Search for Heavy Resonances Decaying to Tau Pairs with the CMS Experiment

by

A. A. Johnson

B.S., Carnegie Mellon University, 2011

M.S., University of Colorado, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

2016

This thesis entitled:
Search for Heavy Resonances Decaying to Tau Pairs with the CMS Experiment
written by A. A. Johnson
has been approved for the Department of Physics

_____

Prof. John P. Cumalat

_____

Prof. Kevin Stenson

_____

Ms. Steven Wagner

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Johnson, A. A. (Ph.D., Physics)

Search for Heavy Resonances Decaying to Tau Pairs with the CMS Experiment

Thesis directed by Prof. John P. Cumalat

Two searches for heavy neutral resonances, $Z\prime$s, decaying to back-to-back tau pairs are carried out using the Compact Muon Solenoid experiment at the Large Hadron Collider. The first, carried out during the 2012 data-taking run at $\sqrt{s} = 8$ TeV, focuses on the fully-leptonic channel, in which one tau decays to an electron and the other to a muon. The second, carried out during the 2015 run at $\sqrt{s} = 13$ TeV, focuses on the fully-hadronic channel, in which each tau decays to a hadronic jet. Each search employs a hybrid of background estimation strategies using both Monte Carlo simulation and data-driven methods. In both searches, observed data is found to be in agreement with Standard Model expectation and no excesses are observed. Limits are placed on the $Z\prime$, culminating in a world-record limit of 2.2 TeV at 95% confidence level.

Dedication

To my parents, Carl and Sharon Johnson. Without your constant love, support, and encouragement, none of this would have been remotely possible.

Acknowledgements

Here's where you acknowledge folks who helped. But keep it short, i.e., no more than one page, as required by the Grad School Specifications.

Contents

Chapter

Appendix

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 The Standard Model

The Standard Model (SM) of particle physics has been phenomenally successful at describing much of the fundamental interactions observed in the universe. It is not without significant deficiencies, however.

OUTLINE

1.2 Standard model

Z'

Why TauTau

Other BSM theories involving ditau pairs

1.3 The CMS Experiment

Pixels

Monitoring tool

Strips

ECAL

HCAL

Muon System

Triggering

L1 HLT

Computing @ CERN

1.4   Taus at CMS

1.5   EMu channel @ 8TeV

MC samples used

Generation methods

Data samples used

Object selection

Validation

Background estimation

Lifetime cuts optimization

Systematics

Limit setting methodology impact of lifetime cuts on limit

1.6   TauTau channel @ 13TeV

MC samples used

Generation methods

Data samples used

Object selection

Validation

Background estimation

Lifetime cuts optimization

Systematics

Limit setting methodology impact of lifetime cuts on limit

Chapter 2

The Compact Muon Solenoid (CMS) Experiment

The CMS Experiment is a multipurpose particle detector located on the LHC ring underneath the Franco-Swiss border at CERN in Geneva, Switzerland. The experiment is located 100 meters underground in Cessy, France. CMS is 28.7 meters long, 15.0 meters in diameter, and weighs approximately 14,000 tonnes. It's arranged in a cylindrical, multi-layered structure consisting of a "barrel" and two endcaps, with the LHC beam passing through the vertical axis of the cylinder. CMS consists of several subdetectors, each designed to measure a different class or property of particle. From the beam line outward, the layers of CMS are the tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the superconducting solenoid, and the muon system (interspersed with the iron return yoke).

The experiment uses a right-handed coordinate system: the origin is set at the pp collision point, with the x-axis pointing towards the center of the LHC ring, the y-axis pointing straight up, and the z-axis pointing along the beam line in the counter-clockwise direction. CMS also uses a pseudo-polar coordinate system, with θ defined as the polar angle from the beam axis, and η as the "pseudorapidity", itself defined as

$$\eta = -ln\left[tan\left(\theta/2\right)\right]$$

<div align="right">(2.1)</div>

2.1 The silicon tracker

The silicon tracker is the innermost detector element in CMS, and is designed to offer the highest resolution measurement of charged particle trajectories (such trajectories are referred to as "tracks"). The tracker is composed of approximately 200m^2 of silicon, and includes arrays of silicon pixels in the inner layer and arrays of silicon strips in the outer layer. The silicon elements are arranged in the densest configuration near to the interaction vertex ($r \approx 10$cm), with pixel size ≈ 100 x 150μm^2. As one moves away from the interaction vertex, the solid angle becomes large enough that the particle flux drops off to a degree that larger silicon elements may be used (silicon strips measuring 10cm x 80μm at 20cm$r$55cm, silicon strips measuring 25cm x 180μm for r55cm. The total number of silicon elements is 66 million pixels and 9.6 million strips. The tracker is divided into a barrel segment and two forward endcaps. The endcaps contain two pixel and nine strip layers each, and the barrel segment is separated into an inner and outer barrel. A schematic of the tracker layout can be seen in Figure fig(TrackerLayout). Particle tracks are reconstructed from hits in the individual silicon pixels/strips, the track being interpolated between hits.

2.1.1 The pixel detector

The pixel detector, shown in Figure fig(PixelLayout), consists of three barrel layers and two endcap layers. The barrel has a length of 53 cm, and the endcap disks range from 6 cm to 15 cm in radius. The pixel modules are arranged in a ladder-like configuration in the barrel, with 768 total modules comprising the barrel. The endcap disks are arranged in a turbine-like fashion, with 24 "blades" per disk, and 7 pixel modules per blade for a total of 672 pixel modules in the endcaps. Each pixel consists of a readout chip (ROC), which are bump bonded to the modules. In total, the pixel detector includes about 16,000 ROCs.

2.1.2   The strip tracker

The strip tracker is divided into a barrel segment and two endcap segments. The barrel segment is itself divided into a Tracker Inner Barrel (TIB) section and a Tracker Outer Barrel (TOB) section. The TIB includes four layers of silicon strips each 320μm thick and ranging in pitch from 80μm and 120 μm. In the TOB, the lower rate of particle flux allows for larger strips (each 500μm thick and ranging in pitch from 120μm to 180μm).

The endcaps are each comprised of a Tracker Endcap (TEC) and Tracker Inner Disks (TIDs). The TIDs are designed to fill the region between the TEC and the TIB. The TECs each contain nine disks, and each TID contains three disks. On each disk, the modules (for both TID and TEC) are arranged in rings centered on the beam line. The TID strips (and three innermost ring strips of the TEC) have thickness 320μm, while the rest of the TEC strips have thickness 500μm. In total the strip tracker contains about 15,400 strip modules.

2.2   The electromagnetic calorimeter

The next layer outward from the silicon tracker is the electromagnetic calorimeter (ECAL), which is designed primarily to measure energies of photons and electrons. The ECAL consists of an array of about 75,000 lead tungstate ($PbWO_4$) crystals, and has a barrel (BE) segment as well as two endcap (EE) segments. The BE segment has 61,200 crystals, while each EE segment contains 7,324 crystals.

$PbWO_4$ was chosen for the ECAL due to its short radiation length ($X_0 = 0.89$ cm), fast response time (80% of light emitted within 25ns), and radiation hardness (up to 10 Mrad).

The barrel (EB) crystals present an apparent cross-section (when viewed from the interaction vertex) of $\approx$ 22x22 mm^2, and are 230 mm (25.8 radiation lengths) thick. The barrel has an inner radius of 129 cm, and covers the pseudorapidity range $0|\eta|1.479$.

The endcap (EE) crystals are each arranged in two D-shaped semicircular aluminum plates. From these plates are cantilevered "supercrystal" structures consisting of 5x5 crystal blocks. Each

crystal presents an apparent cross-section of $\approx$ 28.6x28.6 mm^2, and are 220 mm (24.7 radiation lengths) thick, and the endcap crystals cover the pseudorapidity range 1.479$|\eta|$3.0.

A diagram of the ECAL layout can be seen in Figure fig(ECAL$_l$ayout).

2.3   The hadronic calorimeter

Surrounding the ECAL is the hadronic calorimeter (HCAL). The primary function of the HCAL is to measure the energies of hadrons (particles made of quarks and gluons). Located just inside the solenoid magnet, the HCAL is primarily composed of brass panels made from melted-down artillery shells. Interspersed with the brass panels are plastic scintillation panels, in which are embedded wavelength-shifting (WLS) fibers, which carry the signal to clear fibers outside the scintillators for readout. As hadrons enter the HCAL, they produce secondary particles in the brass which in turn create further particles. These hadron "showers" then interact with the plastic scintillators, where the fibers carry the signal to hybrid photodiodes so the signal can be measured.

HCAL is divided into barrel (HB) and endcap (HE) portions. Due to limited space between the ECAL and solenoid, the HCAL also includes material outside the solenoid: the outer HCAL (HO) lining the solenoid, and the forward calorimeter (HF) outside the muon endcap system. These additions increase the total radiation lengths covered by the HCAL to 10.

While HB and HE use the brass-plastic configuration, HO uses the solenoid itself instead of the brass, although the same plastic scintillators are used. The HF uses steel in place of brass, and quartz fibers in place of plastic scintillators. This is done to preferentially select neutral components of hadron showers, which are shorter and narrower and are thus well suited for the forward environment, which tends to be quite congested with particles.

A diagram of the HCAL layout is shown in Figure fig(HCAL$_l$ayout).

2.4   **The CMS magnet**

2.5   **The muon system**

2.6   **Computing at CMS**

Chapter 3

13 TeV Analysis

3.1   Introduction

At the forefront of the interconnection between particle physics and cosmology are the following questions: (1) What is the origin of the matter-antimatter asymmetry?; (2) What is the origin of neutrino mass?; (3) Are there new fundamental forces in nature?; (4) What is the origin of dark energy; and (5) Is the Higgs boson solely responsible for electroweak symmetry breaking and the origin of mass? Much like the Higgs mechanism is introduced to account for the SU(2)xU(1) symmetry breaking, there are a plethora of theoretical models which incorporate additional gauge fields and interactions to address these questions. For example, string theory is considered a promising candidate for describing gravitational systems at strong coupling and thus plays a prominent role in the description of black holes and evolution of the universe through the understanding of the origin of dark energy. Similarly, models with additional neutrino fields at the TeV scale provide a possible explanation for the mass of light neutrinos. Such models often manifest themselves as new heavy particles that could be observed at the LHC. Surprisingly, some of these new particles predicted on the basis of pure particle physics arguments can even provide the correct dark matter relic density. There are several ways new heavy gauge bosons appear. The most natural possibility is one in which these heavy gauge bosons are the gauge field of a new local broken symmetry. Examples include models with a new U(1) gauge symmetry, little Higgs models, and E6 Grand Unified Theories (GUT). In models with a new U(1) gauge symmetry, the Z' is the gauge boson of the broken symmetry. In little Higgs models, breaking of the global symmetry by gauge and

Yukawa interactions generates Higgs mass and couplings at the TeV scale that cancel off the SM quadratic divergence of the Higgs mass from top, gauge, and Higgs loops. This results in one or more Z' bosons. In Kaluza-Klein models, the Z' bosons are excited states of a neutral, bulk gauge symmetry. From the breadth, scope, and implications of these models, it is apparent that probing these questions and puzzles potentially lies in the physics of new particles at the TeV scale.

Of particular interest for this analysis note are models that include an extra neutral gauge boson that decays to pairs of high-p_T τ leptons. Although many models with extra gauge bosons obey the universality of the couplings, some models include generational dependent couplings resulting in extra neutral gauge bosons that preferentially decay to τ leptons, making this analysis an important mode for discovery. However, even if a new gauge boson decaying to $\mu\mu$ is discovered first, it will be critical to establish the $\tau\tau$ decay channel to establish the coupling relative to $\mu\mu$ channel.

3.2   Strategy

A τ lepton is the heaviest known lepton with a mass of 1.777 GeV/c^2 and a lifetime of 2.9×10^{-13} seconds. Around one third of all τ leptons decay to e/μ, and the remainder decay into hadronic jets (τ_h). In the latter case, a τ_h consists of one, three, or (rarely) five charged mesons usually accompanied by one or more neutral pions.

We consider four distinct analyses for pairs of τ lepton decays, namely $e\mu$ (6.2%), $e\tau_h$ (23.1%), $\mu\tau_h$ (22.5%), and $\tau_h\tau_h$ (42%). We ignore the and channels due to the copious Drell-Yan $Z/\gamma^* \rightarrow e^+e^-, \mu^+\mu^-$ production, although there is ongoing development of algorithms to discriminate prompt e/μ from τ lepton decays to light leptons which will be utilized with the 2016 data/analysis.

The overall strategy of the analysis is similar in all channels. In general, we identify events with two oppositely charged, nearly back–to–back objects. Because the $\tau\tau$ system decays with up to four neutrinos, we expect to have missing transverse energy () present in the event. In contrast to Z′searches in the e^+e^- and $\mu^+\mu^-$ channels, the visible $\tau^+\tau^-$mass does not produce a narrow peak due to the missing neutrinos. Instead, we look for a broad enhancement in the $\tau^+\tau^-$+invariant

mass distribution consistent with new physics. Our selections maintain high efficiency for signal events, provide strong background suppression, and reduce the influence of systematic effects.

As $Z \rightarrow \tau^+\tau^-$ is both background as well as an important validation signal our final selection requirements are such that by removing or reversing just a few cuts we can obtain a clean sample of $Z \rightarrow \tau^+\tau^-$ events. In order to ensure robustness of the analysis and our confidence in the results, whenever possible we rely on the data itself to understand and validate the efficiency of reconstruction methods as well as the estimation of the background contributions. For that purpose we define control regions with most of the selections similar to what we use in our main search but enriched with events from background processes. Once a background enhanced region is created, we measure selection efficiencies in those regions and extrapolate to the region where we expect to observe our signal. In cases where a complete data–driven method is not possible we make use of scale factors, ratio between observed data events and expected MC events in the background enhanced region to estimate the background contribution in the signal region. Although each individual channel could have its own set of requirements, we maintain, wherever possible, consistent definitions and selection criteria between channels.

To quantify the significance of any possible excess or set upper limits on the production rate, we perform a fit of the $m(\tau_1, \tau_2,)$ mass distribution and employ the CL_s technique to interpret the results in terms of the upper 95% credibility level limits for each channel. The joint limit is obtained by combining the posterior probability density functions (likelihood) and taking into account correlation of systematic uncertainties within and across channels.

The structure of this note is as follows: Sections 3.3 and 3.4 describes the data sets used in each analysis. Section 3.5 provides a brief discussion of the reconstruction and identification of the objects used to reconstruct our $\tau\tau$ pairs. Sections 3.6–3.9 describe the specific selections and background extraction methods applied to each individual channel. In Sections 3.10–3.11 we describe the statistical method used to extract the 95% C.L. upper limits and the results of the analysis.

3.3   Triggers

For the final states, we use single lepton triggers instead of $l \times \tau_h$ cross-triggers to maintain a similar strategy across channels and also allow us to use the τ_h isolation sidebands as control and validation samples. For the final state, any trigger with iso-muon requirement, either a e-μ cross-trigger or single-muon trigger, would eliminate the isolation sideband for QCD estimation. Hence, we use the same single-electron trigger as for the final state. Trigger paths are summarized in Table 3.1. Note this constrains the object ID and phase space that we are studying. For example, the use of these triggers requires p_T cuts of 30 GeV, 35 GeV, and 60 GeV for our leading light leptons and τ_h's in the , , and channels, respectively.

Table 3.1: The trigger paths used to collected the data. Emulated trigger paths, in particular those most similar to the paths used to collect the data, are applied to the simulated samples.

Channel	Trigger Path
(data RunC, sim)	HLT_DoubleMediumIsoPFTau40_Trk1_eta2p1_Reg
(data RunD)	HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg
(data RunC)	HLT_IsoMu24_eta2p1
(data RunD)	HLT_IsoMu18
(sim)	HLT_IsoMu17 (L1 μ > 18)
, (data)	HLT_Ele27_eta2p1_WPLoose
, (sim)	HLT_Ele27_eta2p1_WP75

3.3.1   Single Lepton Trigger Efficiency

The single electron and muon trigger efficiencies are measured using a tag and probe method where we selected events with at least one electron and one muon pair satisfying the following requirements:

For electrons:

- $p_T > 13$, $|\eta| < 2.1$, isolation < 0.15, $d_{xy} < 0.045$ cm, $d_z < 0.2$ cm

- passing electron ID (we performed different measurements for "MVANonTrigWP80", "HEEP", "MVATrigWP80")

- no matched conversions

- number of missing hits = 0

For muons:

- $p_T > 20$, $|\eta| < 2.1$, isolation < 0.15, $d_{xy} < 0.045$ cm, $d_z < 0.2$ cm

- passing "Medium" muon ID

For the pair:

- $\Delta R() > 0.3$

- choose the most isolated pair

- choose the opposite sign pair

- 3rd lepton veto

3.3.2   Single Electron Trigger Efficiency

To measure the single-electron trigger efficiency, after the preselection mentioned above, we select (tag) events with a single-muon trigger (HLT_IsoMu18 for data and HLT_IsoMu17_eta2p1 for MC with L1 μ $p_T > 18$) with the offline muon matching the HLT muon that fired the trigger. Then, the trigger efficiency is defined at the fraction of events which also pass (probe) the single-electron trigger (HLT_Ele27_eta2p1_WPLoose for data and HLT_Ele27_eta2p1_WP75 for MC).

The efficiency curves of the single-electron triggers, measured vs. electron , are shown in Figure 3.1. A $p_T > 35$ cut is motivated to avoid the turn-on region. As shown in Figure 3.2, the efficiency curves were further binned into two p_T bins containing similar numbers of events. A weight of 0.94 is measured, and used in the analysis, as the ratio between simulated and observed events in which the selected electron is in the endcap region.

electron trigger turn-on curve

electron trigger turn-on curve

electron trigger turn-on curve

electron trigger turn-on curve

electron trigger turn-on curve

electron trigger turn-on curve
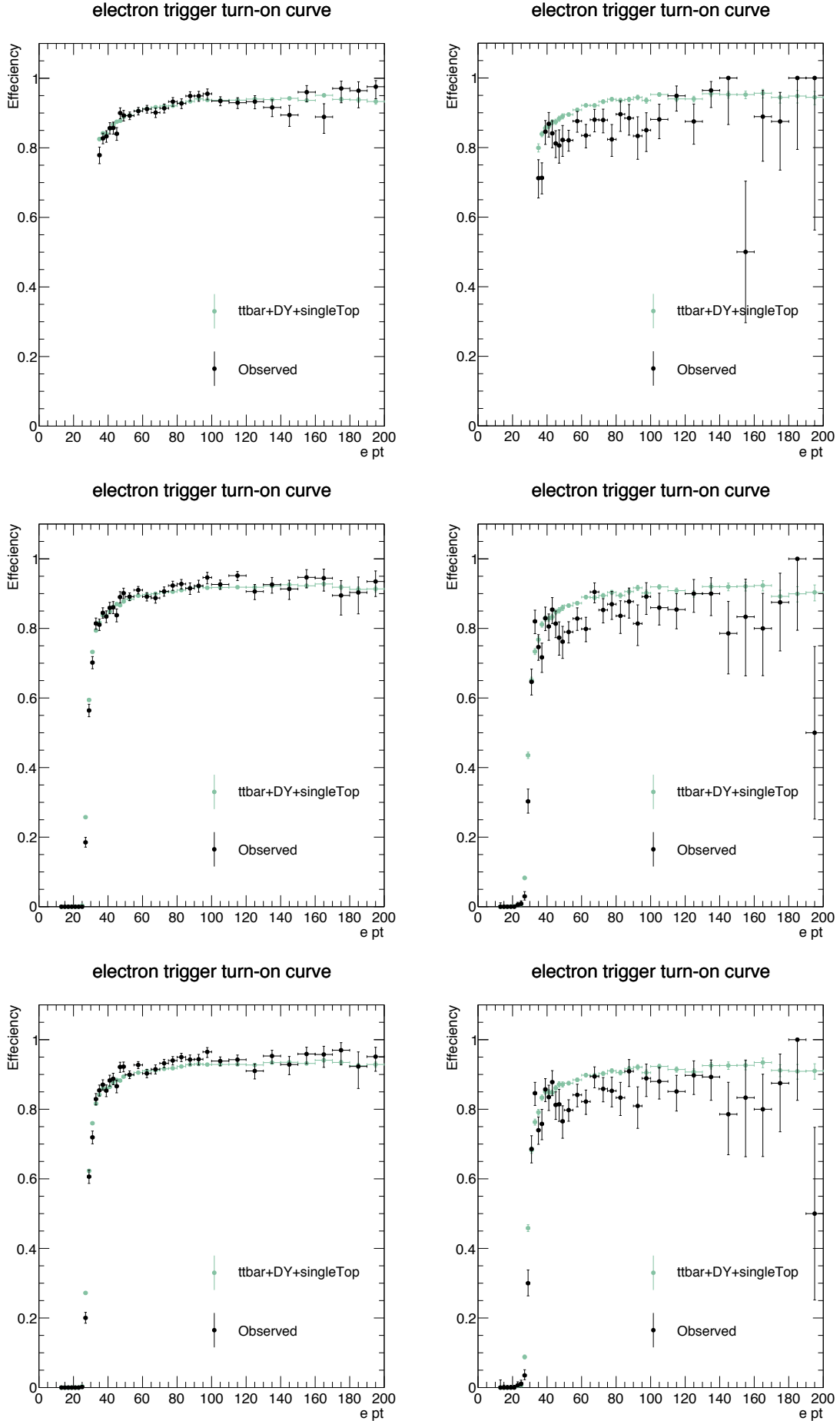
Figure 3.1: The efficiency vs. curves of the single-electron triggers used. Left column: barrel. Right
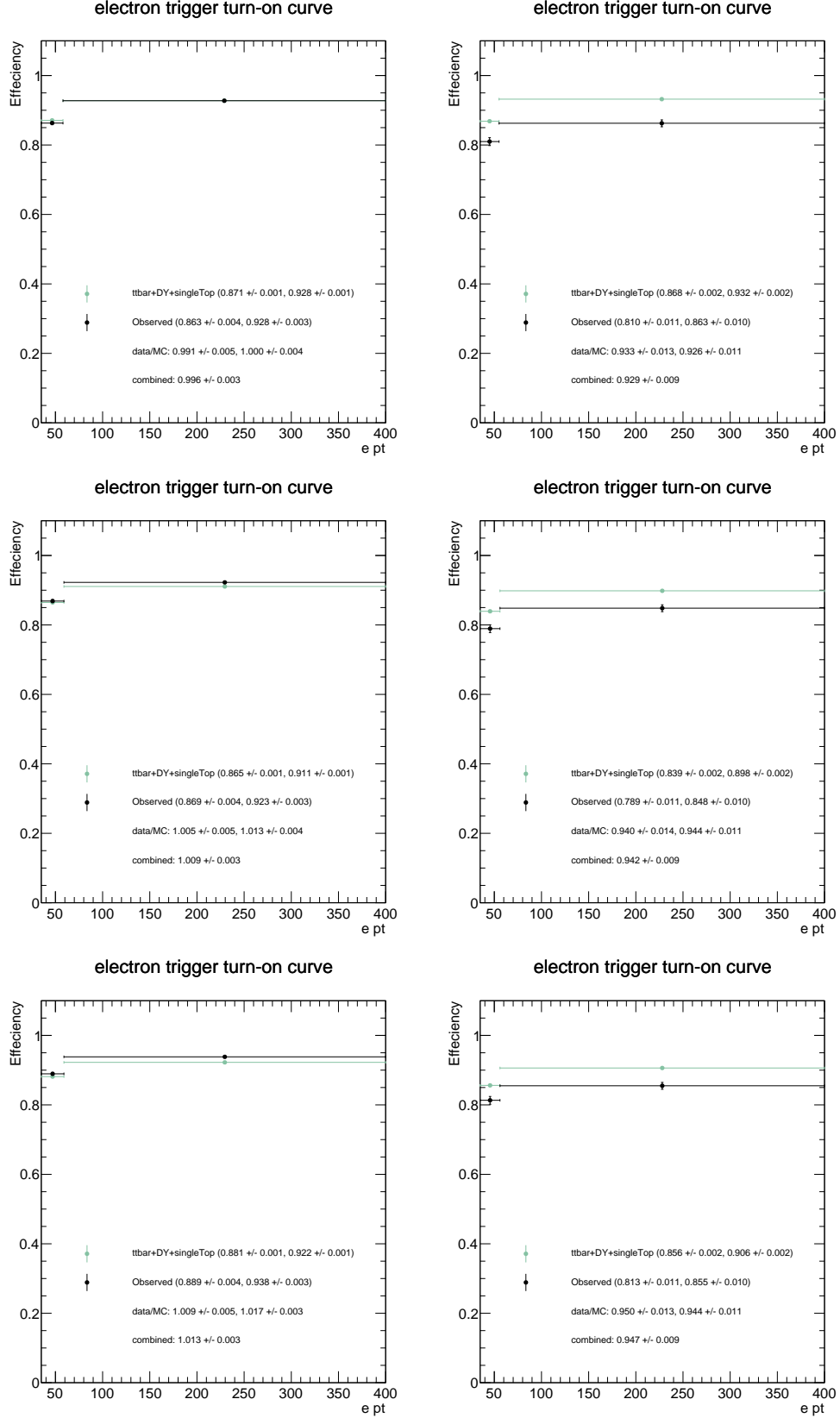
Figure 3.2: The efficiency vs. curves of the single-electron triggers used. Left column: barrel. Right column: endcap. The offline electron ID requirements used are HEEP (top row), MVANonTrig80 (middle row), MVATrigWP80 (bottom row). The two bins each contain half of the events with > 35 .

3.3.3 Single Muon Trigger Efficiency

To measure the single-muon trigger efficiency, after the preselection mentioned above, we select (tag) events with a single-electron trigger (HLT_Ele27_eta2p1_WPLoose for data and HLT_Ele27_eta2p1_WP75 for MC) with the offline electron match the HLT electron that fired the trigger. Then, the trigger efficiency is defined at the fraction of events which also pass (probe) the single-muon trigger (HLT_IsoMu18 for data and HLT_IsoMu17_eta2p1 for MC with L1 μ $p_T > 18$ GeV).

The efficiency curves of the single-muon triggers, measured vs. muon , are shown in Figure 3.3. A > 25 cut is motivated to avoid the turn-on region. As shown in lower two plots, the efficiency curves were further binned into two bins containing similar numbers of events. Weights of 0.991 and 0.986 are measured, and used in the analysis, as the ratios between simulated and observed events in which the selected muon $|\eta| \leq 1.2$ and $|\eta| > 1.2$.

