367–377.

[25] M. D. Schwartz, *Quantum Field Theory and the Standard Model*, p. 110. Cambridge University Press, 3, 2014.

[26] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Physics Letters B* **716** (2012) 1–29.

[27] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* **716** (2012) 30–61.

[28] ATLAS collaboration, *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) 52–59.

[29] CMS collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60–68.

[30] H. Yukawa, *On the Interaction of Elementary Particles. I*, *Progress of Theoretical Physics Supplement* **1** (01, 1955) 1–10, [https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf].

[31] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, *Phys. Rev.* **96** (Oct, 1954) 191–195.

[32] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323.

[33] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132–133.

[34] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509.

[35] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, *Phys. Rev.* **145** (May, 1966) 1156–1163.

[36] K. Freese, *Review of observational evidence for dark matter in the universe and in upcoming searches for dark stars*, *EAS Publications Series* **36** (2009) 113–126.

[37] Muon $g-2$ Collaboration collaboration, *Measurement of the positive muon anomalous magnetic moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (Apr, 2021) 141801.

[38] *High-precision measurement of the w boson mass with the cdf ii detector*, *Science* **376** (2022) 170–176, [https://www.science.org/doi/pdf/10.1126/science.abk1781].

[39] T. D. Lee, *A theory of spontaneous t violation*, *Phys. Rev. D* **8** (Aug, 1973) 1226–1239.

[40] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher and J. P. Silva, *Theory and phenomenology of two-higgs-doublet models*, *Physics Reports* **516** (jul, 2012) 1–102.

[41] L. Evans and P. Bryant, *LHC machine*, *JINST* **3** (2008) S08001.

[42] J. Jowett, *Colliding Heavy Ions in the LHC*, CERN-ACC-2018-121, 2018.

[43] M. Vretenar, J. Vollaire, R. Scrivens, C. Rossi, F. Roncarolo, S. Ramberger et al., *Linac4 design report*, CYRM-2020-006, Geneva, 2020. 10.23731/CYRM-2020-006.

[44] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[45] LHCb collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[46] ATLAS collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, CERN, Geneva, 2019.

[47] ATLAS collaboration, *ATLAS inner detector: Technical Design Report, 1*, CERN-LHCC-97-016, Geneva, 1997.

[48] K. Kleinknecht, *Detectors for particle radiation.* Cambridge Univ. Press, Cambridge, 1986.

[49] C. Biino, *The cms electromagnetic calorimeter: overview, lessons learned during run 1 and future projections*, *Journal of Physics: Conference Series* **587** (feb, 2015) 012001.

[50] ATLAS collaboration, G. Aad et al., *Resolution of the ATLAS muon spectrometer monitored drift tubes in LHC Run 2*, *JINST* **14** (2019) P09011.

[51] I. B. Alonso et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*, CYRM-2020-0010, Geneva, 2020. 10.23731/CYRM-2020-0010.

[52] ATLAS collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer*, CERN-LHCC-2017-017.

[53] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *JINST* **15** (2020) P09015, [2004.13447].

[54] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) 292, [1603.05598].

[55] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** (2021) 578, [2012.00578].

[56] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, [1703.10485].

[57] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [0802.1189].

[58] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, 2211.16345.

[59] ATLAS collaboration, *Identification of hadronic tau lepton decays using neural networks in the atlas experiment*, ATL-PHYS-PUB-2019-033.

[60] ATLAS collaboration, *DeXTer: Deep Sets based Neural Networks for Low-p_T $X \to bb$ Identification in ATLAS*, ATL-PHYS-PUB-2022-042.

[61] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse and A. Salzburger, *The new ATLAS track reconstruction (NEWT)*, *J. Phys.: Conf. Ser.* **119** (2008) 032014.

[62] ATLAS Collaboration, "Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment." ATLAS-CONF-2012-042, 2012.

[63] ATLAS collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, *JINST* **14** (2019) P12006, [1908.00005].

[64] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J. C* **74** (2014) 3071, [1407.5063].

[65] ATLAS Collaboration, *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ *TeV*, *Eur. Phys. J. C* **75** (2015) 303, [1412.7086].