3.3.4 Di-Tau Trigger Efficiency

The efficiency of the $\tau_h\tau_h$ trigger is measured using $Z \to \tau\tau \to \mu\tau_h$ events. The τ_h candidates reconstructed in the selected $Z \to \tau\tau \to \mu\tau_h$ events are required to pass the same τ_h identification used for the final analysis and which will be described in more detail in the sections to follow. The "denominator" selections used to define the $Z \to \tau\tau \to \mu\tau_h$ control sample are summarized below:

- Events must fire the HLT_IsoMu18 trigger (HLT_IsoMu17_eta2p1 with HLT pt cut of 18 GeV for MC)

- ≥ 1 global μ with $|\eta| < 2.1, p_T > 19$ GeV

- "isMediumMuon"

- muon best track $d_{xy} < 0.2$ cm , $d_z < 0.045$ cm w.r.t. PV

- Relative isolation (with $\delta\beta$ corrections) < 0.1

- ≥ 1 HPS τ_h with $|\eta| < 2.1, p_T > 20$ GeV
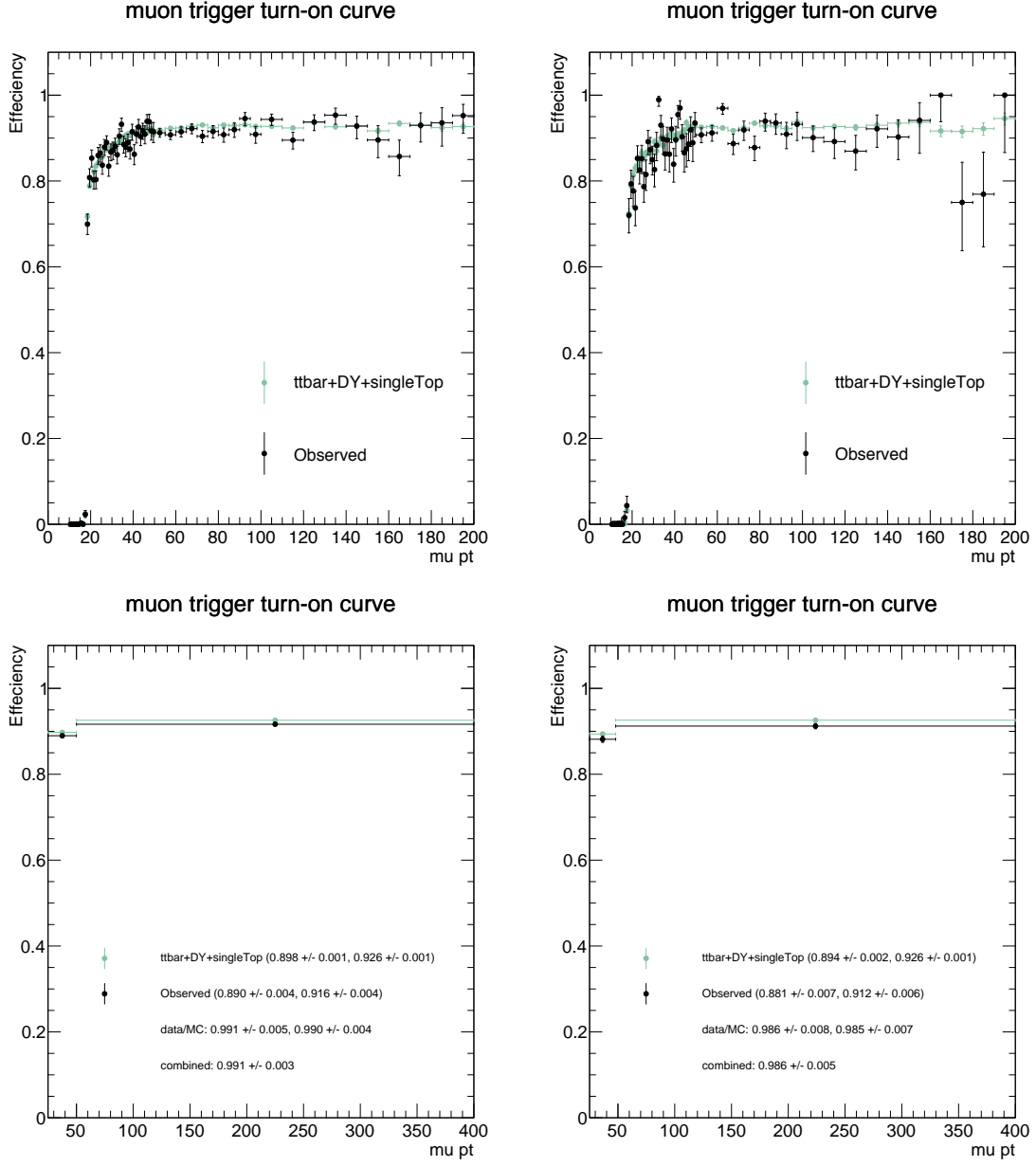
Figure 3.3: The efficiency vs. curves of the single-muon triggers used. Left column: $|\eta| < 1.2$. Right column: $|\eta| > 1.2$. Bottom row: the two bins each contain half of the events with > 25 .

- Muon veto: "againstMuonTight3"

- Electron veto: "againstElectronVLooseMVA5"

- "old" decay mode finding

- Isolation: "byTightCombinedIsolationDeltaBetaCorr3Hits"

- $p_T > 5$. GeV for leading Track of tau with $d_{xy} < 0.2$ cm and $d_z < 0.045$ cm w.r.t. PV

- $\Delta R(\mu, \tau_h) > 0.5$

- $Q(\mu) \times Q(\tau_h) < 0$

- $m_T(\mu,) < 40$ GeV

- 0 jets tagged as b-jets

- 0 tagged electrons

- veto events with a second muon forming opposite-sign dimuon pair

The numerator is defined by additionally requiring those events to pass the HLT_IsoMu17_etap2p1_MediumIsoPFTau40* trigger. The efficiency is measured for each τ_h leg individually and parametrized as function of p_T (Figure 3.4). From the plot it is clear that the emulation of the trigger in simulation does not provide the correct trigger efficiency observed in data. This is mainly due to a difference in the trigger definition in MC, specifically the L1 seed. Therefore, we do not apply the trigger in MC, but instead model the correct per leg trigger efficiency observed in data by weighing the predictions from simulation using the fit of the trigger efficiency curve in data (black solid curve in the plot). The trigger efficiency at the plateau is approximately 90%.

Figure 3.4 (right) shows the trigger efficiency for data and all backgrounds obtained by requiring $Z \to \tau_\mu \tau_h$. Most of the backgrounds are taken from simulation except QCD whose shape and rate have been taken from same sign control region from data by subtracting the contribution
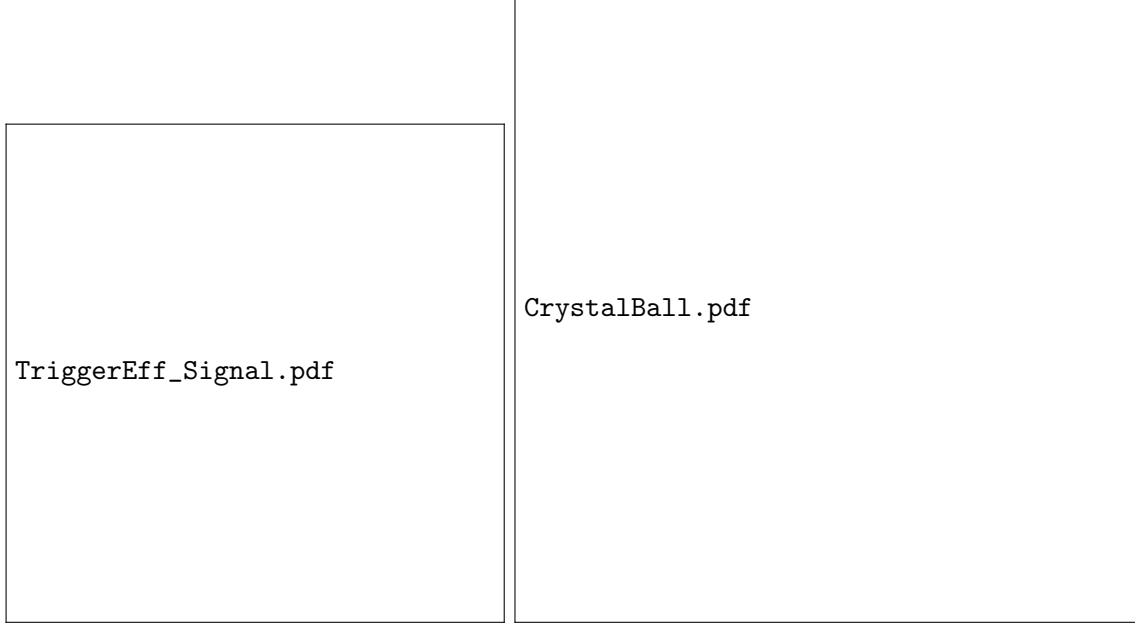
Figure 3.4: The per-leg τ_h trigger efficiency as a function of $p_T(\tau_h)$ for (a) simulated Z' signal sample on left (b) for data and all backgrounds obtained by requiring mu-tau region on right.

of other backgrounds.QCD normalization is then corrected by OS/LS ratio which is taken as $\sim$ 1.05. Table 3.2 shows the event yield in this control region requiring the denominator selections. The "purity", fraction of DYJets events out of all backgrounds, of this region comes out to be $\sim$ 65-66 % . The efficiency curves on Figure 3.4 (right) is fitted with crystall-ball function.

Table 3.2: Event rate in mu-tau control region after requiring denominator level selections

Sample	Events
Data	19578
t $\bar{t}$	170.881±13.072
W+Jets	2127.76±46.128
Z+Jets	12218.5±110.544
QCD	4003.93±63.277

3.4   Data and Monte Carlo Samples

The 13 TeV collision data collected by the CMS detector in year 2015 is used in this analysis. Table 3.3 shows the collision datasets used. The official JSON file is used to select "good" run

ranges and lumi sections. The total integrated luminosity of the collision data samples is 2.11 fb^{-1}.

The official Spring miniAODv2 2015 MC samples are used for all Standard Model processes. The leading order generators, PYTHIA8 and MADGRAPH, were mainly used for Signal and Background MC production. The predicted background yields in simulation were determined using NLO or NNLO cross-sections, while the signal yields and distributions in all plots shown in this AN were normalized using the leading order cross-sections shown in Table 3.4. Table 3.4 shows the entire list of the MC samples used for this analysis.

Table 3.3: Collision Data Samples

Physics Sample	Official CMS Datasets
Run 2015C SingleMu 05 Oct ReMiniAOD	/SingleMuon/Run2015C_25ns-05Oct2015-v1/MINIAOD
Run 2015D SingleMu 05 Oct ReMiniAOD	/SingleMuon/Run2015D-05Oct2015-v1/MINIAOD
Run 2012D SingleMu PromptReco v4	/SingleMuon/Run2015D-PromptReco-v4/MINIAOD
Run 2015C SingleElectron 05 Oct ReMiniAOD	/SingleMuon/Run2015C_25ns-05Oct2015-v1/MINIAOD
Run 2015D SingleElectron 05 Oct ReMiniAOD	/SingleElectron/Run2015D-05Oct2015-v1/MINIAOD
Run 2012D SingleElectron PromptReco v4	/SingleElectron/Run2015D-PromptReco-v4/MINIAOD
Run 2015C Tau 05 Oct ReMiniAOD	/Tau/Run2015C_25ns-05Oct2015-v1/MINIAOD
Run 2015D Tau 05 Oct ReMiniAOD	/Tau/Run2015D-05Oct2015-v1/MINIAOD
Run 2012D Tau PromptReco v4	/Tau/Run2015D-PromptReco-v4/MINIAOD

Because the MC simulated samples contain a pileup (PU) distribution that does not match that of data, the MC needs to be properly weighted to fit the PU distribution observed in data. The reweighting of MC events is performed by determining the probabilities to obtain n interactions in data ($P_{data}(n)$) and MC ($P_{MC}(n)$) and using the event weights

$$w_{PU}(n) = \frac{P_{data}(n)}{P_{MC}(n)} \tag{3.1}$$

to reweigh MC events based on the number of interactions. The recommended minbias cross-section of 69 mb is used to determine these weights.

3.5   Physics Object Reconstruction

3.5.1   Jet Reconstruction

Particle-flow (PF) technique [**?**, **?**] is used to improve the jet p_T and angular resolution in this analysis. The PF technique combines information from different subdetectors to produce a mutually exclusive collection of particles (namely muons, electrons, photons, charged hadrons and neutral hadrons) that are used as input for the jet clustering algorithms. Jets are clustered using the anti-k_T algorithm [**?**], with a distance parameter of $R = 0.4$ in η-ϕ plane (defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) by summing the four-momenta of individual PF particles.

The jets require energy corrections obtained using simulated events that are generated with `PYTHIA`, processed through a detector simulation based on `GEANT4`, and confirmed with in situ measurements of the p_T balance. The overall jet-energy corrections depend on the η and p_T values of jets. The jet-energy corrections are applied by L1 FastJet, L2 Relative, and L3 Absolute corrections. In order to remove the extra energy in jets from underlying event (UE) and pileup (PU), the L1 FastJet corrections use the event-by-event UE/PU densities. The L2 and L3 corrections use jet balancing and photon+jet events to improve and provide a better energy response as a function of jet p_T and η. For data, additional residual corrections are applied.

Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$. For the identification of jets the loose PF ID is used in this analysis. Table 4 shows the selection criteria used for the recommended loose PF ID, which are validated in other studies [**?**]. The jet reconstruction and ID efficiency in simulation is >98%.

3.5.1.1   b-Jet Tagging

In this analysis, b-tagged jets are used for two purposes: to reduce $t\bar{t}$ background in the signal region and to obtain $t\bar{t}$ enriched control samples used to estimate the signal rate.

The CSVv2 algorithm [**?**] is used to identify jet as originating from hadronization of a b-quark. This algorithm combines reconstructed secondary vertex and track-based lifetime information to

build a likelihood-based discriminator to distinguish between jets from b-quarks and those from charm or light quarks and gluons.

The minimum thresholds on these discriminators define loose, medium, and tight operating points with a misidentification probability of about 10%, 1%, and 0.1%, respectively, with an average jet p_T of about 80 GeV. The loose operating point for an efficiency about 85% is used in this analysis. A large sample of pair-produced top quark events is used to measure b-tagging efficiency using several methods [?]. A scale factor is applied to correct for differences in b-tagging efficiency between data and simulation [?].

3.5.2   Electron Reconstruction and Identification

Electrons are reconstructed using information from the tracker and Ecal detectors. Electrons passing through the silicon tracker material lose energy due to Bremsstrahlung radiation. The energy of the radiated photons is spread over several crystals of the Ecal detector along the electron trajectory, mostly in the ϕ direction (the magnetic field is in the z direction). Two algorithms based on energy clustering, "Hybrid" for the barrel and "Island" for the endcaps, are used to measure the energy of electrons and photons [?].

Electron tracks are reconstructed by matching trajectories in the silicon strip tracker to seed hits in the pixel detector. A pixel seed is composed of two pixel hits compatible with the beam spot. A Gaussian Sum Filter (GSF) is used for the reconstruction of trajectories in the silicon strips. In order to minimize the many possible trajectories due to different combinations of hits, the track that best matches an energy supercluster in the Ecal is chosen to be the reconstructed track.

The preselection of primary electron candidates requires good geometrical matching and good agreement between the momentum of the track and the energy of the ECAL supercluster. Two quantities used to estimate the geometrical matching are $\Delta\eta_{in} = \eta_{sc} - \eta_{vertex}^{Track}$ and $\Delta\phi_{in} = \phi_{sc} - \phi_{vertex}^{Track}$. The η_{sc} and ϕ_{sc} coordinates correspond to the supercluster position and are measured using an energy weighted algorithm. The η_{vertex}^{Track} and ϕ_{vertex}^{Track} coordinates are the position of the

track at the interaction vertex extrapolated, as a perfect helix, to the Ecal detector. The good energy-momentum matching is measured by taking the ratio between the corrected energy E_{corr} in the Ecal supercluster and the momentum of the track P_{in} measured in the inner layers of the tracker.

Electron selections have two main components, electron identification (eID) and electron isolation. In this analysis we use the non-triggering MVA electron identification. The MVA cuts used to define the 80% and 90% signal efficiency working points are summarized in Table 3.9. In all channels, the identification and isolation used follows the POG recommended criteria. The exact discriminator names and working points for each channel are listed and described in their respective sections.

3.5.3   Muon Reconstruction and Identification

Muon reconstruction is a multistep process that begins with the information gathered from the muon subdetector. As a first step, standalone muons are reconstructed from hits in the individual drift tube (DT) and cathode strip (CSC) chambers. Hits from the innermost muon stations are combined with hits in the other muon segments using the Kalman fitting technique. The standalone muon trajectory is reconstructed by extrapolating from the innermost muon station to the outer tracker surface. This standalone trajectory is then used to find a matching track reconstructed in the inner silicon tracker. Finally, standalone muons and matching silicon tracks are used to perform a global fit resulting in a "global" muon. Muon reconstruction is described in more detail in [?].

Global muons are reconstructed by combining tracker muons from the inner silicon tracker and standalone muons from the muon chambers. Once a muon is required to have matching tracks in the inner and outer detectors, the main source of background consists of charged hadrons that leave a signature in the inner silicon tracker while also penetrating through the hadronic calorimeter and creating hits in the muon chambers. Charged hadrons that penetrate the hadronic calorimeter and leave hits in the muon system will deposit significant energy in the calorimeters. Therefore, a calorimeter compatibility algorithm can be used to significantly reduce the number of charged pion

fakes. However, calorimeter compatibility is not used in this analysis due to our uncertainty of the performance of such algorithms in the presence of high PU. The presence of punch-throughs often occur due to pions from the fragmentation of quarks and gluons. These punch-throughs can often be discriminated against by requiring isolation. Similarly, non-prompt muons from heavy flavor decays and decays in flights are expected to be within jets and can be discriminated against by imposing an isolation requirement. Muon identification is described in more detail in [?] and [?].

Isolated muons are required to have minimal energy from PF neutral and charged candidates in a cone of $\Delta R = 0.4$ around the lepton trajectory. PF charged candidates considered in the calculation of isolation are required to be near the primary vertex. Isolation for muons is defined as:

$$I = \frac{\sum_i p_T^i}{p_T^\mu} \tag{3.2}$$

where the index i runs over PF neutral and charged candidates. Table 3.10 shows the complete list of for the "isMedium" μ identification criteria used in this analysis. In all channels, the identification and isolation used follows the POG recommended criteria.

3.5.4   Tau Reconstruction and Identification

Responsible: A. Johnson, F. Romeo, J. Cumalat

The challenge in identifying hadronically decaying taus is discriminating against generic quark and gluon QCD jets which are produced with a cross–section several orders of magnitude larger. CMS has developed several algorithms to reconstruct and identify hadronically decaying taus based on Particle Flow (PF) objects. For this analysis, the tau POG recommended Hadron Plus Strips algorithm (HPS) is used. HPS makes use of PF jets as inputs to an algorithm that uses strips of clustered electromagnetic particles to reconstruct neutral pions. The electromagnetic strips ("neutral pions") are combined with the charged hadrons within the PFJets to attempt to reconstruct the main tau decay modes outlined in Table 3.11.

The single hadron plus zero strips decay mode attempts to reconstruct $\tau \rightarrow \nu \pi^\pm$ decays or

$\tau \to \nu\pi^{\pm}\pi^0$ decays where the neutral pion has very low energy. The single hadron plus one or two electromagnetic strips attempts to reconstruct tau decays that produce neutral pions where the resulting neutral pion decays produce collinear photons. Similarly, the single hadron plus two strips mode attempts to reconstruct taus that decay via e.g. $\tau \to \nu\pi^{\pm}\pi^0$ where the neutral pion decays to well separated photons resulting in two electromagnetic strips. The three hadrons decay mode attempts to reconstruct tau decays that occur via $\rho(770)$ resonance. Although it is possible to recover signal in the two hadron decay mode (in the case of one of the three prongs being misidentified), this mode is not considered as its inclusion reduces discrimination performance and hurts the limit. In all cases, electromagnetic strips are required to have $E_T > 1$ GeV/c. Additionally, the particle flow charged hadrons are required to be compatible with a common vertex and have a net charge of $|q| = 1$.

In order to enforce the isolation requirement on the reconstructed tau, a region of size R = 0.5 around the tau decay mode direction is defined. Any PF candidates not used for the reconstruction of electromagnetic strips and charged hadrons not involved in the reconstruction of the tau decay modes are used to calculate isolation. The "Tight", "Medium", and/or "Loose" 3-hits isolation (with $\delta\beta$ corrections) working points are used.

In order to discriminate against muons, HPS taus are required to pass the lepton rejection discriminator which requires the lead track of the tau not be associated with a global muon signature. In order to discriminate against electrons, HPS taus are required to pass a MVA discriminator which uses the amount of HCAL energy associated to the tau with respect to the measured momentum of the track (H/p). Additionally, the MVA discriminator considers the amount of electromagnetic energy in a narrow strip around the leading track with respect to the total electromagnetic energy of the tau. Finally, HPS taus must not reside in the ECAL cracks. In all channels, the identification and isolation used follows the Tau POG recommended criteria. The exact discriminator names and working points for each channel are listed and described in their respective sections.

3.5.5 Efficiency of Tau Identification discriminators

The efficiency of the τ_h ID discriminators used in the analysis are studied using Z'_{SSM} samples with $m(Z') = 2000$ GeV. For this purpose, we require the reconstructed τ_h to have $p_T > 45$ and pseudorapadity $|\eta| < 2.1$. Further, reconstructed τ_h is matched to generator-level tau with $\Delta R(\tau_{reco}, \tau_{gen}) < 0.3$. The efficiency of the Decay Mode Finding (DMF) discriminator "Decay-ModeFinding" is found to be relatively flat at $\sim 80\%$ as shown in Figure 3.11. The individual effiencies of anti-muon , anti-electron and isolation discriminators, relative to the DMF criterion, are shown in Figures 3.12– 3.14. The overall efficiency of τ_h ID selection used in this analysis is $\sim 55\%$. We use the VLoose working point of the anti-electron MVA5 discriminator as the tighter working points have poor performance (i.e. low efficiency which is also not "flat" vs. p_T). The relative flatness of the τ_h identification efficiency with p_T will also facilitate the use of $Z \to \tau\tau$ events as a "standard candle" for comparison with signal.

DMF_4_29_16.pdf   DMF_eta_4_29_16.pdf   DMF_nvtx_4_29_16.pdf

Figure 3.5: Decay Mode Finding efficiency as function of p_T, η, and number of vertices of generator-level taus

3.5.6 Tau Energy Scale and Resolution

Since the resolution and scale of our mass reconstruction depends on the effectiveness of the τ_h reconstruction, in this section we summarize our studies on τ_h response and resolution. We define the response as the difference between the transverse momentum of a reconstructed tau (that has passed all tau ID discriminators) and the transverse momentum of a generated tau that has been

AgainstMuon_4_29_16.pdf	AgainstMuon_eta_4_29_16.pdf	AgainstMuon_nvtx_4_29_16.pdf
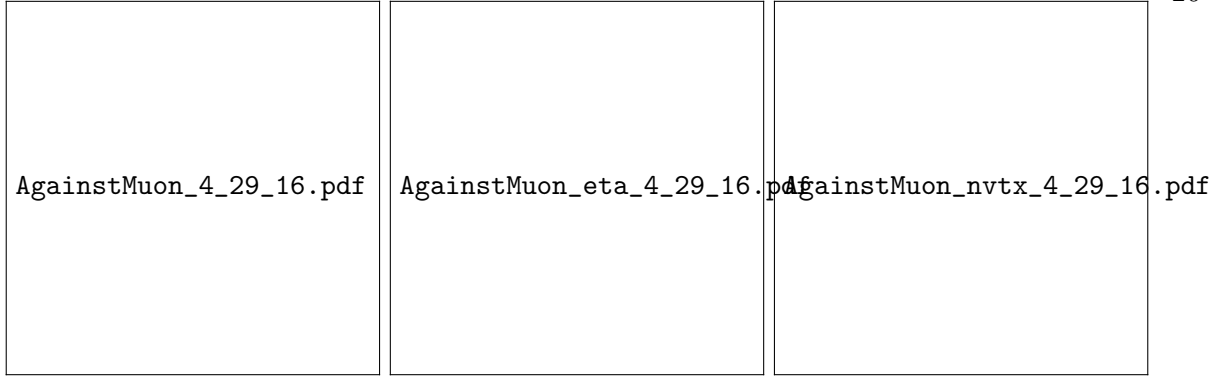
Figure 3.6: Efficiency of anti-muon discriminator (Loose3) as function of p_T, η, and number of vertices of generator-level taus

AgainstElectron_4_29_16.pdf	AgainstElectron_eta_4_29_16.pdf	AgainstElectron_nvtx_4_29_16.pdf
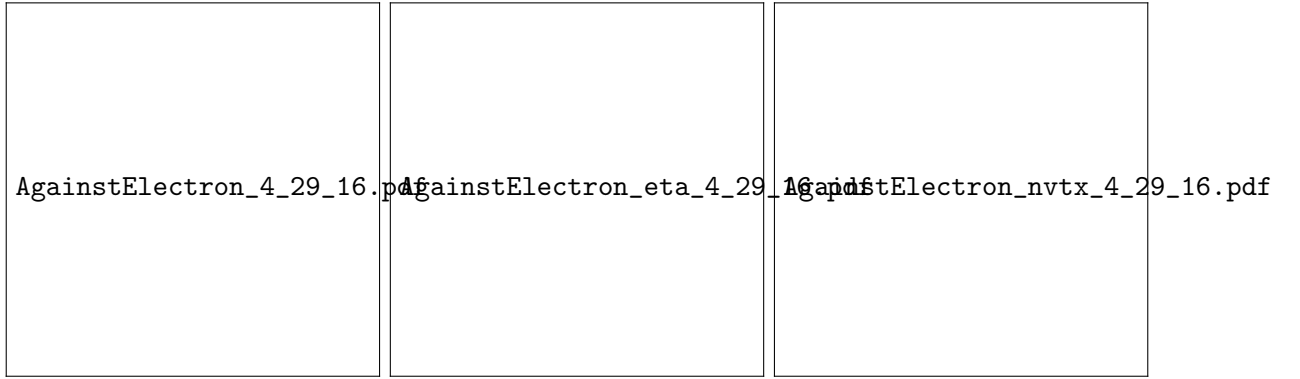
Figure 3.7: Efficiency of anti-electron discriminator (Loose/Medium MVA5) as function of p_T, η, and number of vertices of generator-level taus

CombinedIso_4_29_16.pdf	CombinedIso_eta_4_29_16.pdf	CombinedIso_nvtx_4_29_16.pdf
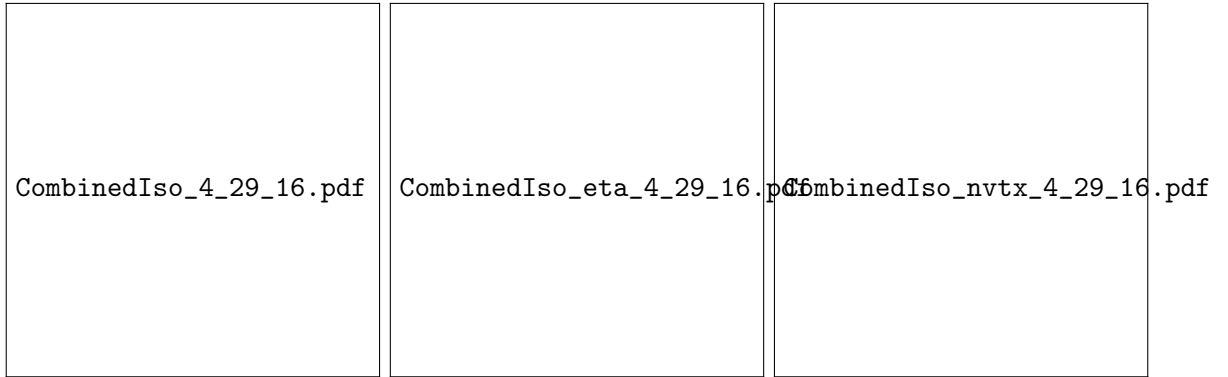
Figure 3.8: Efficiency of isolation discriminator ("CombinedIsoDB3Hits") as function of p_T, η, and number of vertices of generator-level taus

matched ($\Delta R < 0.3$) to the reconstructed tau. We can see from Figure 9 (right) that the response

distribution contains a narrow gaussian like component in the bulk, in addition to a relatively long tail (compared to electrons and muons). While the tails become less substantial at high p_T, the gaussian-like bulk of the response distributions broaden at high p_T.
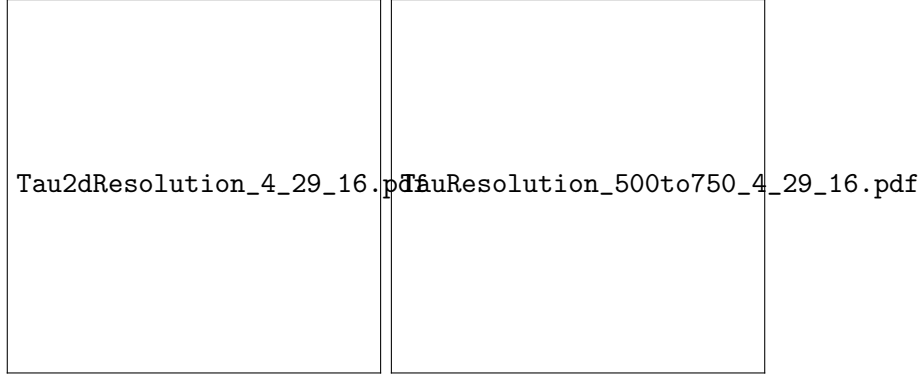
Figure 3.9: Energy response of reconstructed taus, raw and vs. generated tau p_T

3.5.7

The presence of neutrinos in the $\tau\tau$ decays must be inferred from the imbalance of total momentum because they escape from the detector without producing any direct response. The magnitude of the negative vector sum of the visible transverse momentum is known as missing transverse energy, denoted E_T^{miss}.