[66] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton–proton collisions at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1104, [1910.04482].

[67] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, *JINST* **14** (2019) P03017, [1812.03848].

[68] ATLAS Collaboration, "Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV." ATL-PHYS-PUB-2015-015, 2015.

[69] ATLAS Collaboration, "Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration." ATL-PHYS-PUB-2020-009, 2020.

[70] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with* $t\bar{t}$ *events in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **79** (2019) 970, [1907.05120].

[71] ATLAS Collaboration, "Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment." ATL-PHYS-PUB-2017-003, 2017.

[72] ATLAS Collaboration, "Commissioning of the ATLAS high performance b-tagging algorithms in the 7 TeV collision data." ATLAS-CONF-2011-102, 2011.

[73] ATLAS Collaboration, "Secondary vertex finding for jet flavour identification with the ATLAS detector." ATL-PHYS-PUB-2017-011, 2017.

[74] ATLAS Collaboration, "Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC." ATL-PHYS-PUB-2018-025, 2018.

[75] ATLAS Collaboration, "Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV." ATLAS-CONF-2017-029, 2017.

[76] ATLAS Collaboration, *Reconstruction, Identification, and Calibration of hadronically decaying tau leptons with the ATLAS detector for the LHC Run 3 and reprocessed Run 2 data*, .

[77] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) P04008, [1512.01094].

[78] ATLAS Collaboration, "Measurement of the Mis-identification Probability of τ Leptons from Hadronic Jets and from Electrons." ATLAS-CONF-2011-113, 2011.

[79] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, [1905.09127].

[80] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, [0808.3674].

[81] F. Buccioni, J.-N. Lang, J. M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866, [1907.13071].

[82] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, [1111.5206].

[83] A. Denner, S. Dittmaier and L. Hofer, *Collier: A fortran-based complex one-loop library in extended regularizations*, *Comput. Phys. Commun.* **212** (2017) 220–238, [1604.06792].

[84] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* **03** (2008) 038, [0709.1027].

[85] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, [1111.1220].

[86] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) 027, [1207.5030].

[87] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*, *JHEP* **11** (2001) 063, [hep-ph/0109231].

[88] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, [0903.1219].

[89] NNPDF collaboration, R. D. Ball et al., *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040, [1410.8849].

[90] C. Anastasiou, L. Dixon, K. Melnikov and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, *Phys. Rev. D* **69** (2004) 094008, [hep-ph/0312266].

[91] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, [0707.3088].

[92] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [hep-ph/0409146].

[93] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [0709.2092].

[94] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, [1002.2581].

[95] ATLAS Collaboration, "Studies on top-quark Monte Carlo modelling for Top2016." ATL-PHYS-PUB-2016-020, 2016.

[96] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, [1410.3012].

[97] ATLAS Collaboration, "ATLAS Pythia 8 tunes to 7 TeV data." ATL-PHYS-PUB-2014-021, 2014.

[98] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, [1207.1303].

[99] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152.

[100] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111, [0907.4076].

[101] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130, [1207.5391].

[102] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547, [1009.2450].

[103] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, *JHEP* **07** (2008) 029, [0805.3067].

[104] K. Hamilton, P. Nason, E. Re and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, *JHEP* **10** (2013) 222, [1309.0017].

[105] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, *JHEP* **05** (2015) 140, [1501.04637].

[106] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*, *JHEP* **10** (2012) 155, [1206.3572].

[107] J. M. Campbell, R. K. Ellis, R. Frederix, P. Nason, C. Oleari and C. Williams, *NLO Higgs boson production plus one and two jets using the POWHEG BOX, MadGraph4 and MCFM*, *JHEP* **07** (2012) 092, [1202.5475].

[108] K. Hamilton, P. Nason, C. Oleari and G. Zanderighi, *Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching*, *JHEP* **05** (2013) 082, [1212.4504].

[109] S. Catani and M. Grazzini, *Next-to-Next-to-Leading-Order Subtraction Formalism in Hadron Collisions and its Application to Higgs-boson Production at the Large Hadron Collider*, *Phys. Rev. Lett.* **98** (2007) 222002, [hep-ph/0703012].