Missing transverse energy is one of the most important observables for discriminating the signal events from background events which do not contain neutrinos, such as QCD multijet and DY$\rightarrow \mu\mu$/ee events. Because there are many factors which may result in artificial E_T^{miss}, comprehensive studies of E_T^{miss} as measured using the CMS dector are performed [?]. We require events with $E_T^{miss} > 30$ GeV in all channels.

Finally, the standard and recommended "MET Filters" are utilized in this search.

3.5.8   Corrections for Pile-Up

Quantities such as and isolation, where energy depositions are summed up over some range of the detector, can suffer large inefficiencies or systematic effects due to particles from pile-up

interactions. Therefore, a simple and robust method must be employed to subtract off the contribution from secondary interactions. In the case of , this is done by using the FastJet corrections to determine the density of PU on an event by event basis. For isolation, the recommended delta-beta corrections are applied. The recommended PU corrections have been included for all objects in all the channels.

We apply event-by-event the "official" pile-up weights ($\sigma = 69$ mb). Figure 3.10 shows the distributions of the number of reconstructed vertices, before and after applying the pile-up weights.

3.5.9   Electron Reconstruction and Identification

Electrons are reconstructed using information from the tracker and Ecal detectors. Electrons passing through the silicon tracker material lose energy due to Bremsstrahlung radiation. The energy of the radiated photons is spread over several crystals of the Ecal detector along the electron trajectory, mostly in the ϕ direction (the magnetic field is in the z direction). Two algorithms based on energy clustering, "Hybrid" for the barrel and "Island" for the endcaps, are used to measure the energy of electrons and photons [?].

Electron tracks are reconstructed by matching trajectories in the silicon strip tracker to seed hits in the pixel detector. A pixel seed is composed of two pixel hits compatible with the beam spot. A Gaussian Sum Filter (GSF) is used for the reconstruction of trajectories in the silicon strips. In order to minimize the many possible trajectories due to different combinations of hits, the track that best matches an energy supercluster in the Ecal is chosen to be the reconstructed track.

The preselection of primary electron candidates requires good geometrical matching and good agreement between the momentum of the track and the energy of the ECAL supercluster. Two quantities used to estimate the geometrical matching are $\Delta\eta_{in} = \eta_{sc} - \eta_{vertex}^{Track}$ and $\Delta\phi_{in} = \phi_{sc} - \phi_{vertex}^{Track}$. The η_{sc} and ϕ_{sc} coordinates correspond to the supercluster position and are measured using an energy weighted algorithm. The η_{vertex}^{Track} and ϕ_{vertex}^{Track} coordinates are the position of the track at the interaction vertex extrapolated, as a perfect helix, to the Ecal detector. The good

Figure 3.10: Distributions of the number of reconstructed primary vertices. Top: $e\mu$ channel. Bottom: $e\tau_{\mathrm{h}}$ channel. Left: before reweighting. Right: after reweighting ("official" 69 mb). The ratios after reweighting become flattest at an "unofficial" value of 71 mb. The apparent substantial differences in the QCD "integrals" are artifacts of the plotting procedure, in which bins with negative contributions are zero-ed.

energy-momentum matching is measured by taking the ratio between the corrected energy E_{corr} in the Ecal supercluster and the momentum of the track P_{in} measured in the inner layers of the tracker.

Electron selections have two main components, electron identification (eID) and electron isolation. In this analysis we use the non-triggering MVA electron identification. The MVA cuts used to define the 80% and 90% signal efficiency working points are summarized in Table 3.9. In all channels, the identification and isolation used follows the POG recommended criteria. The exact discriminator names and working points for each channel are listed and described in their respective sections.

3.5.10   Muon Reconstruction and Identification

Muon reconstruction is a multistep process that begins with the information gathered from the muon subdetector. As a first step, standalone muons are reconstructed from hits in the individual drift tube (DT) and cathode strip (CSC) chambers. Hits from the innermost muon stations are combined with hits in the other muon segments using the Kalman fitting technique. The standalone muon trajectory is reconstructed by extrapolating from the innermost muon station to the outer tracker surface. This standalone trajectory is then used to find a matching track reconstructed in the inner silicon tracker. Finally, standalone muons and matching silicon tracks are used to perform a global fit resulting in a "global" muon. Muon reconstruction is described in more detail in [?].

Global muons are reconstructed by combining tracker muons from the inner silicon tracker and standalone muons from the muon chambers. Once a muon is required to have matching tracks in the inner and outer detectors, the main source of background consists of charged hadrons that leave a signature in the inner silicon tracker while also penetrating through the hadronic calorimeter and creating hits in the muon chambers. Charged hadrons that penetrate the hadronic calorimeter and leave hits in the muon system will deposit significant energy in the calorimeters. Therefore, a calorimeter compatibility algorithm can be used to significantly reduce the number of charged pion fakes. However, calorimeter compatibility is not used in this analysis due to our uncertainty of the

performance of such algorithms in the presence of high PU. The presence of punch-throughs often occur due to pions from the fragmentation of quarks and gluons. These punch-throughs can often be discriminated against by requiring isolation. Similarly, non-prompt muons from heavy flavor decays and decays in flights are expected to be within jets and can be discriminated against by imposing an isolation requirement. Muon identification is described in more detail in [**?**] and [**?**].

Isolated muons are required to have minimal energy from PF neutral and charged candidates in a cone of $\Delta R = 0.4$ around the lepton trajectory. PF charged candidates considered in the calculation of isolation are required to be near the primary vertex. Isolation for muons is defined as:

$$I = \frac{\sum_i p_T^i}{p_T^\mu} \tag{3.3}$$

where the index i runs over PF neutral and charged candidates. Table 3.10 shows the complete list of for the "isMedium" μ identification criteria used in this analysis. In all channels, the identification and isolation used follows the POG recommended criteria.

3.5.11 Tau Reconstruction and Identification

Responsible: A. Johnson, F. Romeo, J. Cumalat

The challenge in identifying hadronically decaying taus is discriminating against generic quark and gluon QCD jets which are produced with a cross–section several orders of magnitude larger. CMS has developed several algorithms to reconstruct and identify hadronically decaying taus based on Particle Flow (PF) objects. For this analysis, the tau POG recommended Hadron Plus Strips algorithm (HPS) is used. HPS makes use of PF jets as inputs to an algorithm that uses strips of clustered electromagnetic particles to reconstruct neutral pions. The electromagnetic strips ("neutral pions") are combined with the charged hadrons within the PFJets to attempt to reconstruct the main tau decay modes outlined in Table 3.11.

The single hadron plus zero strips decay mode attempts to reconstruct $\tau \to \nu \pi^\pm$ decays or $\tau \to \nu \pi^\pm \pi^0$ decays where the neutral pion has very low energy. The single hadron plus one or

two electromagnetic strips attempts to reconstruct tau decays that produce neutral pions where the resulting neutral pion decays produce collinear photons. Similarly, the single hadron plus two strips mode attempts to reconstruct taus that decay via e.g. $\tau \to \nu \pi^{\pm} \pi^0$ where the neutral pion decays to well separated photons resulting in two electromagnetic strips. The three hadrons decay mode attempts to reconstruct tau decays that occur via $\rho(770)$ resonance. Although it is possible to recover signal in the two hadron decay mode (in the case of one of the three prongs being misidentified), this mode is not considered as its inclusion reduces discrimination performance and hurts the limit. In all cases, electromagnetic strips are required to have $E_T > 1$ GeV/c. Additionally, the particle flow charged hadrons are required to be compatible with a common vertex and have a net charge of $|q| = 1$.

In order to enforce the isolation requirement on the reconstructed tau, a region of size R = 0.5 around the tau decay mode direction is defined. Any PF candidates not used for the reconstruction of electromagnetic strips and charged hadrons not involved in the reconstruction of the tau decay modes are used to calculate isolation. The "Tight", "Medium", and/or "Loose" 3-hits isolation (with $\delta\beta$ corrections) working points are used.

In order to discriminate against muons, HPS taus are required to pass the lepton rejection discriminator which requires the lead track of the tau not be associated with a global muon signature. In order to discriminate against electrons, HPS taus are required to pass a MVA discriminator which uses the amount of HCAL energy associated to the tau with respect to the measured momentum of the track (H/p). Additionally, the MVA discriminator considers the amount of electromagnetic energy in a narrow strip around the leading track with respect to the total electromagnetic energy of the tau. Finally, HPS taus must not reside in the ECAL cracks. In all channels, the identification and isolation used follows the Tau POG recommended criteria. The exact discriminator names and working points for each channel are listed and described in their respective sections.

3.5.12   Efficiency of Tau Identification discriminators

The efficiency of the τ_h ID discriminators used in the analysis are studied using Z'_{SSM} samples with $m(Z') = 2000$ GeV. For this purpose, we require the reconstructed τ_h to have $p_T > 45$ and pseudorapadity $|\eta| < 2.1$. Further, reconstructed τ_h is matched to generator-level tau with $\Delta R(\tau_{reco}, \tau_{gen}) < 0.3$. The efficiency of the Decay Mode Finding (DMF) discriminator "Decay-ModeFinding" is found to be relatively flat at $\sim 80\%$ as shown in Figure 3.11. The individual effiencies of anti-muon , anti-electron and isolation discriminators, relative to the DMF criterion, are shown in Figures 3.12– 3.14. The overall efficiency of τ_h ID selection used in this analysis is $\sim 55\%$. We use the VLoose working point of the anti-electron MVA5 discriminator as the tighter working points have poor performance (i.e. low efficiency which is also not "flat" vs. p_T). The relative flatness of the τ_h identification efficiency with p_T will also facilitate the use of $Z \rightarrow \tau\tau$ events as a "standard candle" for comparison with signal.

Figure 3.11: Decay Mode Finding efficiency as function of p_T, η, and number of vertices of generator-level taus

3.5.13   Tau Energy Scale and Resolution

Since the resolution and scale of our mass reconstruction depends on the effectiveness of the τ_h reconstruction, in this section we summarize our studies on τ_h response and resolution. We define the response as the difference between the transverse momentum of a reconstructed tau (that has passed all tau ID discriminators) and the transverse momentum of a generated tau that has been

AgainstMuon_4_29_16.pdf

AgainstMuon_eta_4_29_16.pdf

AgainstMuon_nvtx_4_29_16.pdf
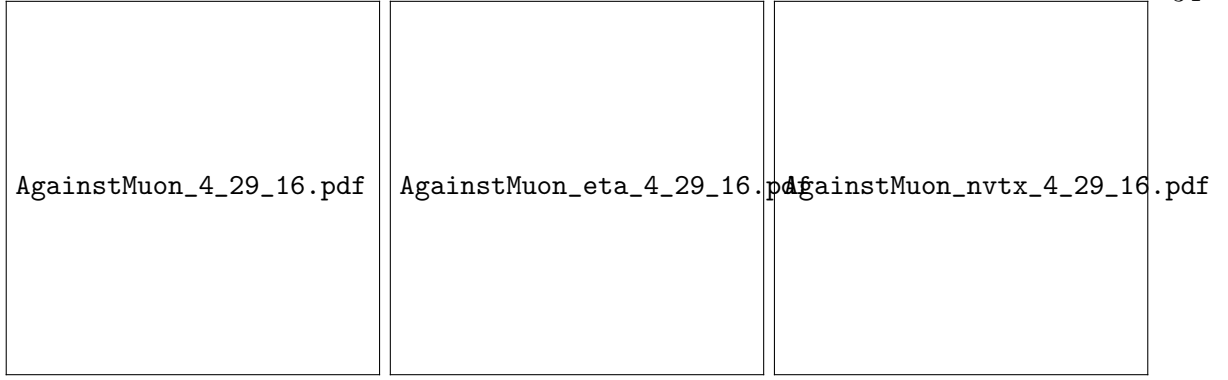
Figure 3.12: Efficiency of anti-muon discriminator (Loose3) as function of p_T, η, and number of vertices of generator-level taus

AgainstElectron_4_29_16.pdf

AgainstElectron_eta_4_29_16.pdf

AgainstElectron_nvtx_4_29_16.pdf

Figure 3.13: Efficiency of anti-electron discriminator (Loose/Medium MVA5) as function of p_T, η, and number of vertices of generator-level taus

CombinedIso_4_29_16.pdf

CombinedIso_eta_4_29_16.pdf

CombinedIso_nvtx_4_29_16.pdf

Figure 3.14: Efficiency of isolation discriminator ("CombinedIsoDB3Hits") as function of p_T, η, and number of vertices of generator-level taus

matched ($\Delta R < 0.3$) to the reconstructed tau. We can see from Figure 9 (right) that the response

distribution contains a narrow gaussian like component in the bulk, in addition to a relatively long tail (compared to electrons and muons). While the tails become less substantial at high p_T, the gaussian-like bulk of the response distributions broaden at high p_T.
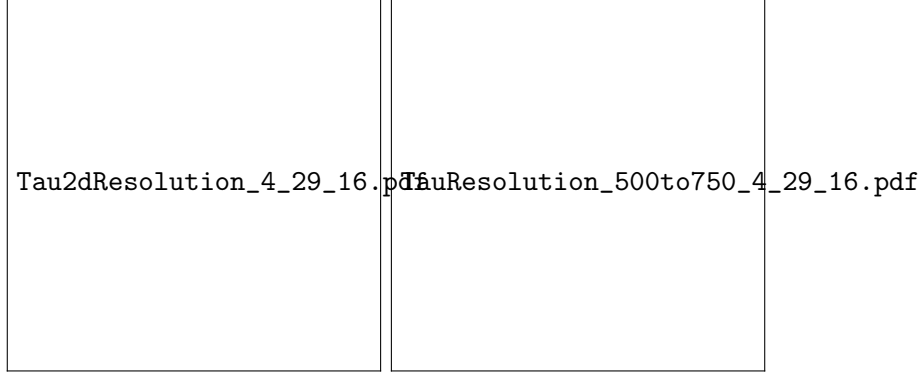
Figure 3.15: Energy response of reconstructed taus, raw and vs. generated tau p_T

3.5.14   Jet Reconstruction

Particle-flow (PF) technique [?, ?] is used to improve the jet p_T and angular resolution in this analysis. The PF technique combines information from different subdetectors to produce a mutually exclusive collection of particles (namely muons, electrons, photons, charged hadrons and neutral hadrons) that are used as input for the jet clustering algorithms. Jets are clustered using the anti-k_T algorithm [?], with a distance parameter of $R = 0.4$ in η-ϕ plane (defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) by summing the four-momenta of individual PF particles.

The jets require energy corrections obtained using simulated events that are generated with PYTHIA, processed through a detector simulation based on GEANT4, and confirmed with in situ measurements of the p_T balance. The overall jet-energy corrections depend on the η and p_T values of jets. The jet-energy corrections are applied by L1 FastJet, L2 Relative, and L3 Absolute corrections. In order to remove the extra energy in jets from underlying event (UE) and pileup (PU), the L1 FastJet corrections use the event-by-event UE/PU densities. The L2 and L3 corrections use jet balancing and photon+jet events to improve and provide a better energy response as a function of jet p_T and η. For data, additional residual corrections are applied.

Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$. For the identification of jets the loose PF ID is used in this analysis. Table 4 shows the selection criteria used for the recommended loose PF ID, which are validated in other studies [**?**]. The jet reconstruction and ID efficiency in simulation is >98%.

3.5.14.1   b-Jet Tagging

In this analysis, b-tagged jets are used for two purposes: to reduce $t\bar{t}$ background in the signal region and to obtain $t\bar{t}$ enriched control samples used to estimate the signal rate.

The CSVv2 algorithm [**?**] is used to identify jet as originating from hadronization of a b-quark. This algorithm combines reconstructed secondary vertex and track-based lifetime information to build a likelihood-based discriminator to distinguish between jets from b-quarks and those from charm or light quarks and gluons.

The minimum thresholds on these discriminators define loose, medium, and tight operating points with a misidentification probability of about 10%, 1%, and 0.1%, respectively, with an average jet p_T of about 80 GeV. The loose operating point for an efficiency about 85% is used in this analysis. A large sample of pair-produced top quark events is used to measure b-tagging efficiency using several methods [**?**]. A scale factor is applied to correct for differences in b-tagging efficiency between data and simulation [**?**].

3.5.15

The presence of neutrinos in the $\tau\tau$ decays must be inferred from the imbalance of total momentum because they escape from the detector without producing any direct response. The magnitude of the negative vector sum of the visible transverse momentum is known as missing transverse energy, denoted E_T^{miss}.

Missing transverse energy is one of the most important observables for discriminating the signal events from background events which do not contain neutrinos, such as QCD multijet and DY$\to \mu\mu$/ee events. Because there are many factors which may result in artificial E_T^{miss}, compre-

hensive studies of E_T^{miss} as measured using the CMS dector are performed [**?**]. We require events with $E_T^{miss} > 30$ GeV in all channels.

Finally, the standard and recommended "MET Filters" are utilized in this search.

3.6   Muon + Hadronic Tau Channel

The motivation for analyzing events where one τ lepton decays to a muon, while the other decays to hadrons is the same for all $\tau^+\tau^-$ related analyses: because muons have the lowest jet misidentification among leptons, the mere requirement of a muon removes a substantial amount of background processes, especially the QCD multijet background. Once this requirement is made, a main source of background is due to Drell-Yan processes giving rise to τ leptons. Because we seek particles with masses much larger than that of the Z boson, this source of background can be easily discriminated against by looking at larger reconstructed $\tau^+\tau^-$ mass regions (reconstructed mass to be defined later). This process, however, can also serve as a "standard candle" to validate the τ_h identification criteria and ensure the robustness of the analysis. Other main sources of background include (1) QCD multijet events where light-quark or gluon jets are misidentified as a τ_h, (2) W + jets events where the W boson decays to a muon and a jet is misidentified as a τ_h, and (3) $t\bar{t}$ events where two leptons can come from the prompt decay of W bosons or one mis-identified τ_h from a jet. The sum of the Drell-Yan and W + jets backgrounds represent approximately 89% of the total background (according to simulation) in this channel. The cuts used to select $\mu\tau_h$ pairs are factorized in to four categories: acceptance, μ identification, τ_h identification, and topological cuts. Acceptance criteria is completely driven by the limitations of the CMS detector and the need to maximize analysis sensitivity while also minimizing systematic effects. For example, in order to minimize systematic effects, the p_T threshold on the muon leg is chosen such that it falls on the plateau of the trigger turn-on curve ($p_T > 30$ GeV ... see section 3). Although it is possible to achieve slightly better sensitivity by increasing the thresholds, the selections are also driven by the need to obtain a sample enriched with $Z \to \tau\tau$ events, with minimal modifications to the final selection criteria. Figure 3.16 shows the p_T distributions for signal and background processes
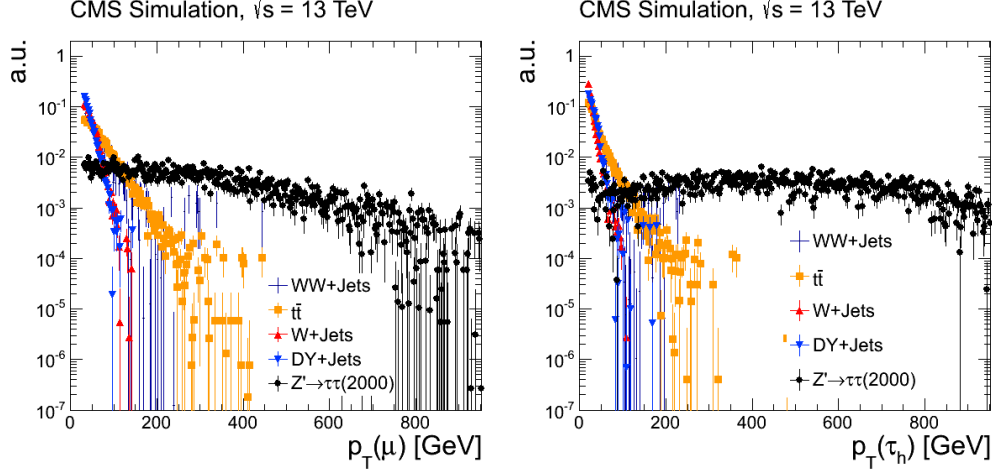
Figure 3.16: Left: $p_T(\mu)$ after acceptance, μ identification, and τ_h identication cuts. Right: $p_T(\tau_h)$ after acceptance, μ identification, and τ_h identication cuts.

relevant to this analysis. As discussed in section 5.3, the muon identification criteria is designed mostly to discriminate against cosmic muons, punch-through pions, and muons associated to jets from decays in flight. The τ_h identification is described in section 5.4 and is mostly designed to discriminate against hadronic jets produced from the fragmentation of quarks and/or gluons. Finally, topological cuts are mainly used to minimize the remaining W + jet(s) and $t\bar{t}$ contributions that remain after the muon and τ_h identification criteria has been imposed.

W boson production in association with jets becomes a dominant background because a clean, well-reconstructed muon is produced by the W boson. Therefore, the requirement of a clean muon signature does not provide additional discrimination. Additionally, the neutrino from the W boson decay acquires an average energy of approximately $m_W/2 \sim 40$ GeV. Because the neutrino will escape the CMS detector undetected, the average measurement of the momentum imbalance will be approximately 40 GeV. Therefore, our requirement on the momentum imbalance in the event (> 30 GeV) does not provide significant discrimination against this background. Figure 3.17 shows the distribution for signal and background processes relevant to this analysis. Finally, the presence of "associated jets" means that the contamination from W + jets in the signal region is highly dependent on the jet$\to \tau_h$ fake rate, which is the largest amongst leptons. Therefore,
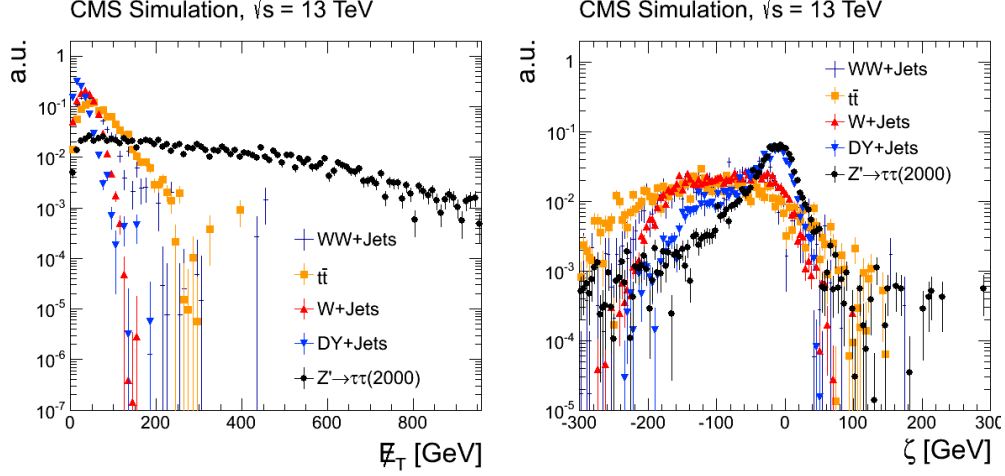
Figure 3.17: Left: after acceptance, μ identification, and τ_h identication cuts. Right: $P_\zeta - 3.1 \times P_\zeta^{vis}$ after acceptance, μ identification, and τ_h identication cuts.

reducing W + jets also requires additional topological requirements.

In W + jet(s) events, unlike $X \to \tau\tau$ resonance production such as $Z' \to \tau\tau$, where the τ-lepton decay products are expected to be back-to-back in ϕ, the presence of the neutrino from the W decay and the uncorrelated jet gives rise to topologies where the jet and the lepton are not back-to-back (Figure 3.18). Therefore, one of the main discriminating variables against W + jets events is the difference in ϕ between the jet and muon directions. Figure 3.22 shows the $\cos\Delta\phi(\mu, \tau_h/jet)$ distributions for Z'$\to \tau\tau$ and the relevant background samples. We require $\cos\Delta\phi(\mu, \tau_h/jet) < -0.95$, which is approximately $> 90\%$ efficient for $Z' \to \tau\tau$ and approximately 25% efficient for W+jets.

For $\tau^+\tau^-$ final states, the in the event is due to the neutrinos from the τ-lepton decays and is expected to point in the direction collinear to the visible tau decay products. Furthermore, the measurement of is completely correlated to the visible tau decay products. In W + jet(s) events, the direction and magnitude of the momentum imbalance is completely correlated to the lepton from the W boson, but uncorrelated to the jet. We require events to be consistent with this signature of a particle dacaying to two τ leptons by defining a unit vector along the bisector of visible tau decay products ($\hat{\zeta}$) and two projection variables, p_ζ and p_ζ^{vis}:
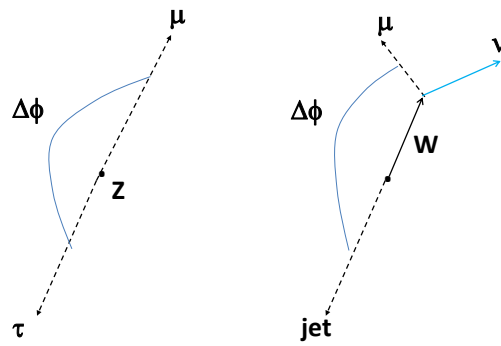
Figure 3.18: Sketch depicting the W+jets rejection power of a $\Delta\phi$ cut.

$$p_\zeta^{vis} = \overrightarrow{p}_{\tau_1}^{vis}\hat\zeta + \overrightarrow{p}_{\tau_2}^{vis}\hat\zeta \tag{3.4}$$

$$p_\zeta = p_\zeta^{vis} + \overrightarrow{\not{E}_{\mathrm{T}}}\hat\zeta \tag{3.5}$$

The sketch in Figure 3.19 displays the definitions for p_ζ^{vis} and p_ζ. Figures 3.20 and 3.21 shows the separation between $Z' \to \tau\tau$ and W + jet(s) events in the p_ζ-p_ζ^{vis} plane. For the case of W+jets, there is no strong correlation between p_ζ^{vis} and p_ζ due to the presence of a jet that is uncorrelated to the μ and ν_μ from the W boson. However, there is a strong correlation for the case of $Z' \to \tau\tau$. To discriminate against W+jet(s) events, requirements on $\Delta\phi(\tau_1, \tau_2)$ and ζ are applied. ζ is defined as a linear combination of p_ζ and p_ζ^{vis}:

- $cos\Delta\phi(\tau_1, \tau_2) < -0.95$

- $p_\zeta - 3.1p_\zeta^{vis} > -50$

Although events containing $t\bar{t}$ contribute to the expected background in all channels containing light leptons, it's contribution to the $\mu\tau_h$ channel is only $\sim 1.4\%$ of the total background (according to simulation). For $\mu\tau_h$ final states the $t\bar{t}$ contribution comes in the form of a real light lepton from the semileptonic decay of the $W^\pm$ and a fake τ_h from the hadronic decay of the second $W^\pm$. These events are characterized by an isolated light lepton, passing all lepton identification and isolation requirements, accompanied by a non–isolated "hadronic tau" due to the larger multiplicity of the hadronically decaying W boson. These events are suppressed with the use of topological cuts.

After only applying lepton identification and isolation requirements, a non-negligible background contribution from $t\bar{t}$ events remain. These events can be further suppressed with cuts that take advantage of the very different topologies between $Z' \to \tau\tau$ and $t\bar{t}$ events. The first, and most important, of these differences is the presence of b–jets in the event. For our purposes, jets with $p_T > 30$ GeV/c and $|\eta| < 2.4$ are counted as b–tagged jets if the "combined secondary vertex" discriminator, described in section 5, returns a value consistent with that of a b–jet. In this analysis, the "loose" operating point of the "combined secondary vertex" discriminator (CSVL) is used.
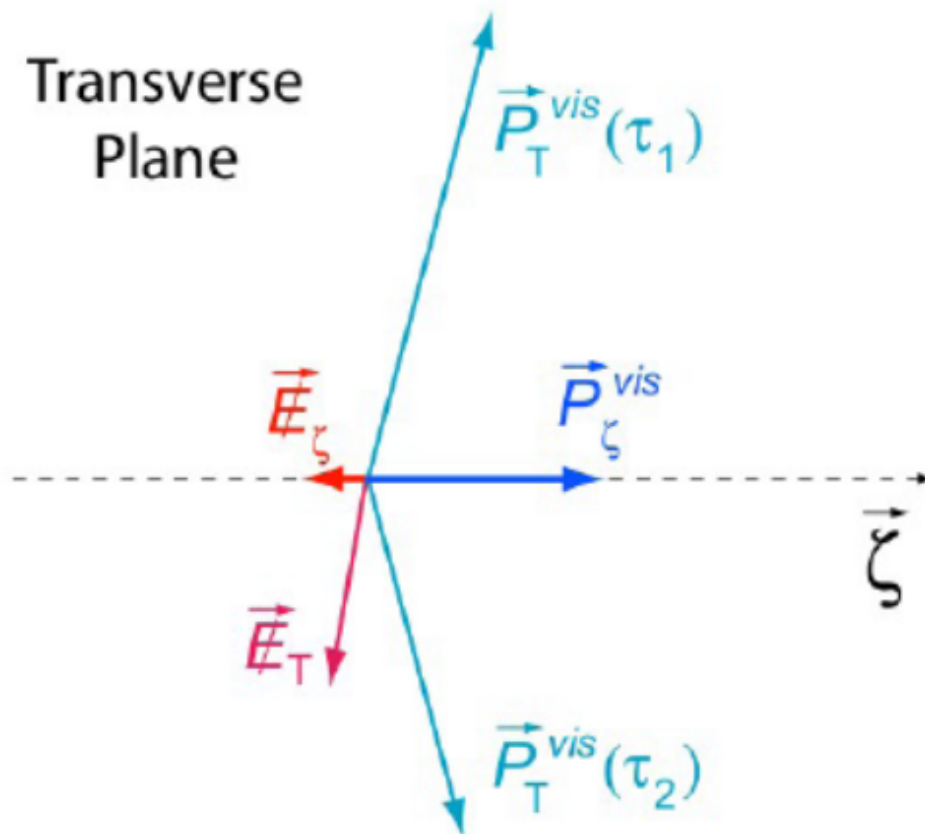
Figure 3.19: Definitions for p_ζ and p_ζ^{vis}.
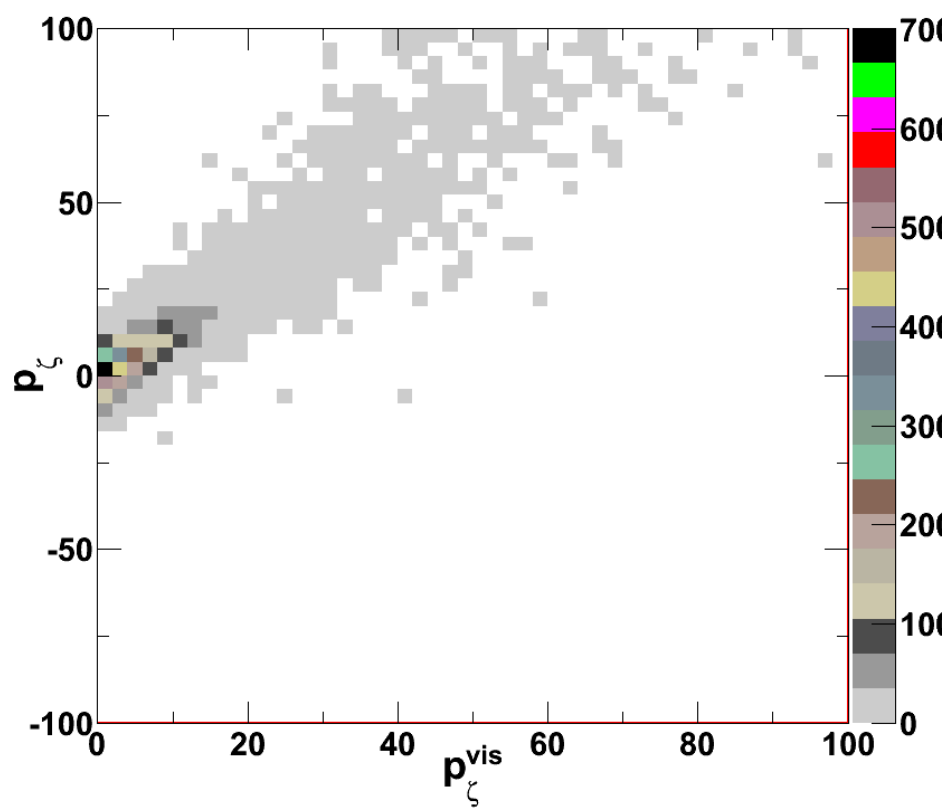
Figure 3.20: $p_\zeta \ p_\zeta^{vis}$ for $Z' \to \tau\tau$.
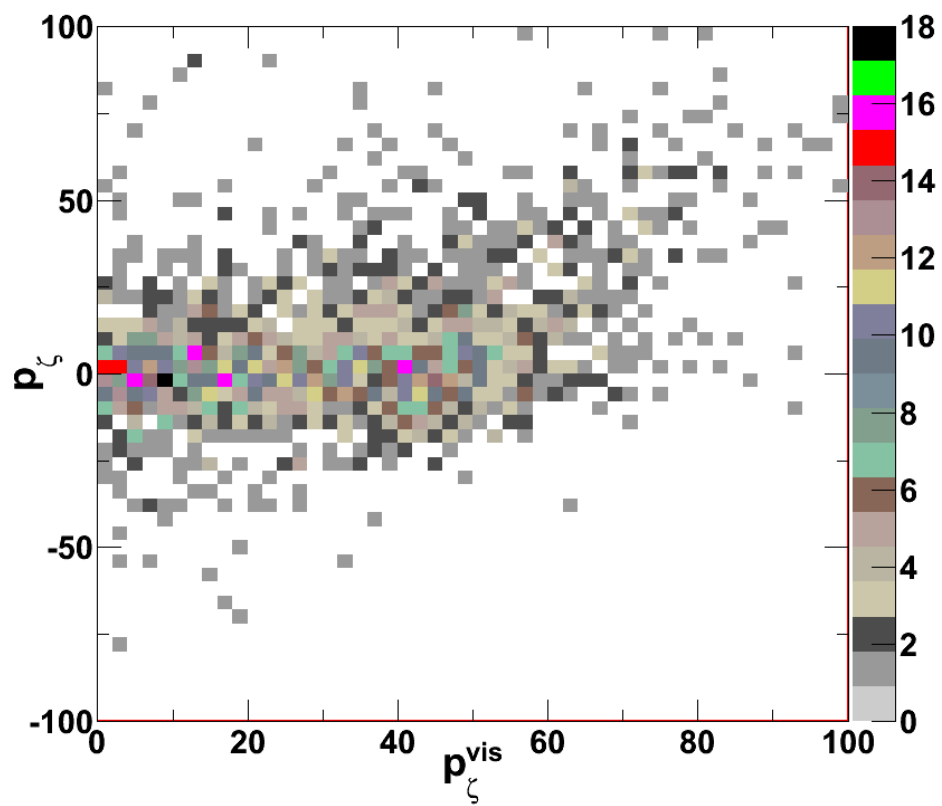
Figure 3.21: p_ζ p_ζ^{vis} for W+Jets.

The entire signal selection criteria are summarized below, while Table 3.13 is the cut flow efficiency table (yields normalized to $\sigma \cdot L \cdot \varepsilon$).

Acceptance Selection:

- Events must fire the HLT_IsoMu18 trigger

- ≥ 1 global μ with $|\eta| < 2.1, p_T > 30$ GeV

- ≥ 1 HPS τ_h with $|\eta| < 2.1, p_T > 20$ GeV

- $\Delta R(\mu, \tau_h) > 0.5$

μ Identification:

- "isMediumMuon"

- Relative isolation (with $\delta\beta$ corrections) < 0.15

τ_h Identification:

- Muon veto: "againstMuonTight3"

- Electron veto: "againstElectronMVAVLooseMVA5"

- "new" decay mode finding with 1 or 3 signal charged hadrons

- Isolation: "byTightCombinedIsolationDeltaBetaCorr3Hits"

Topological requirements:

- $\cos\Delta\phi(\mu, \tau_h) <$ -0.95. The τ_h jet direction is calculated using the sum of the four-momenta of decay mode constituents:

$$\overrightarrow{p}_{\tau_h} = \sum_i \overrightarrow{p}^{\,i}_{signal\ constituents} \tag{3.6}$$

- $Q(\mu) \times Q(\tau_h) < 0$

- > 30 GeV

- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$

- 0 jets tagged as b-jets

Because τ leptons decay to neutrinos which leave the detector undetected, one cannot fully reconstruct the mass resonance with the visible τ decay products. Additionally, because the invariant mass for background processes such as QCD are typically steeply falling distributions in the tails (where new mass resonances are expected), it becomes important to make use of to attempt to separate signal from background and reconstruct the true mass resonance. Historically, several methods such as the collinear approximation have been employed to reconstruct the true mass resonance. However, for the analysis presented, one of the main sources of backgrounds is W + jet(s). In this case, the analysis achieves a high degree of sensitivity by requiring the μ and τ_h candidates to be back-to-back in ϕ (see above description). This is precisely the regime in which the collinear approximation fails. Therefore, the mass is reconstructed as follows:

$$M(\tau_1, \tau_2, \not{E}_{\mathrm{T}}) = \sqrt{(E_{\tau_1} + E_{\tau_2} + \not{E}_{\mathrm{T}})^2 - (\overrightarrow{p_{\tau_1}} + \overrightarrow{p_{\tau_2}} + \overrightarrow{\not{E}_{\mathrm{T}}})^2} \qquad (3.7)$$

where E_τ and $\overrightarrow{p_\tau}$ represent the energy and 4-momentum of the visible objects (in this case the μ and τ_h), respectively. The energy component of $\overrightarrow{\not{E}_{\mathrm{T}}}$ is the scalar quantity , while the z-component is the default "met.pz()" component in the object collection. As can be seen from Figure 3.22, this definition successfully distinguishes between lower mass production of τ-lepton pairs and high-mass τ-lepton pairs from new massive resonant particle production.

Figure 3.23 shows the $p_T(\tau_h)$, $\eta(\tau_h)$, , and $m_T(\mu,)$ control plots using events satisfying the "preselection" requirements (i.e. acceptance + μ identification + τ_h identification cuts described above). Figure 3.24 shows the $p_T(\tau_h)$, $\eta(\tau_h)$, , and $m_T(\mu,)$ control plots in the signal region. In these particular set of plots, the background predictions are entirely based on MC. In general, there is good agreement between the observed distributions and the predictions from simulation.
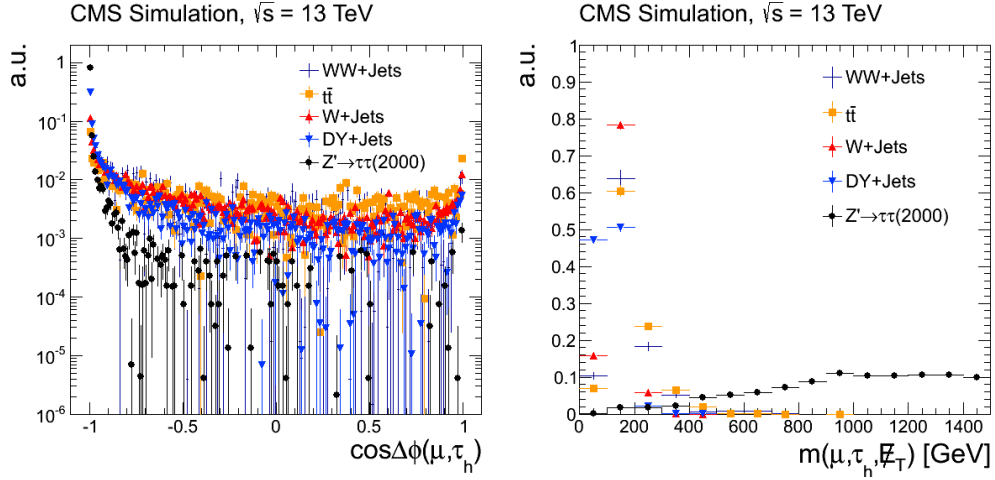
Figure 3.22: Left: $\cos\Delta\phi$ between the muon and tau/jet directions. Right: $m(\mu, \tau_h,)$ for signal Z' and various backgrounds.
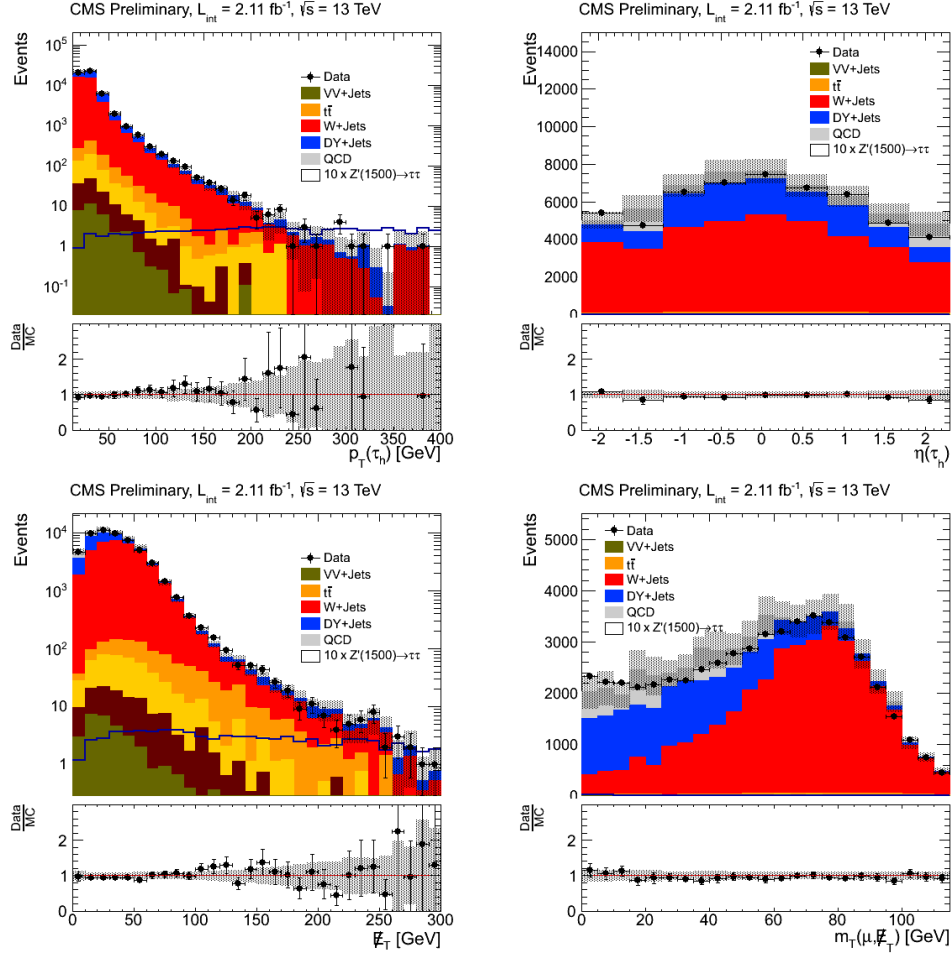
Figure 3.23: $p_T(\tau_h)$, $\eta(\tau_h)$, , and $m_T(\mu,)$ control plots using events satisfying the "preselection" requirements. The background predictions are entirely based on MC.
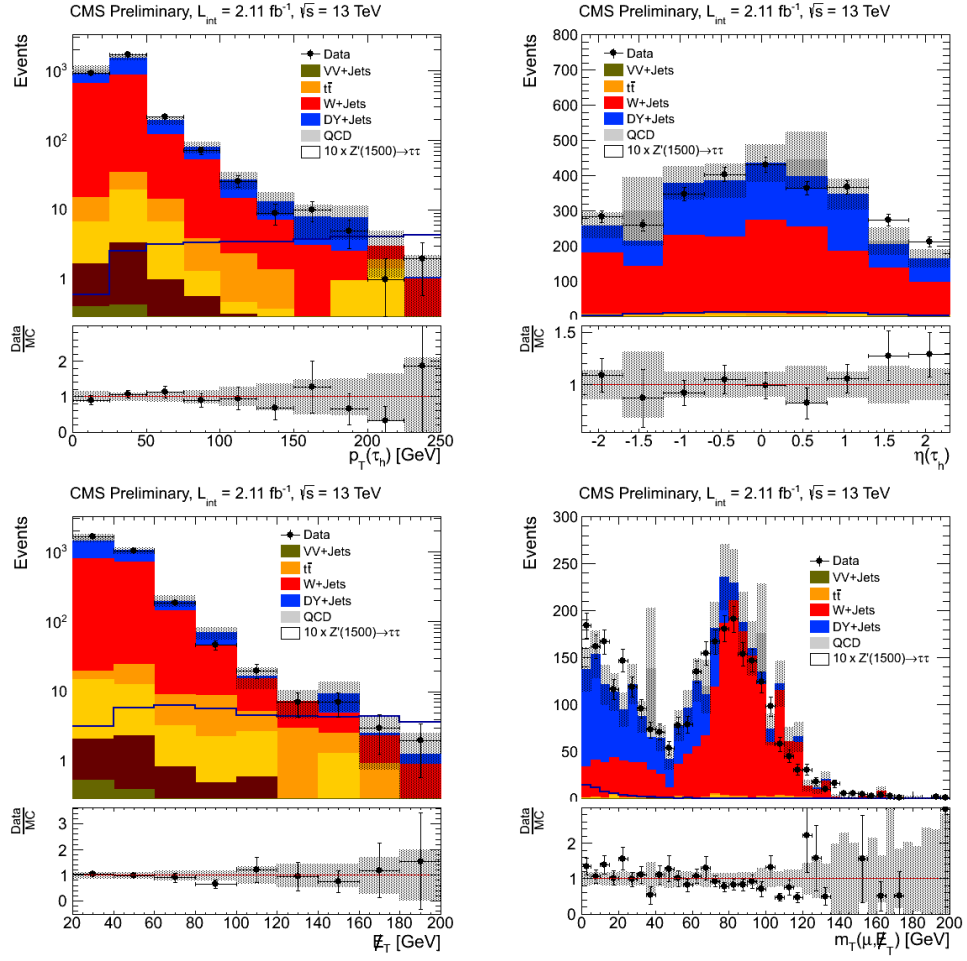
Figure 3.24: $p_T(\tau_h)$, $\eta(\tau_h)$, , and $m_T(\mu,)$ control plots using events satisfying the signal region requirements. The background predictions are entirely based on MC.

3.6.1   W + jets Background Estimation

As discussed above, the main discriminators against W + jets events are the topological variables ζ and $\cos\Delta\phi(\mu, \tau_h)$ in addition to τ_h isolation. Thus, it is only natural to construct the W + jets background estimation methodology using control samples obtained by inverting these requirements. We estimate this background using a completely data-driven approach which relies on the classic ABCD method. The regions ABCD are defined as follows:

- A: fail one or both of the ζ and $\cos\Delta\phi$ cuts; fail "Tight" but satisfy a relaxed τ_h isolation of < 5 GeV

- B: fail one or both of the ζ and $\cos\Delta\phi$ cuts; pass "Tight" τ_h isolation

- C: pass both the ζ and $\cos\Delta\phi$ cuts; fail "Tight" but satisfy a relaxed τ_h isolation of < 5 GeV

- D: pass both the ζ and $\cos\Delta\phi$ cuts; pass "Tight" τ_h isolation

We estimate the W + jets component N_W^i in regions $i = A, B, C$ by subtracting MC non-W backgrounds from data $(N_W^i = N_{\text{Data}}^i - N_{\neq W}^i)$. We then estimate the W + jets component in the signal region D, to be $N_W^D = N_W^C \cdot \frac{N_W^B}{N_W^A}$. Said differently, we take the yield of W + jets events in data containing a non-isolated τ_h and then extrapolate to the signal region by correcting for the "Tight-to-Loose" ratio (also referred to as a "fake rate"), which is measured in a data sample enriched with W + jets events and that is obtained by inverting the ζ and $\cos\Delta\phi$ cuts. The shape of the $m(\mu, \tau_h,)$ distribution is obtained from control region C (nominal selections with non-isolated τ_h).

Figure 3.25 shows a comparison of the ζ and $\cos\Delta\phi$ distributions in W + jets MC, normalized to unity, under two different τ_h isolation conditions: (i) isolated (passing ""Tight"), and (ii) non-isolated (failing "Tight" but passing loosened isolation). The τ_h isolation does not bias the ζ and $\cos\Delta\phi$ distributions, and thus we expect our ABCD background estimation method to nicely model the "true" W + jets yield in the signal region. A closure test for the background estimation method
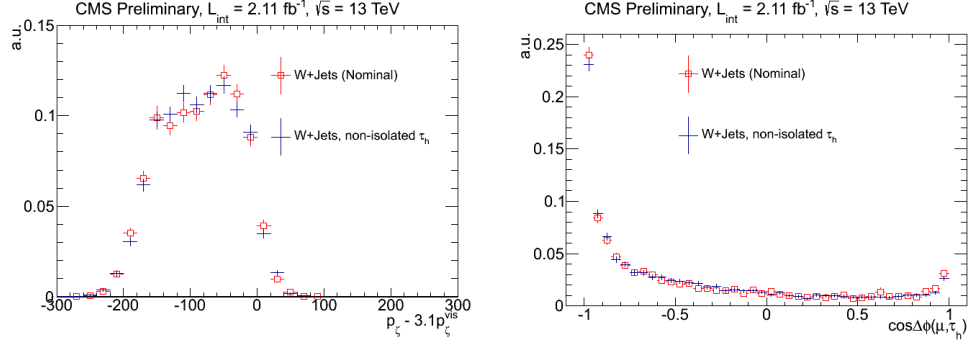
Figure 3.25: Left: Comparison of the ζ distribution in W + jets MC, normalized to unity, for events with isolated and non-isolated τ_h. Right: Comparison of the $\cos\Delta\phi$ distribution in W + jets MC, normalized to unity, for events with isolated and non-isolated τ_h.

outlined above is performed with MC. Two aspects are simultaneously tested: (1) closure on the normalization (i.e. $N_W^D = N_W^C \cdot \frac{N_W^B}{N_W^A}$); (2) correct determination of the $m(\mu, \tau_h,)$ shape. Figure 3.26 shows the closure test in MC. The y-axis of the top left plot is in normal scale so the reader can focus on the level of agreement in the low mass region. The top right plot is similar, except the y-axis is in log scale in order to emphasize the high-mass tails. We observe very good agreement between the nominal yield/shape and the predicted yield/shape. The reader can perform their own cross-check of this conclusion by using the MC-based W + jets yields in Table 3.14, which shows the data and MC background yields in the control samples, and plugging them into the equation $N_W^{\text{Prediction},MC} = N_W^{C,MC} \cdot \frac{N_W^{B,MC}}{N_W^{A,MC}}$. The MC-based prediction using the ABCD method is $N_W^{\text{Prediction},MC} = (1112.06\pm53.65) \cdot \frac{(5235.70\pm115.55)}{(7591.91\pm139.17)} = 766.92\pm43.05$, while the MC-based nominal yield is 757.9 ± 35.0 (see Table 3.13). Therefore, the MC-based nominal yield to predicted yield ratio is $\frac{N_W^{\text{Nominal},MC}}{N_W^{\text{Prediction},MC}} = \frac{757.9\pm35.0}{766.9\pm43.1} = 0.998 \pm 0.05$, which is statistically consistent with unity. Thus, no additional systematic uncertainties are applied due to closure.

As mentioned above, Table 3.14 shows the data and MC background yields in control regions A, B, and C. The purity of W + jets, based on simulation, ranges from $\sim 65 - 81\%$, depending on the sample. The W + jets scale factors, defined as SF$= \frac{N^{\text{Data}}-N_{\neq W}^{MC}}{N_W^{MC}}$, show about a $\sim 20\%$ mis-modeling in simulation. The top row of plots in Figure 3.27 are the $m_T(\mu,)$ distributions in control regions C(top left) and A(top right). The previously mentioned scale factors have been
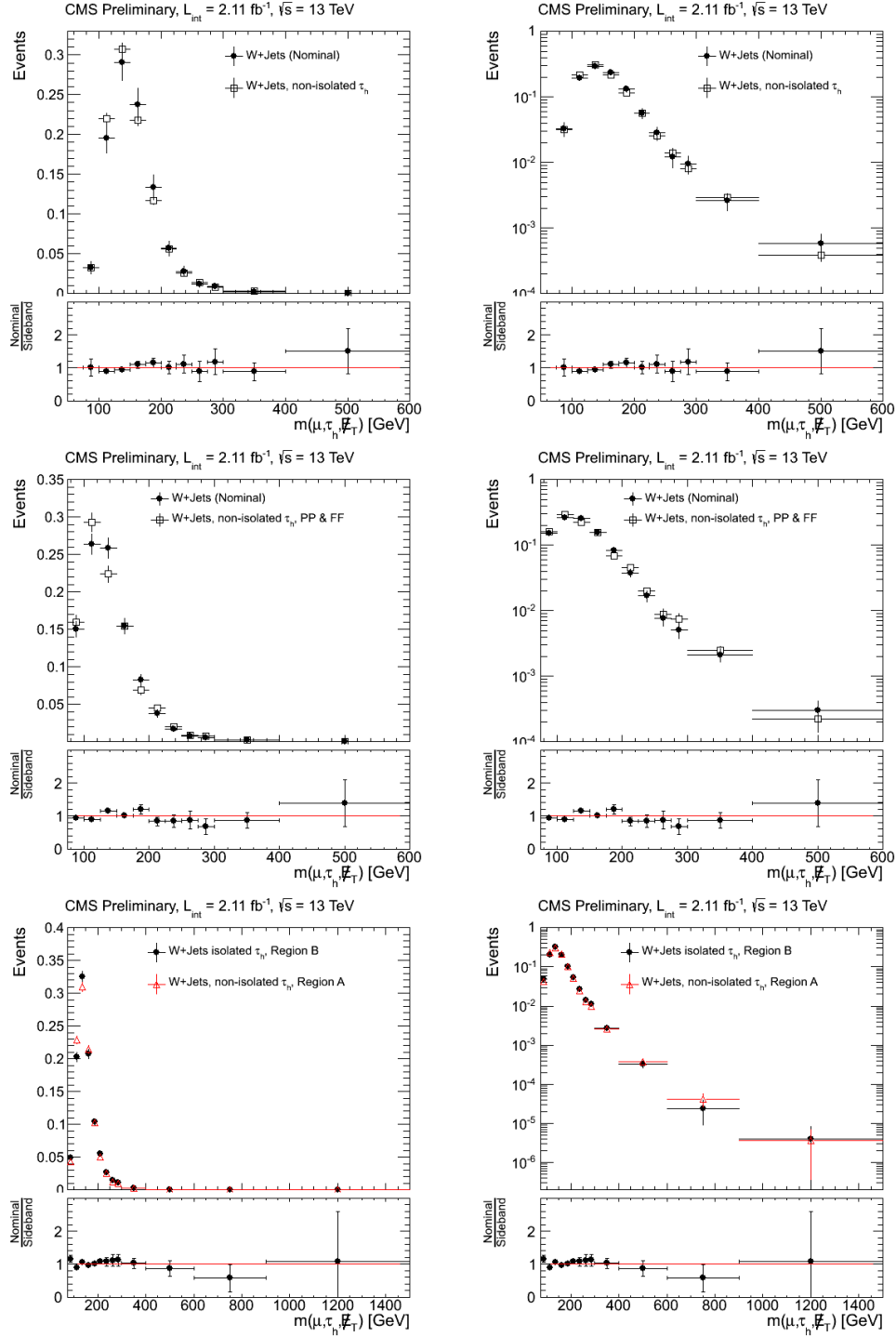
Figure 3.26: Top Left: W + jets closure test performed with simulation (normal scale to focus on low mass) under nominal conditions (i.e. > 30 GeV). Top Right: W + jets closure test performed with simulation (log scale to focus on the high mass tails) under nominal conditions (i.e. > 30 GeV). Middle Left: W + jets closure test performed with simulation (normal scale to focus on low mass), requiring > 0 GeV. Middle Right: W + jets closure test performed with simulation (log scale to focus on the high mass tails), requiring > 0 GeV. Bottom Left: Comparison of the reconstructed mass distributions in regions A and B, in normal scale. Bottom Right: Comparison of the reconstructed mass distributions in regions A and B, in log scale.

applied to the plots.