[110] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43** (2016) 023001, [1510.03865].

[111] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09** (2014) 145, [1406.3660].

[112] LHC Higgs Cross Section Working Group collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[113] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog et al., *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, *JHEP* **05** (2016) 058, [1602.00695].

[114] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog and B. Mistlberger, *Higgs Boson Gluon-Fusion Production in QCD at Three Loops*, *Phys. Rev. Lett.* **114** (2015) 212001, [1503.06056].

[115] F. Dulat, A. Lazopoulos and B. Mistlberger, *iHixs 2 – Inclusive Higgs cross sections*, *Comput. Phys. Commun.* **233** (2018) 243–260, [1802.00827].

[116] R. V. Harlander and K. J. Ozeren, *Finite top mass effects for hadronic Higgs production at next-to-next-to-leading order*, *JHEP* **11** (2009) 088, [0909.3420].

[117] R. V. Harlander and K. J. Ozeren, *Top mass effects in Higgs production at next-to-next-to-leading order QCD: Virtual corrections*, *Phys. Lett. B* **679** (2009) 467–472, [0907.2997].

[118] R. V. Harlander, H. Mantler, S. Marzani and K. J. Ozeren, *Higgs production in gluon fusion at next-to-next-to-leading order QCD for finite top mass*, *Eur. Phys. J. C* **66** (2010) 359–372, [0912.2104].

[119] A. Pak, M. Rogal and M. Steinhauser, *Finite top quark mass effects in NNLO Higgs boson production at LHC*, *JHEP* **02** (2010) 025, [0911.4662].

[120] S. Actis, G. Passarino, C. Sturm and S. Uccirati, *NLO electroweak corrections to Higgs boson production at hadron colliders*, *Phys. Lett. B* **670** (2008) 12–17, [0809.1301].

[121] S. Actis, G. Passarino, C. Sturm and S. Uccirati, *NNLO computational techniques: The cases $H \to \gamma\gamma$ and $H \to gg$*, *Nucl. Phys. B* **811** (2009) 182–273, [0809.3667].

[122] M. Bonetti, K. Melnikov and L. Tancredi, *Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion*, *Phys. Rev. D* **97** (2018) 056017, [1801.10403].

[123] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, *JHEP* **02** (2010) 037, [0911.5299].

[124] M. Ciccolini, A. Denner and S. Dittmaier, *Strong and Electroweak Corrections to the Production of a Higgs Boson + 2 Jets via Weak Interactions at the Large Hadron Collider*, *Phys. Rev. Lett.* **99** (2007) 161803, [0707.0381].

[125] M. Ciccolini, A. Denner and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC*, *Phys. Rev. D* **77** (2008) 013002, [0710.4749].

[126] P. Bolzoni, F. Maltoni, S.-O. Moch and M. Zaro, *Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD*, *Phys. Rev. Lett.* **105** (2010) 011801, [1003.4451].

[127] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003) 073003, [hep-ph/0306234].

[128] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) 149–156, [hep-ph/0307206].

[129] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) 1868, [1111.0761].

[130] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) 078, [1211.5015].

[131] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, *Comput. Phys. Commun.* **195** (2015) 161–171, [1412.5390].

[132] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron colliders*, *Comput. Phys. Commun.* **184** (2013) 998–1003, [1210.5347].

[133] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) 082, [1410.0217].

[134] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [0710.3820].

[135] ATLAS Collaboration, "The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model." ATL-PHYS-PUB-2016-017, 2016.

[136] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, [1005.4568].

[137] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[138] P. Ilten, *Tau Decays in Pythia 8*, *Nucl. Phys. B Proc. Suppl.* **253-255** (2014) 77–80, [1211.6730].

[139] ATLAS Collaboration, *Operation of the ATLAS trigger system in Run 2*, *JINST* **15** (2020) P10004, [2007.12539].

[140] *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, , CERN, Geneva, 2016.

[141] ATLAS collaboration, *Trigger Menu in 2016*, , CERN, Geneva, 2017.