A natural comment or question from the reader is that while the MC closure test looks good, it is possible that mis-modeled in simulation could perhaps pull the closure test in the wrong direction. In other words, how stable is the background estimation method with respect to . The middle row of plots in Figure 3.26 show similar closure tests, but without a cut. We find that closure is observed even without the cut, adding to the robustness of the methodology. Table 3.15 shows the data and MC background yields in control regions A, B, and C obtained without the requirement. The purity of W + jets, based on simulation, has decreased in comparison to the purity observed with > 30 GeV, due to the larger contribution from DY + jets and QCD multijet.

The procedure outlined in this section yields a W + jets estimate of $N_W^{\text{Prediction,Data}} = (1238.31 \pm 111.78) \cdot \frac{(6387.64 \pm 201.46)}{(8633.38 \pm 237.17)} = 916.20 \pm 91.15$ (see Data - $\sum_{i \neq W} BG_i$ in Table 3.14). The uncertainty on the background prediction is $\sim 10\%$ and is mostly dominated by $\sqrt{N_{\text{Data}}}$ in control regions A, B, and C. We note that the main source of contamination in Table 3.14 is DY + jets, which we've normalized using a data-to-MC correction factor described in the next section. However, even if DY + jet(s) yields in the W control samples are taken entirely from simulation, it only provides a small effect on the W + jets prediction in the signal region. This is true because the systematic uncertainty for DY + jets is typically $< 15\%$ (see section 10), while the level of non-W background contamination is at most $\sim 35\%$ (i.e. upper bound on systematic effect is $0.35 \cdot 0.15 \sim 5\%$). The total predicted background $m(\mu, \tau_h,)$ spectrum will be shown in the results section of the AN.

3.6.1.1 W + jets Background Validation with High mT

The standard approach to estimate the W + jets background in the Higgs search is to use a control sample of events with high $m_T(\mu,)$. The W + jets yield in that region is determined by subtracting the non-W predictions from data using simulation, and subsequently extrapolating to the signal region using an extrapolation factor determined from MC. We do not employ a similar approach in this analysis as we are motivated by wanting a fully data-driven approach to
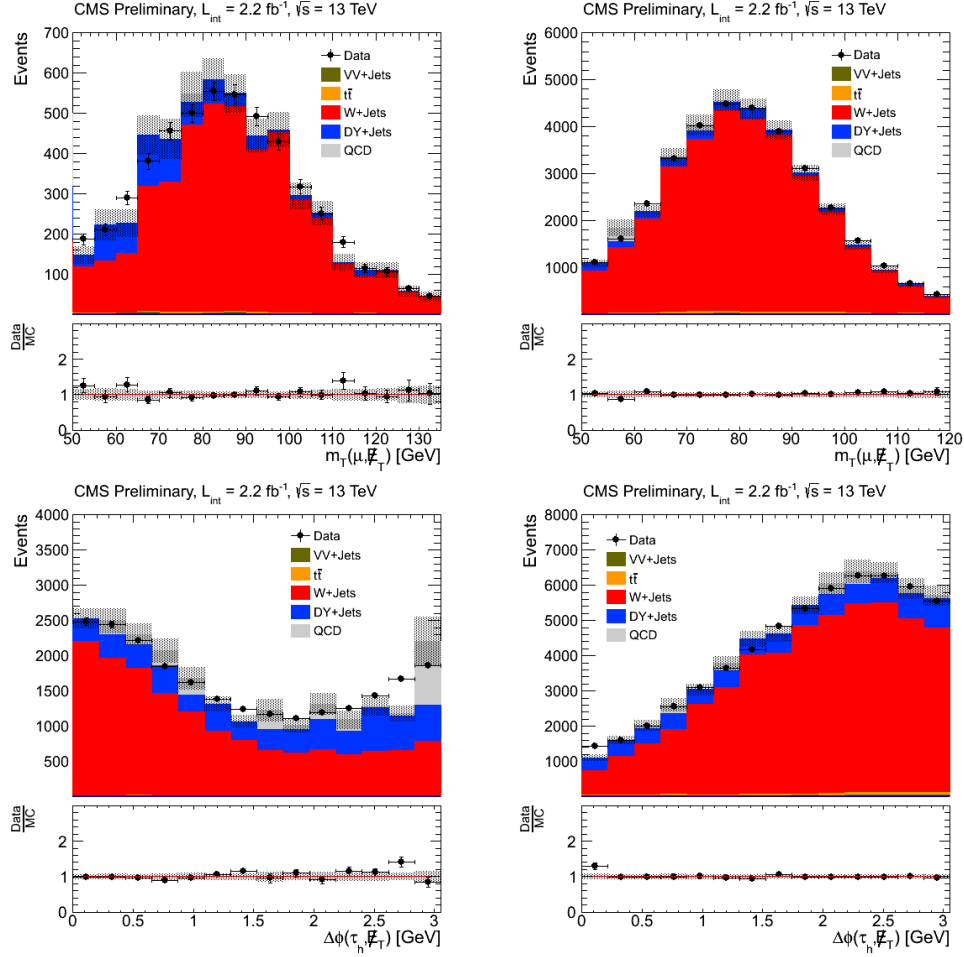
Figure 3.27: Top Left: $m_T(\mu,)$ distributions in W + jets control region C, under nominal conditions (i.e. > 30 GeV). Top Right: $m_T(\mu,)$ distributions in W + jets control region A, under nominal conditions (i.e. > 30 GeV). Bottom Left: $m_T(\mu,)$ distributions in W + jets control region C, requiring > 0 GeV. Bottom Right: $m_T(\mu,)$ distributions in W + jets control region A, requiring > 0 GeV.
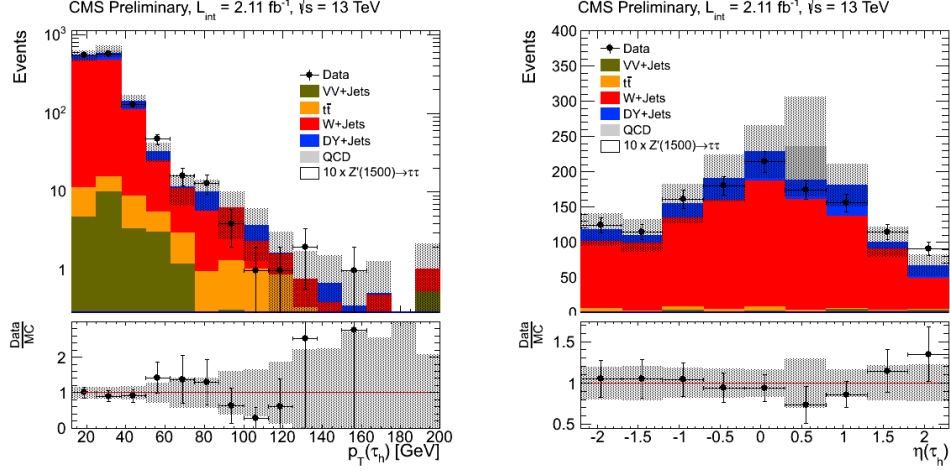
Figure 3.28: $p_T(\tau_h)$ and $\eta(\tau_h)$ distributions in the W + jets high $m_T(\mu,)$ control region.

estimating both the normalization and the mass shape. However, we use a similar control sample of high $m_T(\mu,)$ events to validate our background estimation method and provide further confidence in our method. For this purpose we select events with $m_T(\mu,) > 70$ GeV in addition to the signal region selections. The MC-based W + jets yield in this high $m_T(\mu,)$ control sample is corrected using a scale factor equal to the ratio of the data-driven and nominal W + jets yields in the signal region. Table 3.16 shows the data and background yields in this high $m_T(\mu,)$ control sample, while Figure 3.28 shows the $p_T(\tau_h)$ and $\eta(\tau_h)$ distributions. The measured scale factor, 0.90 ± 0.11, is consistent with unity and therefore provides further validation of our background estimation methodology.

3.6.2   DY + jets Background Estimation

Obtaining a semi-clean sample of $Z \to \tau\tau$ events is one of the most important aspects of the analysis as it needs to be shown the τ_h identification criteria employed is successful at identifying a τ_h. Unlike W + jets, where a jet is mis-identified as a τ_h and whose contribution to the signal region is expected to be mis-modeled by simulation, there is no reason to expect significant disagreement between data and MC for $Z \to \tau\tau$. The efficiency for the requirement of a high quality muon and τ_h candidate is expected to be well modeled by simulation, and has indeed

been validated by the POGs. Moreoever, the in $Z \to \tau\tau$ events is "real" missing momentum from the escaping neutrinos produced by the τ-lepton decay, and thus is also expected to be well modeled by simulation. Therefore, the estimate of the DY + jets contribution ($\sim 98\%$ $Z \to \tau\tau$) is based on both simulation and data. We utilize a scale factor, defined as the ratio of efficiencies measured in data and simulation, to normalize the predicted rate to the data by correcting the expected DY + jets contribution obtained from MC. A semi-clean sample of $Z \to \tau\tau$ events is obtained with the following modifications to the final signal selections:

- < 30 GeV

- $m(\mu, \tau_h,) < 300$ GeV

It is important to note that the analysis has been designed to achieve sensitivity in the high mass region, while also enabling us to achieve a relatively clean $Z \to \tau\tau$ signature in the low mass region with only minor modifications to the final selections. Figure 3.29 shows the $m(\mu, \tau_h,)$ invariant mass in the $Z \to \tau\tau$ control region defined above, while the observed events in data and predicted background yields are listed in Table 3.17 (W + jets and QCD multijet backgrounds are completely data-driven).

Table 3.17 and Figure 3.29 clearly show that both the $Z \to \tau\tau$ rates and shapes are consistent between data and MC, with a SF approximately 0.96 ± 0.06. Therefore, the expected $Z \to \tau\tau$ contribution in the signal region is taken from MC. The expected number of $Z \to \tau\tau$ events in the signal region is 882.9 ± 34.9 (see Table 3.13). Obtaining a semi-clean sample of $Z \to \tau\tau$ events where the number of observed events are consistent with the MC expectation ensures the robustness of the analysis and validates the ability to successfully identify a τ_h. Furthermore, since our τ_h identification efficiency is "flat" vs. $p_T(\tau_h)$, validating correct modeling of $Z \to \tau\tau$ at low $p_T(\tau_h)$ indirectly "validates" that we can correctly identify a high-p_T τ_h. The total predicted background $m(\mu, \tau_h,)$ spectrum will be shown in the results section of the AN.
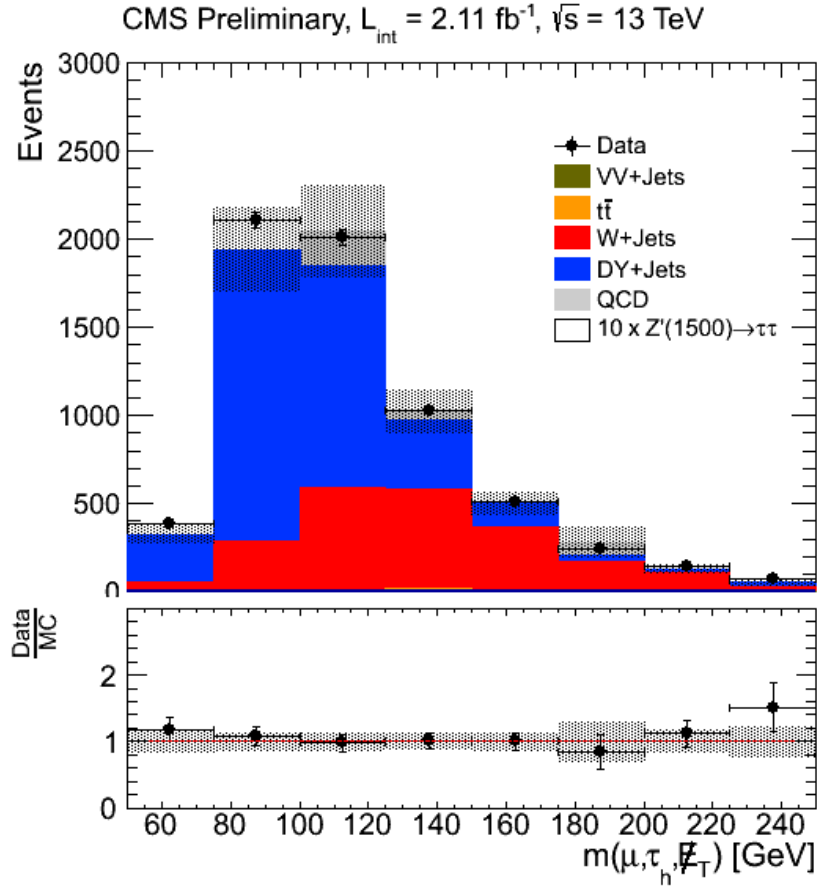
Figure 3.29: $m(\mu, \tau_h,)$ distribution in the $Z \to \tau\tau$ control region defined by < 30 GeV and $m(\mu, \tau_h,) < 300$ GeV, with otherwise similar selections to the signal region.

3.6.3 Other Small Backgrounds

As mentioned in the introduction to section 3.6 and as can be calculated from Table 3.13, the sum of the Drell-Yan and W + jets backgrounds represent approximately 93% of the total background in this channel. The remaining $t\bar{t}$, diboson, and QCD multijet processes compose $\sim 1.3\%$, 1.4%, and 4.4% of the total background respectively. Therefore, these backgrounds are taken directly from simulation with appropriate systematic uncertainties, which are discussed and summarized in section 10. Furthermore, the $t\bar{t}$ and diboson backgrounds are expected to be well modeled by simulation and thus, for added confidence, we only perform a validation of the "good enough" modeling of the QCD multijet background. For this purpose, we obtain a control sample enriched with $\sim 91\%$ purity of QCD multijet events by inverting the isolation on both the τ_h and μ legs as well as inverting the $\mu\tau_h$ charge requirement (select like-sign). Table 3.18 summarizes the observed data events and predicted background yields in this sample. The measured data-to-MC correction factor of SF $= 1.09 \pm 0.27$ is consistent with unity, and thus the QCD multijet yields (in the signal and control regions) are taken directly from simulation. The QCD multijet contribution to the signal region is conservatively assigned an additional 27% systematic uncertainty based on the uncertainty on the correction factor from this control sample (this is on top of the $\sim 63\%$ statistical uncertanty).

3.7 The $\tau_h\tau_h$ Channel

Responsible: A. Johnson, J. Cumalat, A. Kalsi

A search for new heavy gauge bosons is presented using events with two hadronically decaying τ leptons. Because of the resemblance of QCD jets with τ_h, the probability of misidentifying a QCD jet as a τ_h is at least an order of magnitude higher than that for a QCD jet to be misidentified as an electron or a muon. As a result , the final state is highly contaminated by QCD multijet background ($> 90\%$ of the background in the signal region). But, typical signal events are expected to appear at fairly high invariant mass values, where the QCD multi-jet contribution is strongly reduced owing

to its rapidly falling production cross-section versus $\sqrt{s}$. Thus QCD multijet background only moderately affects the sensitivity of the analysis. Apart from QCD multijets, the other prevailing background is Drell-Yan processes giving rise to τ leptons. The large branching ratio of $\tau\tau \to \tau_h\tau_h$ (about 42%) makes its contribution to the sensitivity of the overall analysis highly important.

Similarly to other channels, the selections—designed to discriminate between the signal and background—are divided into: kinematic and geometric acceptance for selecting $\tau_h\tau_h$ pairs, τ_h identification, and topological selections. The main difference with the analyses of the $e\mu$, $\mu\tau_h$ and $e\tau_h$ channels are the substantially tighter $p_T(\tau_h)$ requirements in order to stay efficient with respect to the trigger and targeting at the suppression of QCD multijet backgrounds (the exact selections are described below). Note that surviving pairs of τ_h candidates are preserved at each intermediate stage in the selections. In events in which more than one pair of unique τ_h candidates passes all the selections, only that pair with the highest $m(\tau_h, \tau_h,)$ is selected. This requirement has a very high efficiency for both signal and backgrounds (the fraction of events with more than one pair is $\ll 1\%$).

Events fired by the $HLT\_DoubleMediumIsoPFTau35\_Trk1*$ trigger are considered as the interesting events for offline analysis. Events are required to have at least 2 HPS taus with p_T greater than 60. These taus are required to have pseudorapidity coverage of $|\eta| < 2.1$. A τ_h candidate is also required to satisfy the the reconstruction and identification criteria described in section 3.5. The τ_h candidates are required to pass the following discriminators : "DecayModeFindingNewDMs" , "TightCombinedIsoDB3Hits" , "againstMuonTight3" and "againstElectronLooseMVA5". The efficiency of these dicriminators is shown in section 3.5.12. These discriminators ensure the proper identification of a τ_h against QCD jets, muons and electrons. The candidate $\tau_h\tau_h$ pairs are required to be separated in $\eta - \phi$ space by $\Delta R(\tau_h, \tau_h) > 0.3$. Further, to reduce top pair contamination the event is required not to have any jet identified as a b-quark jet by the btagging algorithms using the combined secondary vertex loose (CSVL) working point. Further, each event must have at least 30 of missing transverse energy to account for the neutrinos present in signal and further discriminate against the QCD multijet background. For consistency with the other channels, only

1- and 3-prong taus are considered.

The $\tau_h\tau_h$ pair is expected to be back-to-back with $\cos(\delta\phi(\tau_h,\tau_h)) < -0.95$. In addition, a "ζ" cut is used ($p_\zeta - 3.1 \cdot p_\zeta^{vis} > -50$) which is explained and optimized in section 6. These are topological cuts which reduce the contamination of backgrounds mainly from $t\bar{t}$ and W+jets processes to negligible levels.

3.7.1   QCD Background Estimation & Validation Strategy

The main background for $\tau_h\tau_h$ final state is QCD multijet events ($> 90\%$) and evaluated by a data-driven approach as a MC program does not model the background properly, nor does it provide sufficient statistics to make any strong conclusions. The number of events in the signal region is given by the following equation:

$$N_{\text{Signal}}^{\text{QCD}} = N_{\text{LS}}^{\text{QCD}} \cdot R_{OS/LS}, \tag{3.8}$$

Here $N_{\text{LS}}^{\text{QCD}}$ and $R_{OS/LS}$ are evaluated using the following approach. For its estimation, we rely on the classical method of counting events selected in a similar way as the signal events but selecting $\tau_h\tau_h$ pairs with like-sign electrical charge. This leads to events heavily dominated by the QCD multi-jet background. Assuming that QCD dijets are charge-blind, the number of like-sign events N_{LS} should be equal to the number of opposite-sign N_{OS} QCD multi-jet events after correcting N_{LS} measured in data for known contamination from electroweak backgrounds using simulation. However, the assumption of the charge symmetry in events with two jets is not always true. An asymmetry in the charges of jets in multi-jet events can arise from the remaining correlation between the quark charge and the leading track charge of the jet in events where quark charges are correlated. Note that the correlation between the charge of the quark and the charge of the track becomes stronger in jets, in which the entire jet fluctuates into just a few high momentum tracks. Calculation of QCD N_{LS} and N_{OS} terms are made in control regions of the lower sidebands for the calculation of the asymmetry factor $R_{OS/LS}$ (shown in the next section). The contamination from signal in the LS control regions are small since the charge mis-measurement is small ($\sim 1-5\%$)

for $m(Z') < 2.5$ TeV.

The QCD estimation and validation strategy used in this analysis is shown in Figure 3.30 .

QCD_Strategy.jpg

Figure 3.30: QCD estimation and validation strategy for the $\tau_h\tau_h$ channel.

The shape of the $m(\tau_h, \tau_h,)$ distribution is obtained from control region C (same-sign $\tau_h\tau_h$ with nominal). To extract the OS/LS ratio from data , two control regions 1B and 1D are obtained by keeping the same selections as signal selections but inverting the cut (< 30 GeV) and requiring OS and LS $\tau_h\tau_h$ pairs respectively. The contribution of non-QCD MC backgrounds are subtracted from data in these control regions and then the $R_{OS/LS}$ is measured:

$$\mathrm{N}^{\mathrm{QCD}}_{\mathrm{OS}} = N^{\mathrm{Data}}_{\mathrm{OS}} - N^{\text{non-QCD MC}}_{\mathrm{OS}}$$
$$\mathrm{N}^{\mathrm{QCD}}_{\mathrm{LS}} = N^{\mathrm{Data}}_{\mathrm{LS}} - N^{\text{non-QCD MC}}_{\mathrm{LS}}$$
$$\mathrm{R}_{\mathrm{OS/LS}} = N^{\mathrm{QCD}}_{\mathrm{OS}} / N^{\mathrm{QCD}}_{\mathrm{LS}}$$

Table 3.19 shows the data and MC yields in controls regions 1B and 1D. The purity of QCD multijet, defined by Data - $\sum_i BG_i$, is approximately $96 - 99\%$ depending on the sample. The measured OS/LS ratio is 1.64 ± 0.21. The above equation shows the mathematical procedure used to obtain this ratio.

Closure and validation tests for the background estimation method outlined above is performed with real data, since there are insufficient statistics to perform such a test with simulation. Two aspects are simultaneously tested: (1) closure on the normalization (i.e. $N^A_{\mathrm{QCD}} = N^C_{\mathrm{QCD}} \cdot \frac{N^B_{\mathrm{QCD}}}{N^D_{\mathrm{QCD}}}$); (2) correct determination of the $m(\tau_h, \tau_h,)$ shape. In order to check (with data) whether same-sign $\tau_h\tau_h$ events can correctly model the mass shapes in the opposite-sign regions, we perform a shape closure/validation test by taking the shape from region 1D (obtained as Data - $\sum_i BG_i$) and normalize it to the QCD yield in control region 1B. By comparing the shape for the QCD prediction in region 1B with the observed mass spectrum in the same region, we can determine whether same-sign $\tau_h\tau_h$ correctly models the mass shapes in the opposite-sign region. Furthermore, any disagreement in the shape between data and the QCD prediction can be used to assign systematic uncertainty on the shape. Figure 3.31 shows the $m(\tau_h, \tau_h,)$ mass distribution for this closure test in control region 1B. We observe very good agreement between the observed shape and the predicted shape and thus no additional systematic uncertainties are applied due to this particular closure test on the shape.

Two additional control samples are utilized in order to provide further closure for this method. Control regions 2B and 2D are obtained by using the τ_h isolation sidebands (passing the "loose" isolation working point and failing the "tight" isolation working point) with otherwise similar selections to control regions 1B and 1D (low). The QCD multi-jet prediction in control region 2B
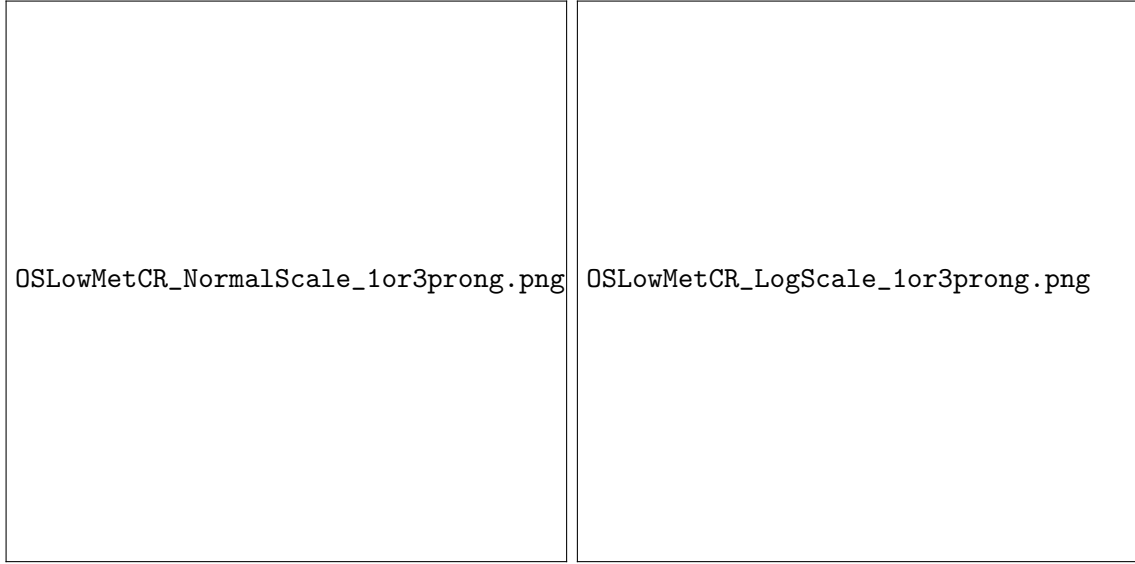
Figure 3.31: $\tau_h - \tau_h$ mass distribution in isolated OS, low- sideband (Left: normal scale. Right: log scale).
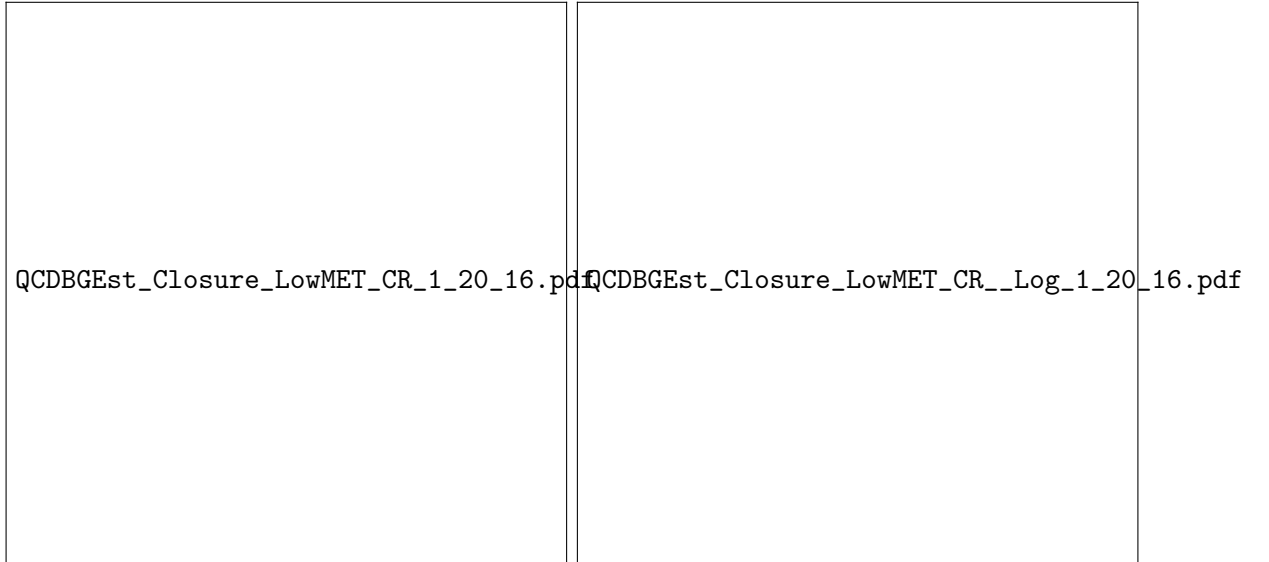
Figure 3.32: $m(\tau_h, \tau_h,)$ mass distribution in anti-isolated OS, low- sideband (Left: normal scale. Right: log scale).

is determined as data minus non-QCD backgrounds. The QCD shapes and rates extracted from control region 2D (after subtracting non-QCD backgrounds) are normalized to the QCD yield in control region 2B. Figure 3.32 summarizes this closure test. Nice agreement between the data and predicted QCD rate in control region 2B (Figure 3.32) validates the ability to extract the shapes of distributions in OS from LS samples (we will extract the QCD signal region $m(\tau_{h,1}, \tau_{h,2},)$ shape from LS data).

Next we perform the full-blown closure test for the ABCD method by comparing the observed yield and mass spectrum in control region $2A$ (OS, non-isolated $\tau_h \tau_h$ with nominal) with the QCD multijet prediction in that same region obtained by using the QCD shape from region $2C$ and normalizing it to $N_{\text{QCD}}^{2A} = N_{\text{QCD}}^{2C} \cdot \frac{N_{\text{QCD}}^{2B}}{N_{\text{QCD}}^{2D}}$. Figure 3.33 shows the closure test in region $2A$, showing agreement between data and prediction. The QCD multijet prediction in control region 2A, using the ABCD method, is 427 ± 26, while the observed yield is 429. Thus, no additional systematic uncertainties are applied due to closure.
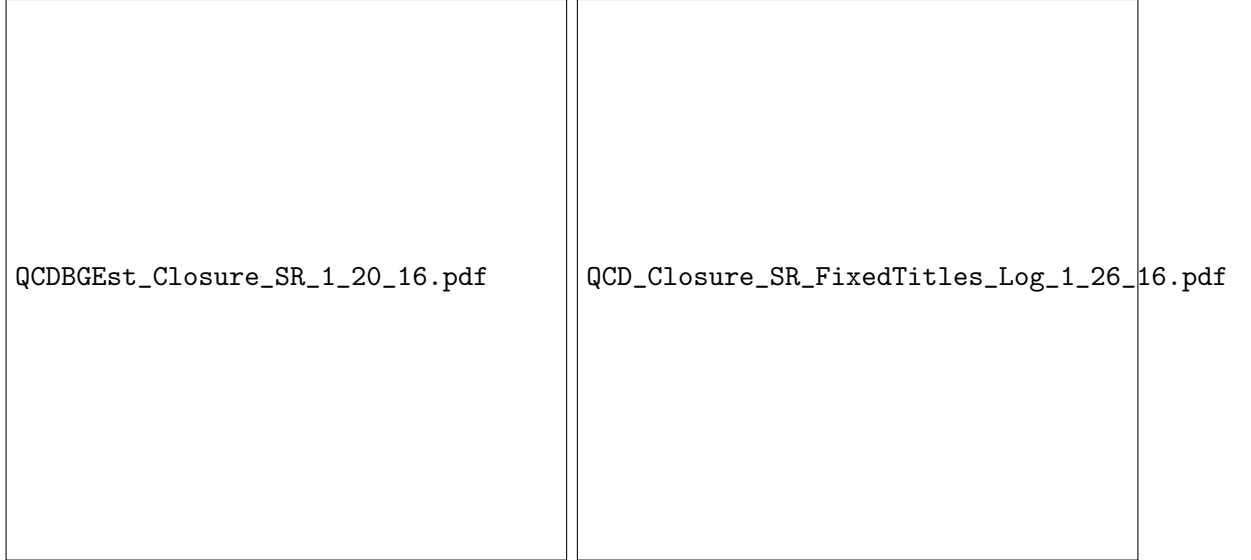
Figure 3.33: $m(\tau_h, \tau_h,)$ mass distribution in anti-isolated OS region with nominal (Left: normal scale. Right: log scale).

Table 17 represents the event rate of data and backgrounds in regions 1A (signal region) and 1C (same-sign, isolated $\tau_h \tau_h$ with nominal). QCD shapes/rates are extracted in data-driven

way from control region 1C and scaled by the OS/LS ratio determined from control region 1B and 1D. The non-QCD backgrounds are taken directly from MC in Table 17 (see the next section for a validation of the DY + jets background yield). Other processes such as W + jets, $t\bar{t}$, and diboson represent only $\sim 1\%$ of the total background rate in the signal region, and are thus taken directly from MC. The final prediction of QCD events in the signal region 1A is given in the right-most column of Table 18. The procedure outlined in this section yields a QCD estimate of $N_{\text{QCD}}^{\text{Signal}} = 48.72 \pm 11.0$. The uncertanty is based on the statistics of the data and MC samples. We stress that this is the QCD predicted rate over the entire $m(\tau_h, \tau_h,)$ spectrum. As was mentioned in the strategy section of this note, we fit for a potential signal that would appear as an excess of events over the standard model expectation in the high $m(\tau_h, \tau_h,)$ part of the distribution. The total predicted background $m(\tau_h, \tau_h,)$ spectrum will be shown in the results section of the AN.

3.7.2   Background Estimation for $Z(\to \tau\tau)$ + jets

We do not employ a complete data-driven measurement of the $Z \to \tau\tau$ + jets contribution to the signal region. Instead, we take an approach based on both simulation and data. The efficiency for the requirement of at least two high quality τ_hs is expected to be well modeled by simulation. Therefore, the estimate of the $Z \to \tau\tau$ + jets contribution is determined by obtaining a control sample used to validate the correct modeling of the requirement of at least two high quality τ_hs. Since the DY + jets background in this channel is only $< 10\%$ of the total background in the signal region, the above approach is sufficient.

As discussed above, the typical probability of misidentifying a QCD jet as a τ_h is at least an order of magnitude higher than that for a QCD jet to be misidentified as a light lepton. As a result the QCD multijet background in the $\tau_h \tau_h$ channel (the focus of this note) is substantially higher than in lepton plus tau or dilepton channels. One should note that the presence of large multijet background mainly complicates the definition of suitable control regions for validating the agreement between collision data and simulation for other backgrounds. For this purpose, the events are selected using the "pre-selection" cuts, and additionally requiring $\tau_h \tau_h$ pairs with invariant mass

less than 100 GeV in order to obtain a semi-clean sample of Z→ $\tau\tau$ events. Figure **??** shows the $m(\tau_h\tau_h)$ distribution for this validation sample where the QCD contribution has been determined using the method discussed above. One can see that the rate and shape between data and MC is consistent. The measured Z→ $\tau\tau$ data-to-MC scale factor is $SF_{\textbf{preselection}} = 0.97 \pm 0.19$.

3.8   Electron + Hadronic Tau Channel

3.8.1   Event selection

Events must fire the single-electron trigger described in Section 3.3. We select reconstructed electrons satisfying:

- $p_T > 35$ and $|\eta| < 2.1$

- distance of closest approach to the leading sum-p_T^2 primary vertex of less than 0.045 cm (transvese) and 0.2 cm (longitudinal)

- passing the tight working point of the e/γ POG non-triggering MVA ID

- having no matched conversion nor missing hits

- within $\Delta R < 0.5$ of the HLT electron that fired the trigger

Offline τ_h's are required to have:

- $p_T > 20$ and $|\eta| < 2.1$

- distance of closest approach to the leading sum-p_T^2 primary vertex of less than 0.2 cm (longitudinal)

- pass the new Decay Mode Finding requirement as either a 1-prong or 3-prong τ_h

- pass the "againstElectronVLooseMVA5" and "againstMuonTight3" identification requirements

We build pairs of electrons and τ_h's in which the electron and τ_h are separated by at least $\Delta R > 0.5$. In events with more than one such pair, we select the pair with the two most isolated leptons, considering first the electron, and then the τ_h. This criterion was seen to have good efficiency for signal samples. In the rare case of multiple such pairs having identical isolation values, the reconstructed p_T's are considred, preferring higher values.

After a pair has been chosen for an event, we apply the following isolation requirements on the leptons, for an event to enter the signal region: electron relative isolation < 0.15; τ_h isolation "byTightCombinedIsolationDeltaBetaCorr3Hits." In order to keep the different final states exclusive, an event is rejected if there is an additional electron satisfying the above identification requirements and with relative isolation < 0.3, or a muon satisfying the identification requirements described in Section 3.9.1 with relative isolation < 0.3. To reduce further possible di-electron events in the channel, an event is rejected if there is an opposite-charge electron pair with $\Delta R > 0.15$ in which both of the electrons satisfy $p_T > 15$, $|\eta| < 2.5$, the e/γ POG "veto" ID, and relative isolation < 0.3.

As for the other channels, the signal region is defined as having

- $\cos \Delta\phi(e, \tau_h) < -0.95$

- $Q(e) \times Q(\tau_h) < 0$

- > 30

- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$

- no jet with $p_T > 30$ tagged as a b-jet (CSV loose)

The distributions of these variables after preselection are shown in Figures 3.35, **??**, 3.37.

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Figure **??**.

The Standard Model processes considered as backgrounds are Drell-Yan, di-boson production, top quark single and pair production, W+jets production, and QCD multi-jet production.
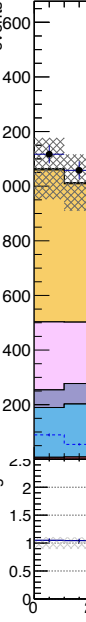
3.8.2 Genuine dilepton events

Studies of simulated events indicate that for Drell-Yan process, top quark single and pair production, and di-boson production, the reconstructed and selected electrons and hadronic taus are typically associated with genuine simulated leptons. The nominal expected event rates are estimated by scaling the simulated samples by the best available cross sections, listed in Table 3.4, and by the integrated luminosity of the data samples.

3.8.2.1 Drell-Yan process

Due to large W+Jets and QCD contamination, as shown in the left panel of Figure 3.39 with the following selections:

- < 30 ;

- no jet with $p_T > 30$ tagged as a b-jet (CSV loose);

- < 200 GeV   ,

we use the Drell-Yan rate systematic uncertainty (12%) estimated from the final state in 3.9.2.1.

3.8.2.2 $t\bar{t}$ and single top processes process

For , we estimate the $t+\bar{t}+$ single top production rate systematics (8%) using the final state, as described in 3.9.2.2, due to its higher top purity. However we cross check this in a top-rich region, defined by the following selections and shown in the right panel of Figure 3.39:

- $\cos\Delta\phi(e, \tau_h) < -0.95$

- > 30

- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$

- at least one jet with $p_T > 30$ tagged as a b-jet (CSV loose)

The $t\bar{t}$ + single top production rate systematics estimated to be:

$$t\bar{t} + \text{single top systematics} = |1 - t\bar{t} + \text{single top}\overline{\frac{}{Data-otherbackgrounds=5\%}}| \quad ,(3.9)$$ in agreement with the estimate from .

3.8.2.3 Di-boson process

We take di-boson processes directly from simulation with a 15% production uncertainty.

3.8.3 QCD multi-jet background

For a given variable and binning, e.g. the effective mass variable used for signal extraction, we construct a data-driven template for the shape of the QCD multi-jet background, i.e. the processes lacking prompt leptons. Based on the charge of the final state and τ_h isolation, we split the events into four regions shown in Figure ?? and described below:

- A (Signal) Region: e and τ_h have opposite charge and τ_h pass "Tight" isolation requirement.

- B Region: e and τ_h have same charge and τ_h pass "Tight" isolation requirement.

- C Region: e and τ_h have opposite charge and τ_h pass anti-isolation requirement.

- D Region: e and τ_h have same charge and τ_h pass anti-isolation requirement.

In each region (B, C, D), QCD events are estimated by subtracting events with genuine leptons (estimated by simulation) bin-by-bin from data. QCD events are assumed to be charge blind, thus, the amount of QCD events in region B should be comparable to that of in the signal region. However, with the freedom to define the anti-isolation region, we choose an anti-isolation definition such that there are much more QCD in region C compared to the signal region. Taking the QCD shape from region C will provide us a much smoother template for QCD estimation.

Hence, QCD events in the signal region are estimated with the shape from region C and multiplying a scale factor derived from regions B and D. The factor is defined as:

$$f_{\text{LT}}^{\text{QCD}} = \left(N_{\text{data}}^{\text{B}} - N_{\text{MC}}^{\text{B}}\right) / \left(N_{\text{data}}^{\text{D}} - N_{\text{MC}}^{\text{D}}\right) \quad . \tag{3.10}$$

This QCD estimation method is valid only if the QCD shape in anti-isolated correctly models the QCD shape in the isolated region. The check is done by comparing the observation and background estimation in region B with the QCD shape taking from region D and normalized to the QCD in region B. An example of this test is shown in the right panel of Figure **??** with the anti-isolation definition as: $\tau_h$isolation failing the "Tight" working point but below 5.0.

With the freedom to define the $\tau_h$anti-isolation region, before the signal region selections so there is enough statistics, we scan through different anti-isolation definitions as shown in Figure 3.41 where the x-axis labels the start of the anti-isolation region and the y-axis labels the end. With each anti-isolation definition, we perform the check mentioned above and calculate χ^2 between observation and background estimation. The p-value of the χ^2 tests for each anti-isolation definition is shown on the cells in the right panel of Figure 3.41. The left panel of Figure 3.41 shows the same test but with $f_{\text{LT}}^{\text{QCD}}$ instead in each cell.

Based on the following criteria:

- good results from χ^2 test

- low "Loose-to-Tight" scale factor ($f_{\text{LT}}^{\text{QCD}}$) to ensure a smooth template

The range of isolation failing the tight working point, but below 5.0, was chosen as the sideband with. After the signal region selection the "Loose-to-Tight" scale factor is estimated to be: 0.13 ± 0.02 where an additional 15% uncertainty is added to the QCD systematics on top of the bin-by-bin systematics.

3.8.4   W+jets background

The simulated W+jets samples, especially at low , were not generated with large MC statistics. Using the samples directly would have two disadvantages: (a) avoiding non-smooth templates

in the signal extraction would restrict somewhat the choice of signal selection and binning; (b) the poorly sampled shapes degrade the ability to validate the background estimates. As a workaround, we use a region of relaxed $\tau_h$isolation to obtain the simulated shape: $\tau_h$isolation is required to fail the "tight" working point but be less than 5. This shape template is then normalized to the integrated MC yield when requiring tight $\tau_h$isolation.

Figure 3.43 compares the direct and relaxed predictions in the signal region and in the control region discussed below. It also shows the compatibility of the predicted shapes when the isolation sideband is split into to regions: $\tau_h$isolation failing "Tight" working point but < 2.5; $2.5 < \tau_h$isolation < 5.

To evaluate the estimated background rate in a signal-depleted and W-enriched region, we make the same requirements as for the signal region, except require $p_\zeta - 3.1 p_{\zeta\mathrm{vis.}} < -50$ (i.e., inverted w.r.t. the signal selection), and allowing any value of $\cos \Delta \phi_{e\tau_h}$ (i.e., relaxed w.r.t. the signal selection). We then scan the W+jets event rate from 1.0 to 1.25, in steps of 0.05. For each assumed rate, we redetermine the QCD prediction in this control region. Figure **??** shows the agreement between observations and background estimates at each scan point. A scale factor of 1.15 ± 0.10 gives the best agreement in this control region. The distribution of in the W-enriched control region, requiring also $\cos \Delta \phi_{e\tau_h} < -0.8$ (to be somewhat closer to the signal region), is shown in Figure 3.45.

Figure 3.45:
described in

Table 3.4: MC Samples

Process	cross-section (pb)	Official CMS Datasets (MINIAODSIM)
$Z \to ll$	6025.2	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
HT binned	147.4*1.23	/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1/
LO samples	40.99*1.23	/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	5.678*1.23	/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v2
	2.198*1.23	/DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z \to ll$	6025.2	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
mass binned	7.67*0.987	/DYJetsToLL_M-200to400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
NLO samples	0.423*0.987	/DYJetsToLL_M-400to500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	0.24*0.987	/DYJetsToLL_M-500to700_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v3
	0.035*0.987	/DYJetsToLL_M-700to800_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	0.03*0.987	/DYJetsToLL_M-800to1000_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	0.016*0.987	/DYJetsToLL_M-1000to1500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-Asympt25ns_74X_mcRun2_asymptotic_v2-v1
$W + jets$	61526.7	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
HT binned	1345*1.21	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
LO samples	359.7*1.21	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	48.91*1.21	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	18.77*1.21	/WJetsToLNu_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$t\bar{t}$	831.76	/TT_TuneCUETP8M1_13TeV-powheg-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2_ext3-v1
single Top samples	35.6	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	35.6	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v2
	136.02*0.108*3	/ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
	80.95*0.108*3	/ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
VV	11.95	/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$ZZ \to 2l2q$	3.22	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$ZZ \to 4l$	1.212	/ZZTo4L_13TeV_amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$WW \to l\nu 2q$	49.997	/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$WZ \to 2l2q$	5.595	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$WZ + jets$	5.26	/WZJets_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$WZ \to l3\nu$	3.05	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$WZ \to l\nu 2q$	10.71	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
QCD samples	720648000	/QCD_Pt20toInf_MuEnrichedPt15_TuneCUETP8M1_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(500)$	9.33	/ZprimeToTauTau_M_500_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(1000)$	0.468	/ZprimeToTauTau_M_1000_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(1500)$	0.0723	/ZprimeToTauTau_M_1500_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(2000)$	0.0173	/ZprimeToTauTau_M_2000_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(2500)$	0.00554	/ZprimeToTauTau_M_2500_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1
$Z'(3000)$	0.00129	/ZprimeToTauTau_M_3000_TuneCUETP8M1_tauola_13TeV_pythia8/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1

Table 3.5: Loose Jet-ID Selections.

Selection	Cut
Neutral Hadron Fraction	< 0.99
Neutral EM Fraction	< 0.99
Number of Constituents	> 1
And for $\eta < 2.4$, $\eta > -2.4$ in addition apply	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Table 3.6: Electron ID Selections.

Category	$\text{MVA}_{\text{min cut}}$ (80% signal eff)	$\text{MVA}_{\text{min cut}}$ (90% signal eff)
Barrel ($\eta < 0.8$) p_T 5–10	0.287435	-0.083313
Barrel ($\eta > 0.8$) p_T 5–10	0.221846	-0.235222
Endcap p_T 5–10	-0.303263	-0.67099
Barrel ($\eta < 0.8$) $p_T > 10$	0.967083	0.913286
Barrel ($\eta > 0.8$) $p_T > 10$	0.929117	0.805013
Endcap $p_T > 10$	0.726311	0.358969

Table 3.7: μ Identification

Cut
muon::isLooseMuon(recoMu)
recoMu.innerTrack()-> validFraction()> 0.8
AND
recoMu.globalTrack()-> normalizedChi2()< 3
recoMu.combinedQuality().chi2LocalPosition< 12
recoMu.combinedQuality().trkKink< 20
muon::segmentCompatibility(recoMu)> 0.303
OR
muon::segmentCompatibility(recoMu)> 0.451
RelIso < 0.15

Table 3.8: Reconstructed Tau Decay Modes

HPS Tau Decay Modes
Single Charged Hadron + Zero Strip
Single Charged Hadron + One Strip
Single Charged Hadron + Two Strips
Two Charged Hadrons
Three Hadrons

Table 3.9: Electron ID Selections.

Category	$MVA_{min\ cut}$ (80% signal eff)	$MVA_{min\ cut}$ (90% signal eff)
Barrel ($\eta < 0.8$) p_T 5–10	0.287435	-0.083313
Barrel ($\eta > 0.8$) p_T 5–10	0.221846	-0.235222
Endcap p_T 5–10	-0.303263	-0.67099
Barrel ($\eta < 0.8$) $p_T > 10$	0.967083	0.913286
Barrel ($\eta > 0.8$) $p_T > 10$	0.929117	0.805013
Endcap $p_T > 10$	0.726311	0.358969

Table 3.10: μ Identification

Cut
muon::isLooseMuon(recoMu)
recoMu.innerTrack()-> validFraction()> 0.8
AND
recoMu.globalTrack()-> normalizedChi2()< 3
recoMu.combinedQuality().chi2LocalPosition< 12
recoMu.combinedQuality().trkKink< 20
muon::segmentCompatibility(recoMu)> 0.303
OR
muon::segmentCompatibility(recoMu)> 0.451
RelIso < 0.15

Table 3.11: Reconstructed Tau Decay Modes

HPS Tau Decay Modes
Single Charged Hadron + Zero Strip
Single Charged Hadron + One Strip
Single Charged Hadron + Two Strips
Two Charged Hadrons
Three Hadrons

Table 3.12: Loose Jet-ID Selections.

Selection	Cut
Neutral Hadron Fraction	< 0.99
Neutral EM Fraction	< 0.99
Number of Constituents	> 1
And for $\eta < 2.4$, $\eta > -2.4$ in addition apply	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Process	Trigger: HLT_IsoMu18	1μ	1τ_h	b-jet Veto	> 30 GeV	μτ_h topology cuts
Data	47386261	15478176	61718	53254	28131	1807
WW	24673.4 ± 77.5	20329.8 ± 70.4	431.9 ± 10.3	371.9 ± 9.5	257.1 ± 7.9	24.1 ± 2.4
WZ	7650.2 ± 27.9	6515.0 ± 25.7	123.6 ± 3.5	100.1 ± 3.2	64.4 ± 2.6	4.5 ± 0.7
ZZ	1825.7 ± 8.0	1642.5 ± 7.6	39.2 ± 1.2	31.6 ± 1.0	14.0 ± 0.7	0.7 ± 0.2
$t\bar{t}$	231246.4 ± 189.0	193282.7 ± 172.8	4369.2 ± 26.0	606.8 ± 9.7	510.8 ± 8.9	26.1 ± 2.0
W+Jets	16090363.1 ± 5271.1	11990572.1 ± 4538.0	30234.3 ± 215.4	28230.5 ± 210.9	18078.6 ± 167.7	757.9 ± 35.0
DY+Jets	2668493.1 ± 2584.4	2140958.6 ± 2086.4	12872.6 ± 180.7	11907.2 ± 175.1	3611.3 ± 98.5	882.9 ± 34.9
QCD	6296741.5 ± 17403.0	1058693.6 ± 7136.0	3992.2 ± 438.3	3030.2 ± 381.0	865.8 ± 204.0	143.1 ± 89.8
Total BG	25320993.1 ± 18367.6	15411993.1 ± 8712.3	52063.0 ± 521.5	44278.3 ± 469.6	23402.0 ± 282.1	1839.12 ± 102.60
Z'→ ττ(500)	3521.2 ± 21.5	3219.6 ± 20.5	960.8 ± 11.2	868.7 ± 10.7	715.9 ± 9.7	502.3 ± 8.1
Z'→ ττ(1000)	213.8 ± 1.2	201.9 ± 1.1	73.3 ± 0.7	64.9 ± 0.7	59.4 ± 0.6	40.8 ± 0.5
Z'→ ττ(1500)	36.0 ± 0.1	34.2 ± 0.1	13.2 ± 0.1	11.6 ± 0.1	10.9 ± 0.1	7.2 ± 0.07
Z'→ ττ(2000)	8.7 ± 0.05	8.2 ± 0.04	3.3 ± 0.03	2.9 ± 0.03	2.8 ± 0.03	1.8 ± 0.02
Z'→ ττ(2500)	2.8 ± 0.01	2.6 ± 0.01	1.1 ± 0.01	0.9 ± 0.008	0.9 ± 0.008	0.6 ± 0.006
Z'→ ττ(3000)	0.6 ± 0.003	0.6 ± 0.003	0.3 ± 0.002	0.2 ± 0.002	0.2 ± 0.002	0.1 ± 0.001

Table 3.13: Signal and background yields after various stages of the $\mu\tau_h$ selection.

Process	$\mu\tau_h$ W + jets CR B	$\mu\tau_h$ W + jets CR A	$\mu\tau_h$ W + jets CR C
ZZ + jets	8.52 ± 2.97	4.19 ± 2.08	0.71 ± 0.86
WZ + jets	24.42 ± 5.19	18.24 ± 4.48	1.85 ± 1.43
WW + jets	141.18 ± 13.25	86.18 ± 10.35	17.15 ± 4.62
QCD	188.50 ± 96.15	631.34 ± 176.91	140.46 ± 83.44
$t\bar{t}$	249.55 ± 16.97	189.68 ± 14.80	22.80 ± 5.13
Z + jets	1278.18 ± 71.00	871.00 ± 46.45	425.72 ± 40.21
W + jets	5235.70 ± 115.55	7591.91 ± 139.17	1112.06 ± 53.65
Data	8278	10434	1847
Purity	73.47%	80.83%	64.63%
Data - $\sum_{i \neq W} BG_i$	6387.64 ± 201.46	8633.38 ± 237.17	1238.31 ± 111.78
SF	1.22 ± 0.05	1.14 ± 0.04	1.11 ± 0.11

Table 3.14: Background and data yields in W + jets control regions A, B, C, under nominal conditions (i.e. > 30 GeV).

Process	$\mu\tau_h$ W + jets CR B	$\mu\tau_h$ W + jets CR A	$\mu\tau_h$ W + jets CR C
ZZ + jets	15.99 ± 4.07	9.60 ± 3.15	1.77 ± 1.35
WZ + jets	37.44 ± 6.42	26.20 ± 5.37	3.99 ± 2.10
WW + jets	183.63 ± 15.11	114.33 ± 11.92	29.02 ± 6.01
$t\bar{t}$	284.32 ± 18.12	222.57 ± 16.03	29.70 ± 5.86
QCD	411.63 ± 141.90	1002.89 ± 221.70	442.10 ± 147.19
Z + jets	3282.89 ± 101.15	2754.53 ± 84.50	1870.10 ± 83.63
W + jets	7586.89 ± 139.27	10894.85 ± 166.72	2699.93 ± 84.62
Data	13896	17235	5984
Purity	64.28%	72.51%	53.18%
Data - $\sum_{i \neq W} BG_i$	9680.10 ± 394.34	13104.88 ± 400.66	3607.34 ± 267.70
SF	1.28 ± 0.06	1.20 ± 0.04	1.34 ± 0.11

Table 3.15: Background and data yields in W + jets control regions A, B, C obtained without a requirement.

Table 3.16: Events in the W + jets high $m_T(\mu,)$ control region for data and MC.

Data	731
W + jets	580.163 ± 39.4722
Z + jets	135.905 ± 17.0143
WZ + jets	2.88218 ± 1.7809
WW + jets	15.2605 ± 4.35526
ZZ + jets	0.217323 ± 0.474434
QCD	36.0738 ± 39.3694
$t\bar{t}$	17.7127 ± 4.52124
Purity	73.6047%
Data - $\sum_i BG_i$	522.948 ± 51.1199
SF	0.901382 ± 0.107354

Table 3.17: Events in the $Z \to \tau\tau$ control region for data and MC.

Data	6658
W + jets	2586.87 ± 168.197
Z + jets	3806.14 ± 100.685
WZ + jets	8.95952 ± 3.14143
WW + jets	48.7744 ± 7.78855
ZZ + jets	2.82066 ± 1.70842
QCD	319.837 ± 125.08
$t\bar{t}$	32.9836 ± 6.1709
Purity	55.9202%
Data - $\sum_i BG_i$	3657.76 ± 225.177
SF	0.961015 ± 0.0643923

Table 3.18:   Yields in the $\mu\tau_h$ QCD control region for data and MC.

Process	non-isolated τ_h, non-isolated μ, same-sign $\mu\tau_h$
Data	442
W + jets	33.1253 ± 9.20042
Z + jets	1.41175 ± 1.258
WZ	0.340759 ± 0.612045
WW	0 ± 0
ZZ	$0.000177035 \pm 0.0135337$
QCD	371.641 ± 90.8108
$t\bar{t}$	1.67439 ± 1.39031
Purity	91.0453%
Data - $\sum_i BG_i$	405.448 ± 23.0334
SF	1.09097 ± 0.273689

Table 3.19:   Yields in the controls region 1(B) and 1(D) used for calculation of OS/LS ratio.