[142] ATLAS collaboration, *Trigger Menu in 2017*, , CERN, Geneva, 2018.

[143] ATLAS collaboration, *Trigger menu in 2018*, , CERN, Geneva, 2019.

[144] A. Elagin, P. Murat, A. Pranko and A. Safonov, *A new mass reconstruction technique for resonances decaying to $\tau\tau$*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654** (2011) 481–489.

[145] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson, *Parameterized neural networks for high-energy physics*, *The European Physical Journal C* **76** (2016) .

[146] CMS Collaboration, *Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions*, *JHEP* **10** (2014) 160, [1408.3316].

[147] Objax Developers, *Objax*, 2020.

[148] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *CoRR* **abs/1412.6980** (2014) .

[149] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554, [1007.1727].

[150] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *JINST* **14** (2019) P12006, [1908.00005].

[151] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2020) 689, [2007.02645].

[152] ATLAS Collaboration, "$E_{\mathrm{T}}^{\mathrm{miss}}$ performance in the ATLAS detector using 2015–2016 LHC pp collisions." ATLAS-CONF-2018-023, 2018.

[153] ATLAS Collaboration, *DeXTer: Deep Sets based Neural Networks for Low-p_{T} $X \to b\bar{b}$ Identification in ATLAS*, , CERN, Geneva, 2022.

[154] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, 2212.09379.

[155] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* **13** (2018) P07017.

[156] E. Gross and O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, *European Physical Journal C - EUR PHYS J C* **70** (05, 2010) .

[157] CMS collaboration, *Search for exotic Higgs boson decays to a pair of pseudoscalars in the $\mu\mu$bb and $\tau\tau$bb final states in proton-proton collisions with the CMS experiment*, , CERN, Geneva, 2023.

[158] CMS collaboration, *Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV*, .

[159] ATLAS collaboration, *Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System*, CERN-LHCC-2017-020, https://cds.cern.ch/record/2285584.

[160] H. Kroha, R. Fakhrutdinov and A. Kozhin, *New High-Precision Drift-Tube Detectors for the ATLAS Muon Spectrometer*, *JINST* **12** 12.

[161] K. Lang, J. Ting and V. Vassilakopoulos, *A technique of direct tension measurement of a strung fine wire*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **420** (1999) 392–401.

[162] Y. Guo, J. Wang, Y. Liang, X. Xiao, X. Hu, Q. An et al., *Design of a Time-to-Digital Converter ASIC and a mini-DAQ system for the Phase-2 upgrade of the ATLAS Monitored Drift Tube detector*, *Nucl. Instrum. Meth. A* **988** (2021) 164896.

[163] P. Farthouat and P. Gällnö, *TTC-VMEbus INTERFACE*, https://ttc.web.cern.ch/TTCviSpec.pdf.

[164] ATLAS collaboration, C. Posch, E. Hazen and J. Oliver, *MDT-ASD, CMOS front-end for ATLAS MDT*, ATL-MUON-2002-003, https://cds.cern.ch/record/684217.

[165] C. Bacci, C. Bini, G. Ciapetti, G. De Zorzi, P. Gauzzi, F. Lacava et al., *Autocalibration of high precision drift tubes*, *Nuclear Physics B - Proceedings Supplements* **54** (1997) 311–316.

[166] R. Carnegie, M. Dixit, J. Dubeau, D. Karlen, J.-P.Martin, H. Mes et al., *Resolution studies of cosmic ray tracks in a TPC with GEM readout*, *Nucl. Instrum. Meth. A* **538** (2005) 372–383.

[167] J. D. Jackson, *Classical electrodynamics; 2nd ed.* Wiley, New York, NY, 1975.

[168] ATLAS collaboration, K. R. Schmidt-Sommerfeld, *Study of Muon Drift Tube Detectors and Fast Readout Electronics for Very High Counting Rates*, CERN-THESIS-2020-016, https://cds.cern.ch/record/2713422.

[169] ATLAS collaboration, S. Hadzic, *Test of High-Resolution Muon Drift-Tube Chambers for the Upgrade of the ATLAS Experiment*, CERN-THESIS-2019-150, https://cds.cern.ch/record/2692069.