Process	OS $\tau_h\tau_h$ isolated + MET 30 GeV	SS $\tau_h\tau_h$ isolated + MET 30 GeV
Data	207	113
W + jets	0.897455 ± 1.57954	1.20229 ± 1.26803
Z + jets	21.8385 ± 4.67317	0.132112 ± 0.363472
WZ + jets	0 ± 0	0 ± 0
WW + jets	0.522162 ± 0.800846	0 ± 0
ZZ + jets	0.0522376 ± 0.232475	0 ± 0
$t\bar{t}$	0.0111017 ± 0.112979	0 ± 0
Data - $\sum_i BG_i$	183.679 ± 15.2329	111.666 ± 10.7117
OS/LS ratio	1.64 ± 0.208	

Process	OS $\tau_h\tau_h$ isolated + MET 30 GeV	SS $\tau_h\tau_h$ isolated + MET 30 GeV
Data	55	30
W + jets	0.0935669 ± 0.352639	0.281734 ± 0.612064
Z + jets	8.117 ± 2.84904	0.101807 ± 0.319072
WZ + jets	0 ± 0	0 ± 0
WW + jets	0.487436 ± 0.768961	0 ± 0
ZZ + jets	0 ± 0	0 ± 0
$t\bar{t}$	0 ± 0	0 ± 0
Data - $\sum_i BG_i$	46.302 ± 7.98953	29.6165 ± 5.52055

Table 3.20: Background and data yields in QCD control regions A and C under nominal isolation and conditions (i.e. isolated + > 30 GeV).

Region	OS $\tau_h\tau_h$ + MET 30 GeV	SS $\tau_h\tau_h$ + MET 30 GeV	SS $\tau_h\tau_h$ + MET 30 GeV	OS $\tau_h\tau_h$ + MET 30 GeV
isolated	183.679 ± 15.4261	111.666 ± 10.7129	29.6165 ± 5.52111	48.7161 ± 11.0027

Table 3.21: QCD yields in the isolated regions A (signal region), B, C, and D.

3.8.5   Overlays of observations and SM predictions

The expected SM event yields in the signal region are shown in Figure 3.46.

Distributions of and η are shown in Figure 3.47.

3.8.6   Post-unblinding: further checks of the backgrounds

3.8.6.1   Checks of MC-based W+Jets

Based on m_T distribution shown in Figure 3.46, we require $60\ GeV < m_T < 120\ GeV$ to compare data v.s estimated background in a W+Jets rich region. Figure 3.48 shows that the W+Jets shape estimated from the relaxed $\tau_h$isolation region provides a smoother template and agrees better with observations.

For the channel, 2-prong τ_h's were rejected due low signal presence and high W+Jets and QCD contamination. Hence, 2-prong τ_h's provide a testing group for the background estimation methods. As shown in Figure 3.49 there is a good agreement between observations and estimated background.

3.8.6.2   Data-driven W+jets checks

Similar to the channel, by requesting e and $\tau_h$to have opposite charge, we have the following four regions:

- A (Signal) Region: pass "ζ" and "$\cos\Delta\phi$" and $\tau_h$pass "Tight" isolation requirement.

- B Region: fail "ζ" or "$\cos\Delta\phi$" and $\tau_h$pass "Tight" isolation requirement.

- C Region: pass "ζ" and "$\cos\Delta\phi$" and $\tau_h$pass anti-isolation requirement.

- D Region: fail "ζ" or "$\cos\Delta\phi$" and $\tau_h$pass anti-isolation requirement.

where the W+jets in the signal region is estimated from region C by subtracting all other background from data and multiplying a scale factor estimated among region B and D.

$$N_{W+jets}^{signal} = N_{data-otherbkg}^{C} \times \frac{N_{data-otherbkgs}^{B}}{N_{data-otherbkg}^{D}} \tag{3.11}$$
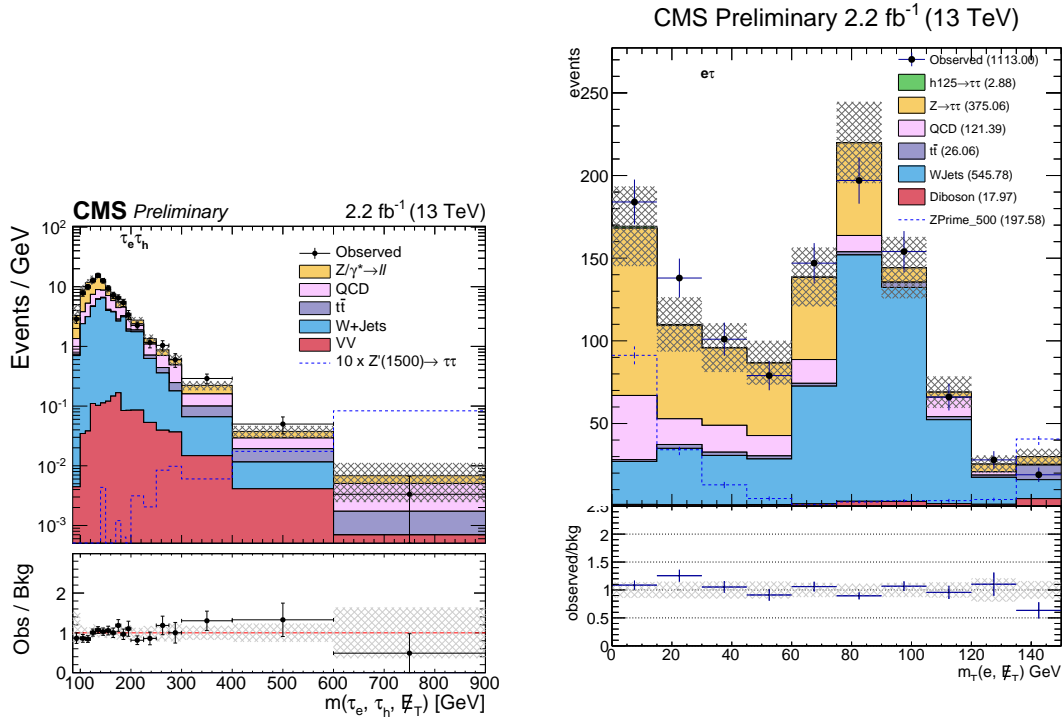
Figure 3.46: Left: predicted background yields and observed event yields in the channel after signal selection. Right: the distribution of transverse mass.
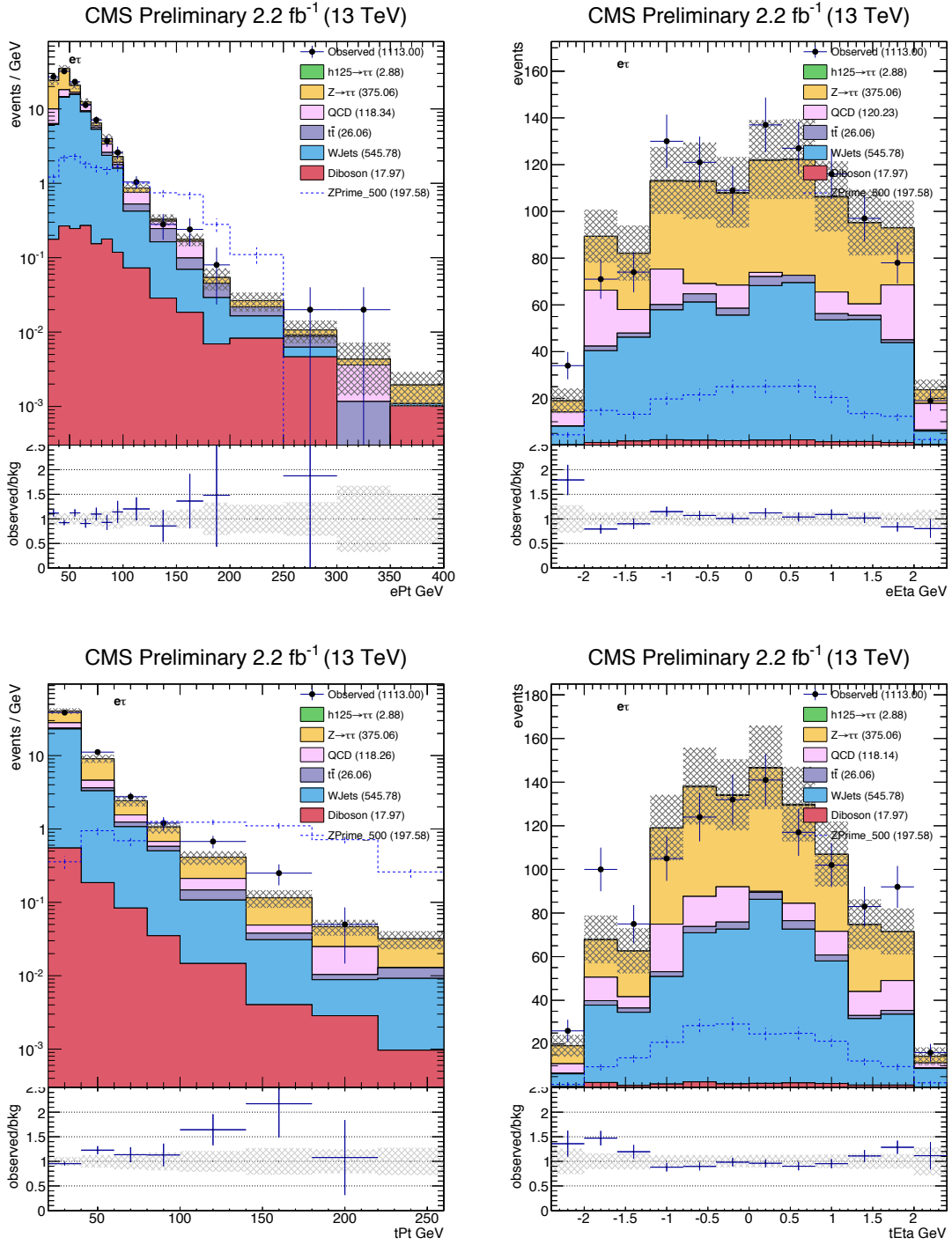
Figure 3.47: Distributions, after final selection, of electron (top left), electron pseudo-rapidity (top right), τ (bottom left), τ pseudo-rapidity (bottom right).
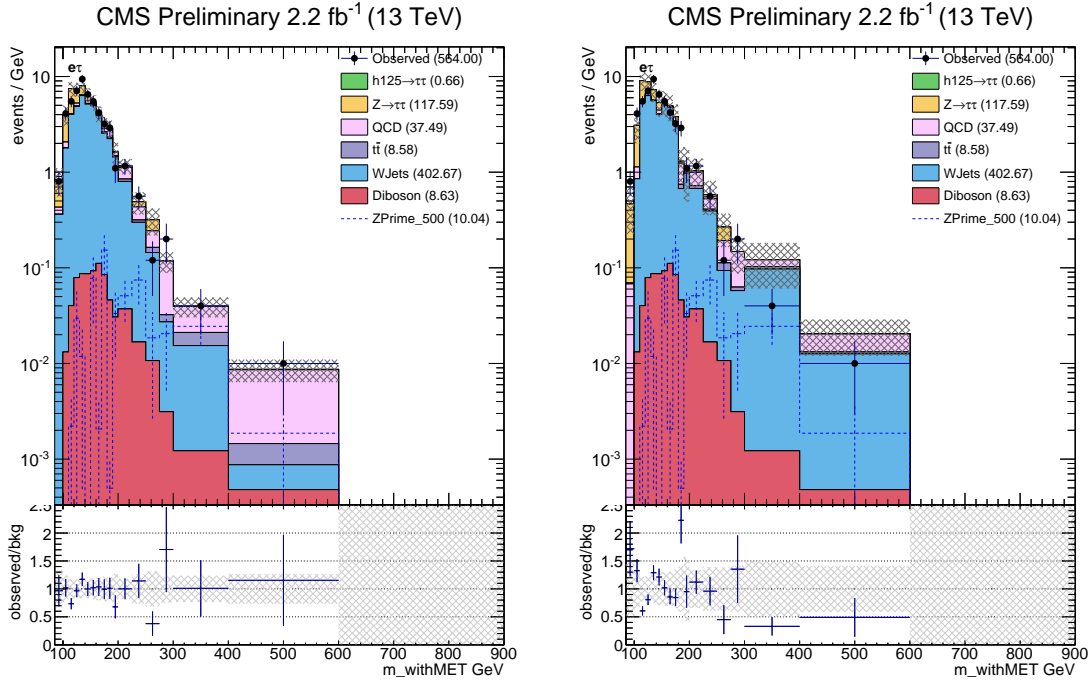
Figure 3.48:   channel, in the signal region, requiring also 60 $GeV < m_T < 120$ GeV. Left: The distributions of with W+Jets shape estimated from relaxed τ_h isolation region. Right: The distributions of with W+Jets shape estimated from tight τ_h isolation region.

Figure 3.49: In the signal region, but requiring 2-prong τ_h's: the distribution of reconstructed parent mass, , with W+Jets shape estimated from relaxed $\tau_h$isolation region.

QCD in each region is estimated from data by inverting the charge requirement. Thus, we have four more regions A', B', C' and D' with similar requirements as ABCD but requesting e and τ_h to have the same charge. Hence, QCD in B, C and D regions are estimated in the following way:

- QCD in B Region: shape taken from data - all MC bkg (including W+jets MC) in D' region. Yield normalized to data - all MC bkg (including W+jets MC) in D region with a "SStoOS" scale factor of 1.07 ± 0.05 estimated from the channel.
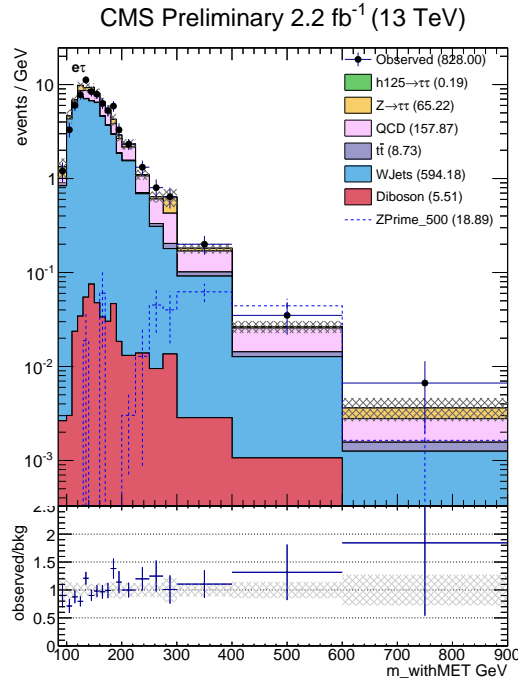
- QCD in C Region: shape taken from data - all MC bkg (including W+jets MC) in C' region. Yield normalized to data - all MC bkg (including W+jets MC) in C region

- QCD in D Region: shape taken from data - all MC bkg (including W+jets MC) in D' region. Yield normalized to data - all MC bkg (including W+jets MC) in D region.

The right panel of Figure 3.50 shows the observation and background estimation comparisons with the data-driven W+jets and QCD estimations mentioned above. Compared with the MC W+jets estimation on the left panel, we see negligible difference between the two.

3.9 Electron + Muon Channel

3.9.1 Event selection

The electron selection is identical to that described in Section 3.8.1. Muons are required to have:

- $p_T > 10$ GeV and $|\eta| < 2.1$

- distance of closest approach to the leading sum-p_T^2 primary vertex of less than 0.045 cm (transvese) and 0.2 cm (longitudinal)

- satisfy the muon POG medium muon requirement

We build pairs of electrons and muons in which the electron and muon are separated by at least $\Delta R > 0.3$. In events with more than one such pair, we select the pair with the two most
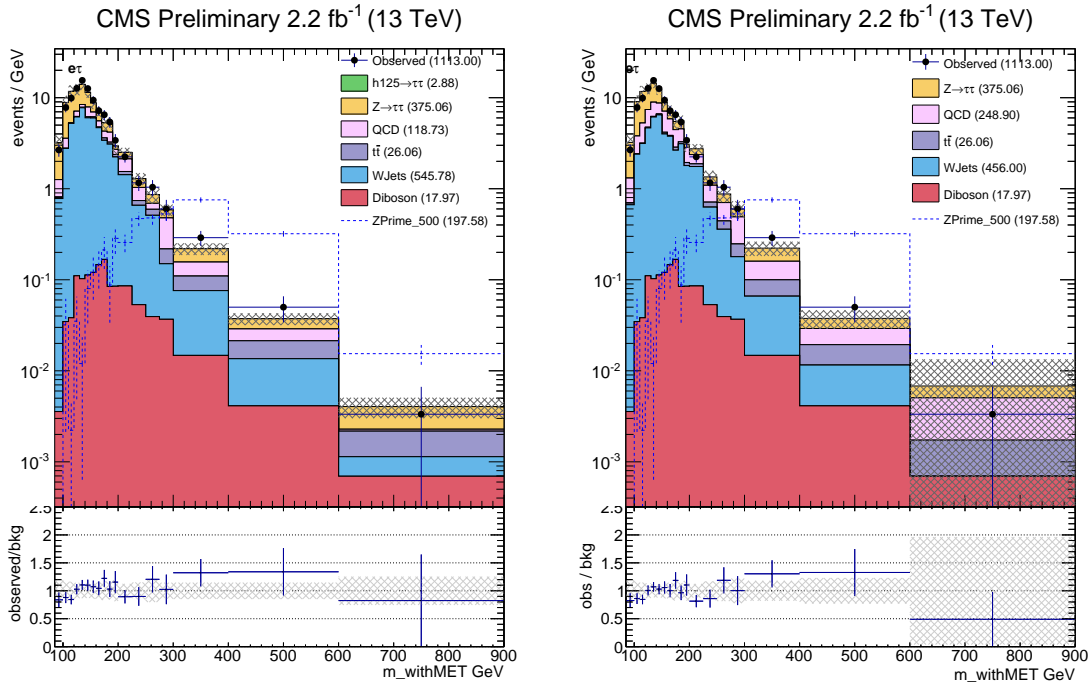
Figure 3.50: Left: W+jets shape and yield from simulation Right: Data-driven W+jets estimation channel:

isolated leptons, considering first the muon, and then the electron. This criterion was seen to have good efficiency for signal samples. In the rare case of multiple such pairs having identical isolation values, the reconstructed p_T's are considred, preferring higher values.

After a pair has been chosen for an event, we require both the electron and muon relative isolations to be < 0.15, for an event to enter the signal region. To reduce a possible Drell-Yan background, events are rejected if there is an additional electron satisfying the requirements described in Section 3.8.1 and with relative isolation < 0.3, or an additional muon satisfying the above identification requirements with relative isolation < 0.3.

As for the other channels, the signal region is defined as having

- $\cos \Delta\phi(e, \mu) < -0.95$

- $Q(e) \times Q(\mu) < 0$

- > 30

- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$

- no jet with $p_T > 30$ tagged as a b-jet (CSV loose)

The distributions of these variables after preselection are shown in Figures 3.51, 3.52, 3.53.

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Figure 3.54.

3.9.2   Genuine dilepton events

Studies of simulated events indicate that for Drell-Yan process, top quark single and pair production, and di-boson production, the reconstructed and selected muons and electrons are typically associated with genuine simulated leptons. The nominal expected event rates are estimated by scaling the simulated samples by the best available cross sections, listed in Table 3.4, and by the integrated luminosity of the data samples.

3.9.2.1   Drell-Yan process

Systematics for Drell-Yan process is estimated in an Drell-Yan rich region with the following selections and shown in the left panel of Figure 3.55:

- $Q(e) \times Q(\mu) < 0$

- < 30

- no jet with $p_T > 30$ tagged as a b-jet (CSV loose)

- < 125 GeV

The Drell-Yan production rate systematic uncertainty is estimated to be:

$$Drell - Yan \, systematics = \left| 1 - \frac{Drell - Yan}{Data - other backgrounds} \right| = 12\% \tag{3.12}$$

which we apply both to and final states.

3.9.2.2 $t\bar{t}$ and single top processes

Systematics for $t\bar{t}$ and single top processes are estimated in a top quark rich region with the following selections and shown in the right panel of Figure 3.55:

- $Q(e) \times Q(\mu) < 0$

- > 30

- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$

- at least one jet with $p_T > 30$ tagged as a b-jet (CSV loose)

The $t\bar{t}$ + single top production rate systematics estimated to be:

$t\bar{t}$ + single top systematics $= |1 - t\bar{t}$ + single top $\frac{}{Data - other backgrounds = 8\%}$ (3.13) which we apply both to and final states.

3.9.2.3   Di-boson process

We take di-boson processes directly from simulation with a 15% production uncertainty.

3.9.3   QCD multi-jet background

The estimation of the QCD background for the channel is directly analogous to that in the channel, except that the sideband is defined by the muon isolation. Figure 3.56 shows the results of the scanning for a sideband. The range of relative isolation from 0.15 to 0.95 was chosen as the sideband. After the signal region selection the "Loose-to-Tight" scale factor is estimated to be: 0.20 ± 0.08 where this 40% rate uncertainty is applied to the QCD process (in addition to the bin-by-bin systematic uncertainties).

3.9.4   W+jets background

The W background is very small. However, as in the channel, the W+jets simulated sample was not generated with high statistics. As a workaround, the W+jets shape is taken from the simulated sample in the muon isolation sideband, and scaled to match the simulated yield in the tight muon isolation. The "loose-to-tight" factor is 0.073 ± 0.03.

3.9.5   Overlays of observations and SM predictions

The expected SM event yields in the signal region are shown in Figure 3.58.

Distributions of and η are shown in Figure 3.59.

3.10   Systematics

The following systematics have been considered (summarized in Table 3.22):

- **Luminosity:** We consider a 5% uncertainty on the measured luminosity [?]. It is considered 100% correlated across MC based backgrounds within a channel. It is also considered 100% correlated across channels (for MC based backgrounds).

- **Trigger, Reconstruction, and Selection:** An overall uncertainty is applied for the trigger uncertainties determined on the correction factors described in Section 3 and which are measured using tag-and-probe methods. The uncertainty for muons and electrons is estimated to be 1% each independent of η or p_T, while we consider 5% uncertainty per hadronic tau leg in the $\tau_h\tau_h$ channel (assuming each leg is 100% correlated). The trigger uncertainty is considered 100% correlated across MC based backgrounds within a channel. It is also considered 100% correlated across channels using the same trigger. For the case of the $\tau_h\tau_h$ trigger, where the trigger efficiency uncertainty is measured per τ_h leg, the total trigger uncertainty is calculated by assuming both legs are 100% correlated. For example, if the per leg τ_h trigger uncertainty is 5%, the total trigger uncertainty for the $\tau_h\tau_h$ channel will be 10% $(2 \cdot 5\%)$.

- **b-Tagging Efficiency:** We consider a 30% uncertainty on the mis-tag rate as measured by the b-tagging POG [?]. For the case of our signal, the systematic uncertainty on the requirement of 0 jets mis-tagged as b-jets is determined by propagating the 30% uncertainty on the mis-tag rate through the following equation (which represents the signal efficiency for requiring 0 jets mis-tagged as b-Jets):

$$\epsilon^{\mathrm{NBtag}<1} = 1 - \sum_{n=1} P(n) \cdot \sum_{m=1}^{n} C(n,m) \cdot f^m \cdot (1-f)^{n-m} \qquad (3.14)$$

where $P(n)$ is the probability to obtain n additional jets (non-tau and non-lepton) in the event, $C(n,m)$ the combinatorial of n *choose* m, and f the mis-tag rate. The probability to obtain at least one additional jet in the event is $\sim 10\%$. Therefore, based on the above equation, the mis-tag rate and uncertainty, and the probability to obtain at least one additional jet we calculate a systematic effect of $\sim 5\%$ on our signal due to the mis-tag rate. The b-tagging/mis-tagging systematics are considered 100% correlated across MC based backgrounds with similar composition (e.g. W + jets and DY + jets where there are typically no real b-jets), but completely uncorrelated to backgrounds that have different

composition (e.g. $t\bar{t}$ vs. DY + jets).

- **Electron Energy Scale:**We consider the effect on the signal acceptance efficiency of a 1% (2.5%) shift on the electron energy scale in the barrel (endcap) region. The resultant systematic uncertainty on signal and MC based backgrounds is < 1%.

- **Muon Momentum Scale:** We consider the effect on the signal acceptance efficiency of a 1% momentum scale uncertainty on the muon momentum. The resultant systematic uncertainty on signal and MC based backgrounds is < 1%.

- **Tau Energy Scale:** We consider the effect of the 5% tau energy scale uncertainty measured by the tau POG on the signal acceptance. The tau 4-momentum is scaled by a factor of $k = 1.05$ ($p_{smeared} = k \cdot p_{default}$) and variables are recalculated using $p_{smeared}$. We find that by using $p_{smeared}$ calculated with a factor of $k = \pm 1.05$, the signal and MC baded backgrounds fluctuates by up to $\sim 11\%$.

- **Jet Energy Scale:** We consider the effect of a 3-5% jet energy scale uncertainty on the signal acceptance (depending on the η and p_T of the considered jet as prescribed by the *JetMET POG*). The jet 4-momentum is scaled by a factor of $k = 1.05$ ($p_{smeared} = k \cdot p_{default}$) and variables are recalculated using $p_{smeared}$. We find that by using $p_{smeared}$ calculated with a factor of $k = \pm 1.05$, the signal and MC baded backgrounds fluctuates by up to $\sim 12\%$.

- **PDF Systematics Uncertainty:** We consider the effect of the PDF uncertainties on the signal acceptance. Due to the lack of PDF weights in the Z' signal samples, we sliced up mass binned DY samples to mock up Z' samples in the following way:

Z' mass point (GeV)	500	1000	1500	2000	2500	3000
DY mass slice	500to550	1000to1050	1500to1550	2000to2050	2450to2550	2800to3000

Following the "PDF4LHC recommendations for LHC Run II" [?], the PDF uncertainties are computed from the 68% confidence level with the PDF4LHC15_mc sets. The PDF

uncertainties for our main backgrounds, $t + \bar{t}$, W+Jets and DY, are much smaller than their bin-by-bin statistical uncertainties thus is neglected. The PDF uncertainties on the signal acceptance range from 0.7% for Z' at 500 GeV up to 12% for Z' at 3 TeV.

- **Background Estimates:** The uncertainty on the data-driven background estimations are driven by the statistics in data in the various control samples. There is also a mostly negligible contribution from the level of contamination from other BGs in the control regions. In cases where MC based BGs must be subtracted off, the uncertanties in the MC bgs due to the above listed systematic uncertanties are propagated throughout the subtraction and used to assign a systematic uncertainty on the background prediction.

Table 3.22: Summary of systematic uncertainties. Values are given in percent. "s" indicates template variations ("shape" uncertainties).

Source	QCD	W	DY	$t + \bar{t}$	VV	Signal
	$hh,\ \mu h,\ eh,\ e\mu$	$hh,\ \mu h,\ eh,\ e\mu$	$hh,\ \mu h,\ eh,\ e\mu$	$hh,\ \mu h,\ eh,\ e\mu$	$hh,\ \mu h,\ eh,\ e\mu$	$hh,\ \mu h,\ eh,\ e\mu$
Lumi	–,–,–,–	5,–,5,5	5,5,5,5	5,5,5,5	5,5,5,5	5,5,5,5
μ Trig	–,–,–,–	–,–,–,–	–,1,–,–	–,1,–,–	–,1,–,–	–,1,–,–
μ ID	–,–,–,–	–,–,–,1	–,1,–,1	–,1,–,1	–,1,–,1	–,1,–,1
e Trig	–,–,–,–	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1
e ID	–,–,–,–	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1
τ_h Trig	–,–,–,–	10,–,–,–	10,–,–,–	10,–,–,–	10,–,–,–	10,–,–,–
τ_h ID	–,–,–,–	30,–,6,–	12,6,6,–	12,6,6,–	12,6,6,–	12,6,6,–
b ID	–,–,s,s	10,–,s,s	3,3,s,s	10,12,s,s	3,3,s,s	3,3,s,s
JES	–,–,s,s	12,–,s,s	8,s,s,s	12,s,s,s	8,s,s,s	2,2,s,s
MMS	–,–,–,–	–,–,–,1	–,1,–,1	–,1,–,1	–,1,–,1	–,1,–,1
EES	–,–,–,–	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1	–,–,1,1
TES	–,–,s,s	11,–,s,s	11,s,s,s	11,s,s,s	8,s,s,s	3,3,s,s
Closure+Norm.	8,68,16,37	5,8,10,41	19,7,12,12	8,8,8,8	15,15,15,15	

3.11   Results and Conclusions

Figure 3.60 shows the background predictions as well as the observed $m(\tau_1, \tau_2,)$ spectrum, in log scale, for the four channels considered in this analysis note: $\mu\tau_h$(top left), $\tau_h\tau_h$(top right), $e\tau_h$(bottom left), $e\mu$(bottom right). Table 3.11 (top) lists the number of estimated background

events compared with the total number of observed events in data for each final state considering the whole mass spectrum, while Table 3.11 (bottom) those considering $M(\tau_1, \tau_2, \not{E}_T) > 300$. The observed $m(\tau_1, \tau_2,)$ spectrum in the signal region does not reveal any evidence for $Z' \to \tau\tau$ production (**this sentence and what follows is just placeholder text since we have not unblinded**).

Number of observed events in data and estimated background events, for the whole mass range (top) and in the region $M(\tau_1, \tau_2, \not{E}_T) > 300$ GeV (bottom). The uncertainties quoted on the number of background events represent the combined statistical and systematic uncertainty.

Process	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
Drell-Yan	8 ± 3	882 ± 127	375 ± 118	321 ± 99
W+jets	0.1 ± 0.1	916 ± 96	546 ± 86	19 ± 11
Diboson	0.5 ± 0.5	29 ± 7	18.0 ± 4	108 ± 17
$t\bar{t}$	$-$	26 ± 7	26 ± 8	223 ± 45
Multijet	49 ± 13	122 ± 84	117 ± 72	32 ± 24
Total	58 ± 13	1976 ± 180	1082 ± 162	703 ± 113
Observed	55	1807	1113	728

Process	$\tau_h\tau_h$	$\mu\tau_h$	$e\tau_h$	$e\mu$
Drell-Yan	5 ± 2	16 ± 4	9 ± 3	4 ± 2
W+jets	0.004 ± 0.004	23 ± 9	17 ± 7	1 ± 1
Diboson	0.02 ± 0.02	6 ± 3	2 ± 1	23 ± 4
$t\bar{t}$	$-$	4 ± 2	5 ± 1	65 ± 13
Multijet	18 ± 6	4 ± 3	6 ± 2	1 ± 1
Total	23 ± 6	51 ± 11	39 ± 8	94 ± 14
Observed	20	42	40	96

An upper bound at 95% confidence level (CL) is set on $\sigma \cdot BR$, where σ is the cross-section for pair production of $pp \to Z'$ and BR the branching fraction for $Z' \to \tau\tau$.

The calculation of the exclusion limit is obtained by using each bin of the $m(\tau_1, \tau_1,)$ distribution to construct one bin of the likelihood and computing the 95% confidence level (CL) upper limit on the signal cross-section using the asymptotic CL_s method. Said differently, a shape based analysis is performed, using the $m(\tau_1, \tau_1,)$ distribution as the fit discriminant to determine the likelihood of observing signal in the presence of the predicted background rate, given the observed yield in data. Systematic uncertainties are represented by nuisance parameters, which are profiled, assuming a log normal prior for normalization parameters, and Gaussian priors for mass-spectrum

shape uncertainties.

The above procedure is performed using the Higgs limit calculation tool "combine". The tool takes as input data cards with the yields and nuisance parameters in each $m(\tau_1, \tau_1,)$ bin of the search channels. These data cards are provided by the four channels being considered, and the only further input that is required is the correlations within and across channels. As only an example, if all the channels had 10 bins from $0 < m(\tau_1, \tau_1,) < 5000$ GeV (500 GeV/bin), this means there will be 10 cards per channel and therefore 40 cards in total for the four channels (if all the channels have the same bin size). The cards corresponding to a specific final state, e.g. $\mu\tau_h$, were then combined using the "CombineCards.py" tool provided by the Higgs limit tool, resulting in a single *combined* data card. This procedure was performed for all the final states considered in the analysis, resulting in X *combined* cards. The individual limits, per channel, were obtained by running the combine tool over each *combined* card separately. The final combined limit was obtained by combining the four resulting *combined* cards per channel described above, using the "CombineCards.py" tool. In order to handle correlations within and across channels, the following approach was used. Each nuisance parameter was defined with a convention of two indices: the first index, i, denoted the channel ($i = \mu\tau_h = 0$, $i = \tau_h\tau_h = 1$, $i = e\tau_h = 2$, $i = e\mu = 3$) and the second one the type of process ($j =$ Signal$= 0$, $j =$ W+jets$= 1$, etc.). Since the limit tool handles nuisance parameters with the same name as fully correlated, correlations across channels and processes were specified by utilizing the same i and process index j, respectively.

Figure 3.11 shows the expected limits as well as the theoretical cross-section as functions of $m(Z')$ mass for each channel. Figure 3.63 show the combined limit. A k factor of 1.3 has been used to scale the leading order (LO) signal cross-section. We exclude Z' (decaying through $Z' \to \tau\tau$) masses below approximately 2.0 TeV.

Table 3.23: Event summary table after signal region selection

Process				
Z' (500)	307.4 ± 35.3	502.3 ± 57.7	197.6 ± 22.7	218.6 ± 27.3
Z' (1000)	34.6 ± 2.6	40.8 ± 3.1	14.7 ± 1.1	19.0 ± 1.5
Z' (1500)	6.6 ± 0.3	7.2 ± 0.3	2.3 ± 0.1	3.6 ± 0.2
Z' (2000)	1.6 ± 0.07	1.8 ± 0.08	0.59 ± 0.03	0.91 ± 0.04
Z' (2500)	0.55 ± 0.02	0.60 ± 0.02	0.19 ± 0.01	0.30 ± 0.01
Z' (3000)	0.13 ± 0.01	0.14 ± 0.01	0.04 ± 0.00	0.07 ± 0.00
Drell-Yan	8.4 ± 3.1	882.4 ± 127.0	375.1 ± 117.6	321.2 ± 99.2
W+jets	0.1 ± 0.1	916.2 ± 96.1	545.8 ± 85.6	18.9 ± 11.4
Diboson	0.5 ± 0.5	29.2 ± 7.4	18.0 ± 4.4	108.3 ± 17.4
$t\bar{t}$	–	26.1 ± 6.7	26.1 ± 7.5	222.8 ± 44.8
Multijet	48.7 ± 13.0	121.8 ± 83.5	116.7 ± 71.5	31.9 ± 24.3
Observation	55	1807	1113	728
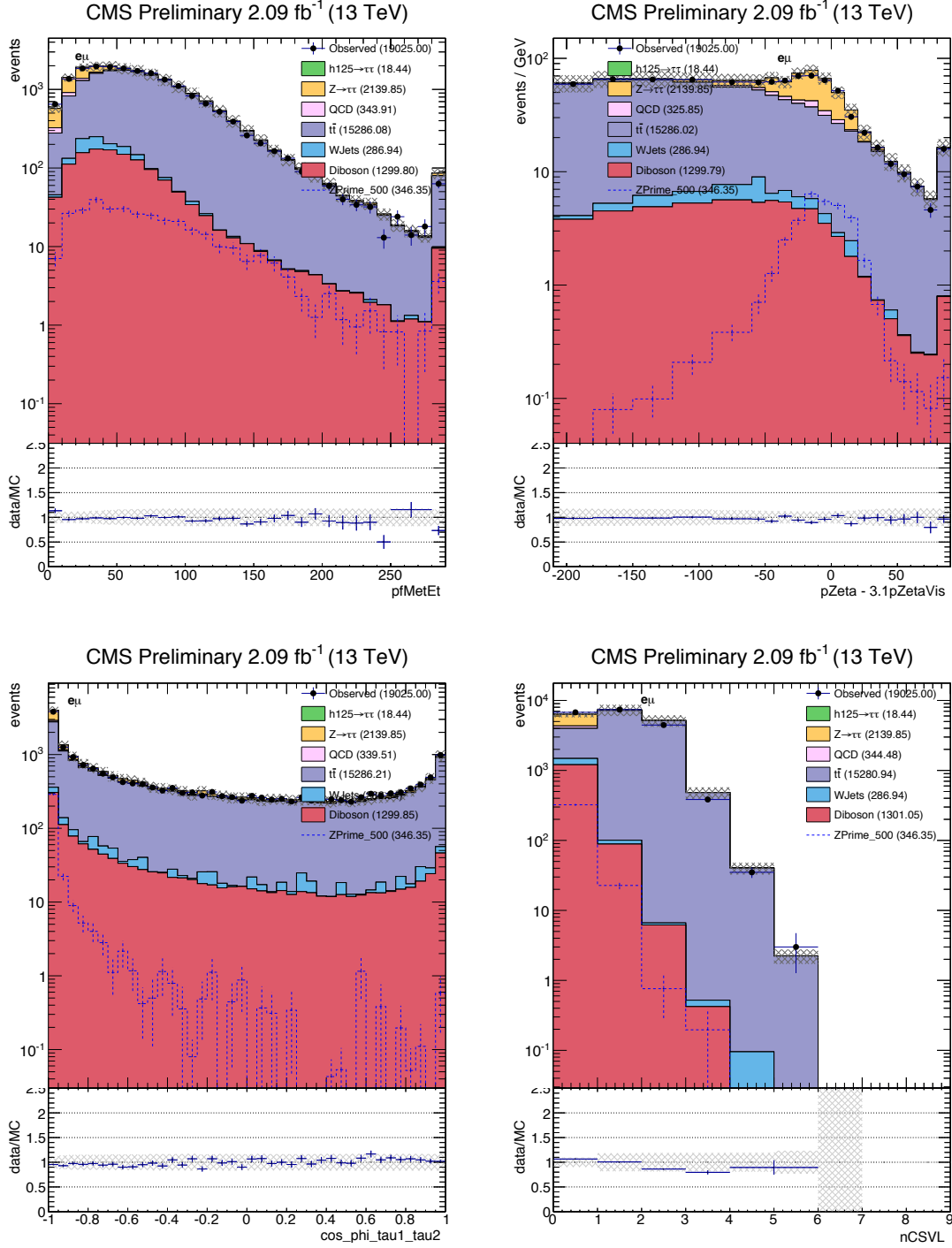
Figure 3.51: Distributions, after preselection, of the variables used for the signal selection: (top left), "ζ" (top right), cos Δφ(e, μ) (bottom left), and n_b (bottom right).

Figure 3.52: Distributions, after preselection, of electron (top left), electron pseudo-rapidity (top right), muon (bottom left), muon pseudo-rapidity (bottom right).
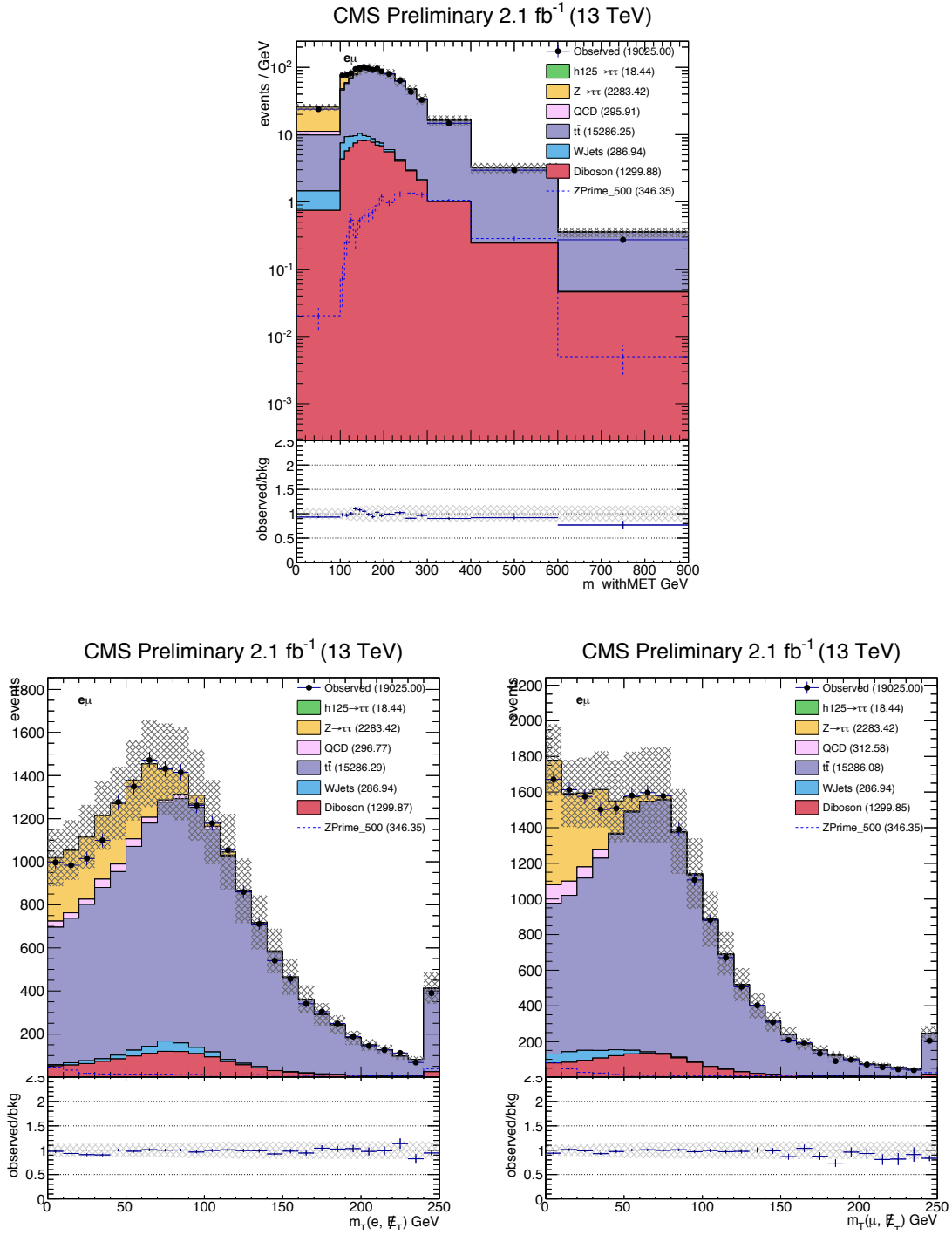
Figure 3.53: Distributions, after preselection, of $m\,(e,\mu,)$ (top), $m_T\,(e,)$ (bottom left), and $m_T\,(\mu,)$ (bottom right).
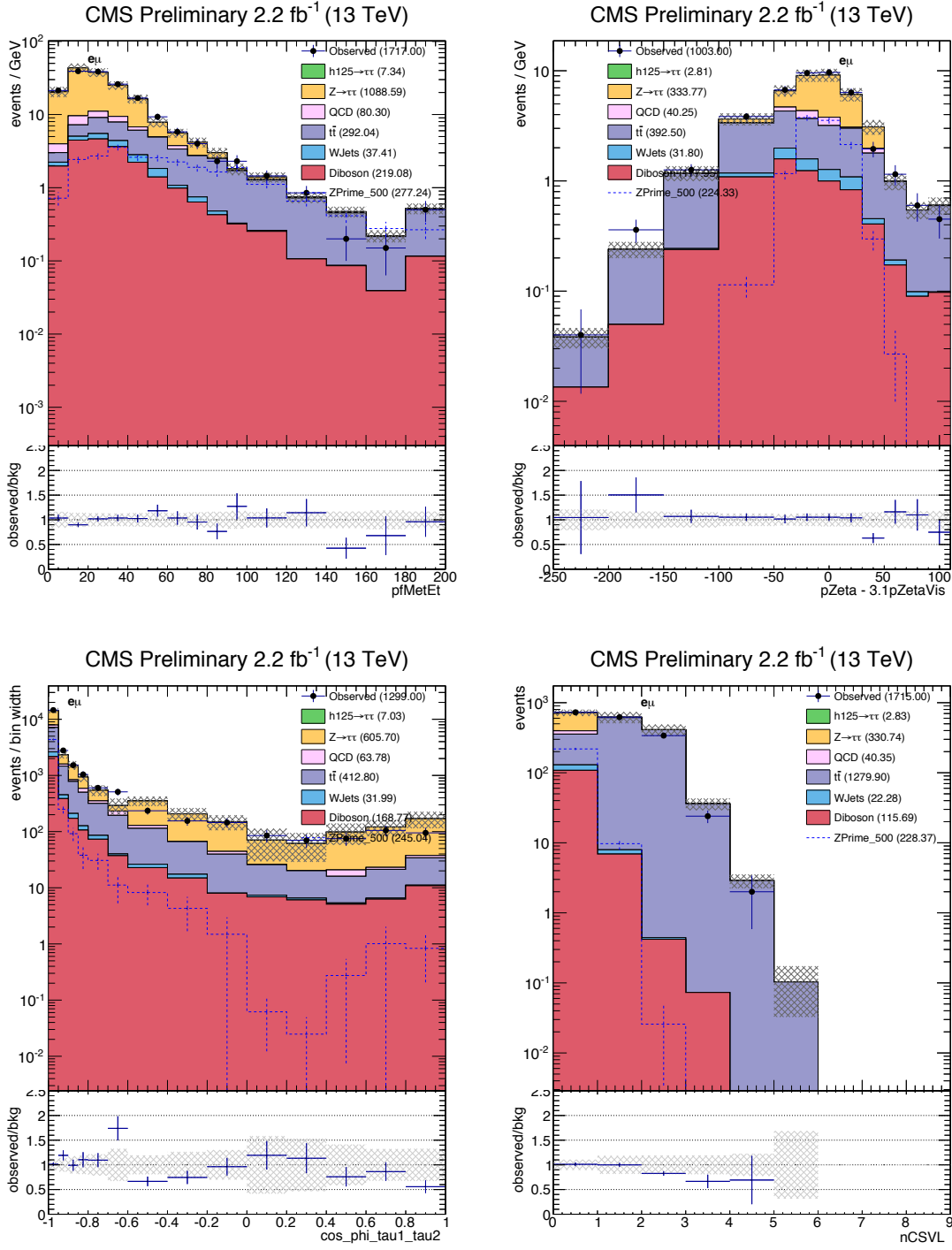
Figure 3.54: Distributions of the variables used for signal selection, after all other signal selection requirements on variables other than the one plotted: (top left), "ζ" (top right), $\cos\Delta\phi(e,\mu)$ (bottom left), and n_b (bottom right).
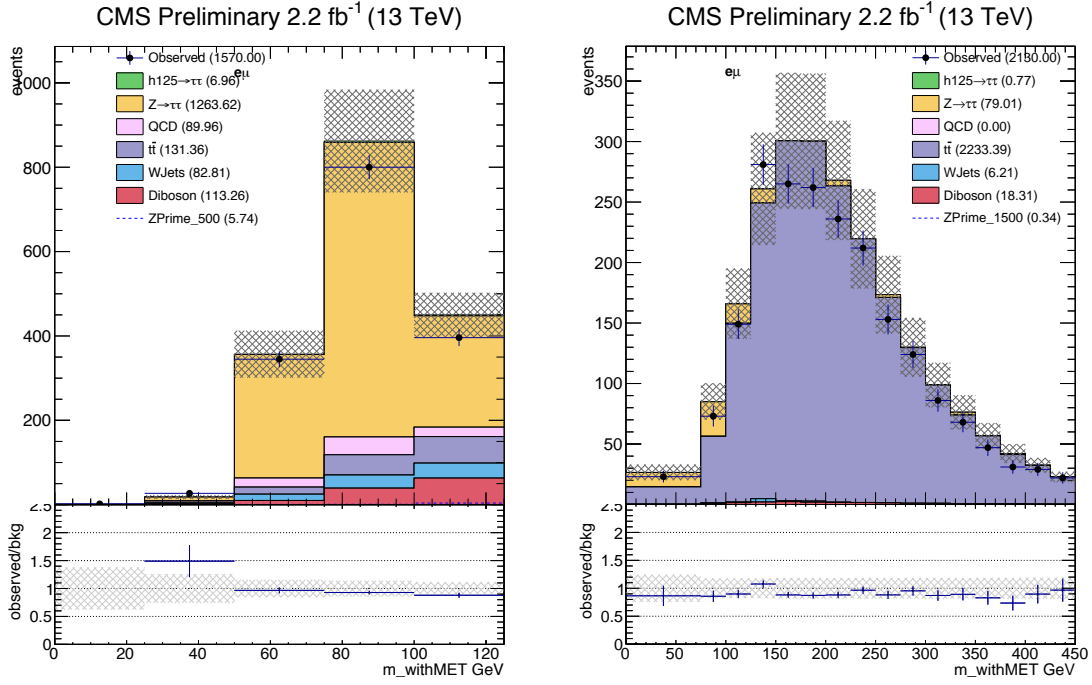
Figure 3.55: Distributions of . Left: validation region with < 30 , $n_b = 0$ and < 125 GeV. Right: validation region with > 30 , $n_b \geq 1$ and $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$.
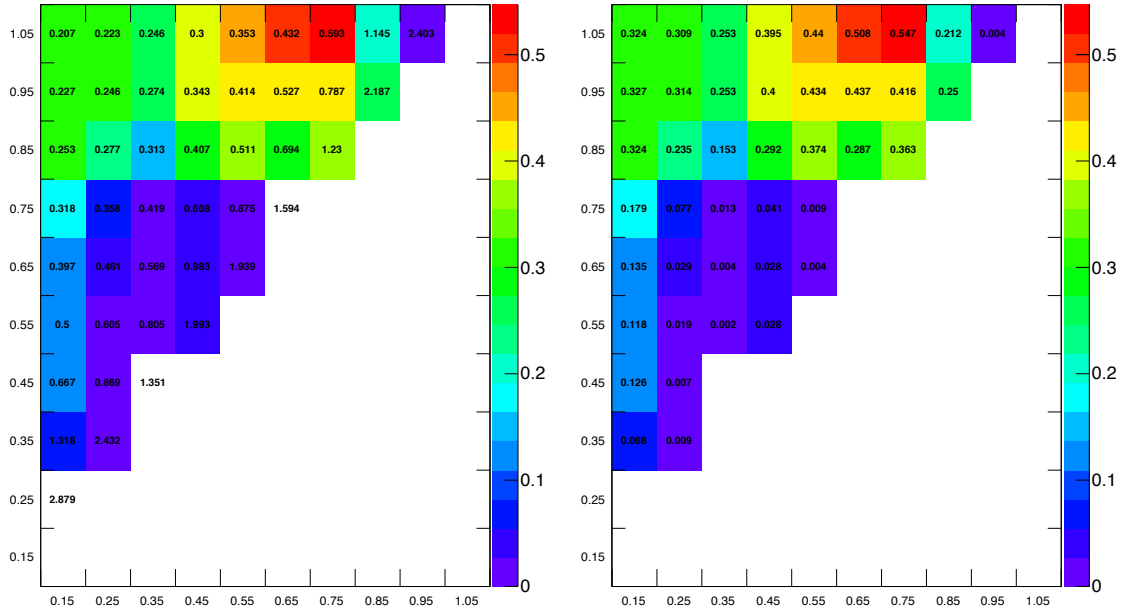
Figure 3.56: channel: scan of the range of relaxed relative isolation for the muon. Left: text within each bin gives the normalization factor applied to same-charge iso-relaxed events ("loose to tight" factor); the color axis matches the right plot. Right: χ^2 of the agreement between the predicted and observed distributions of tightly-isolated same-charge events.
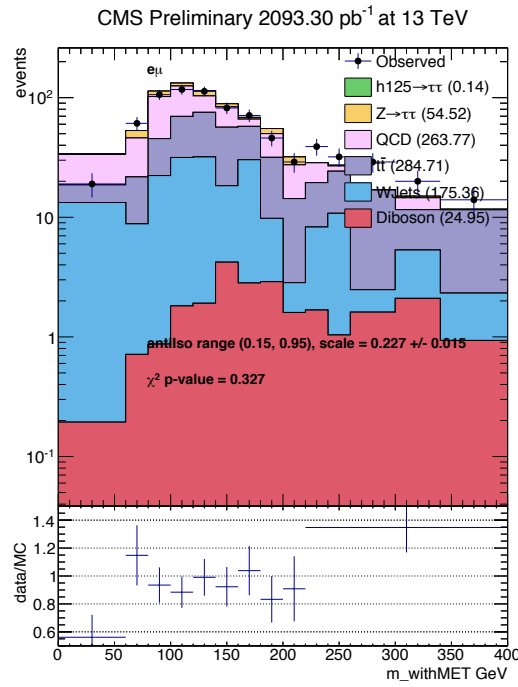
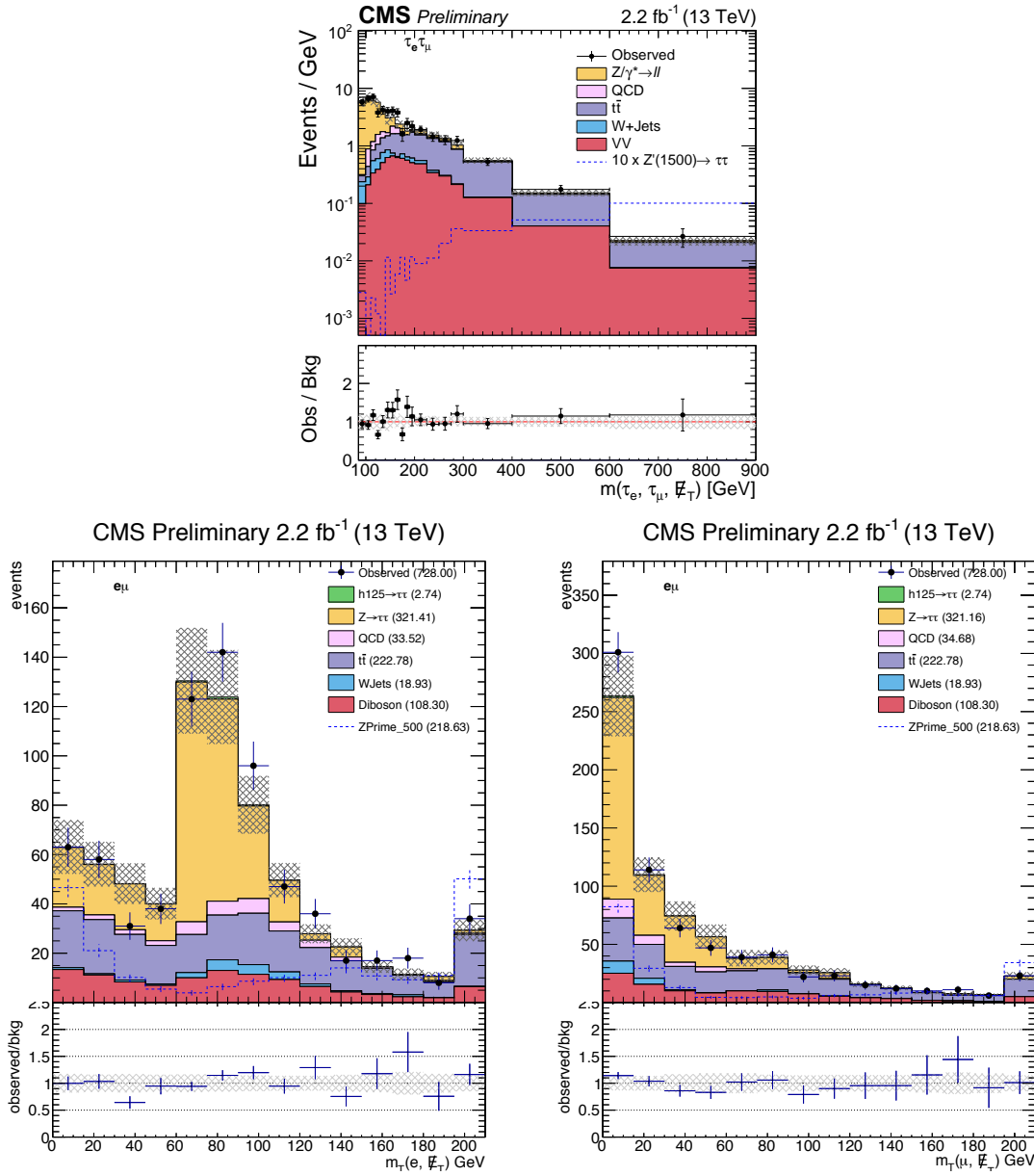Figure 3.57: The distribution of reconstructed parent mass, , in the same-charge, tightly-isolated sample: channel.

Figure 3.58: Top: predicted background yields and observed event yields in the channel after signal selection. Bottom: distributions of transverse mass.
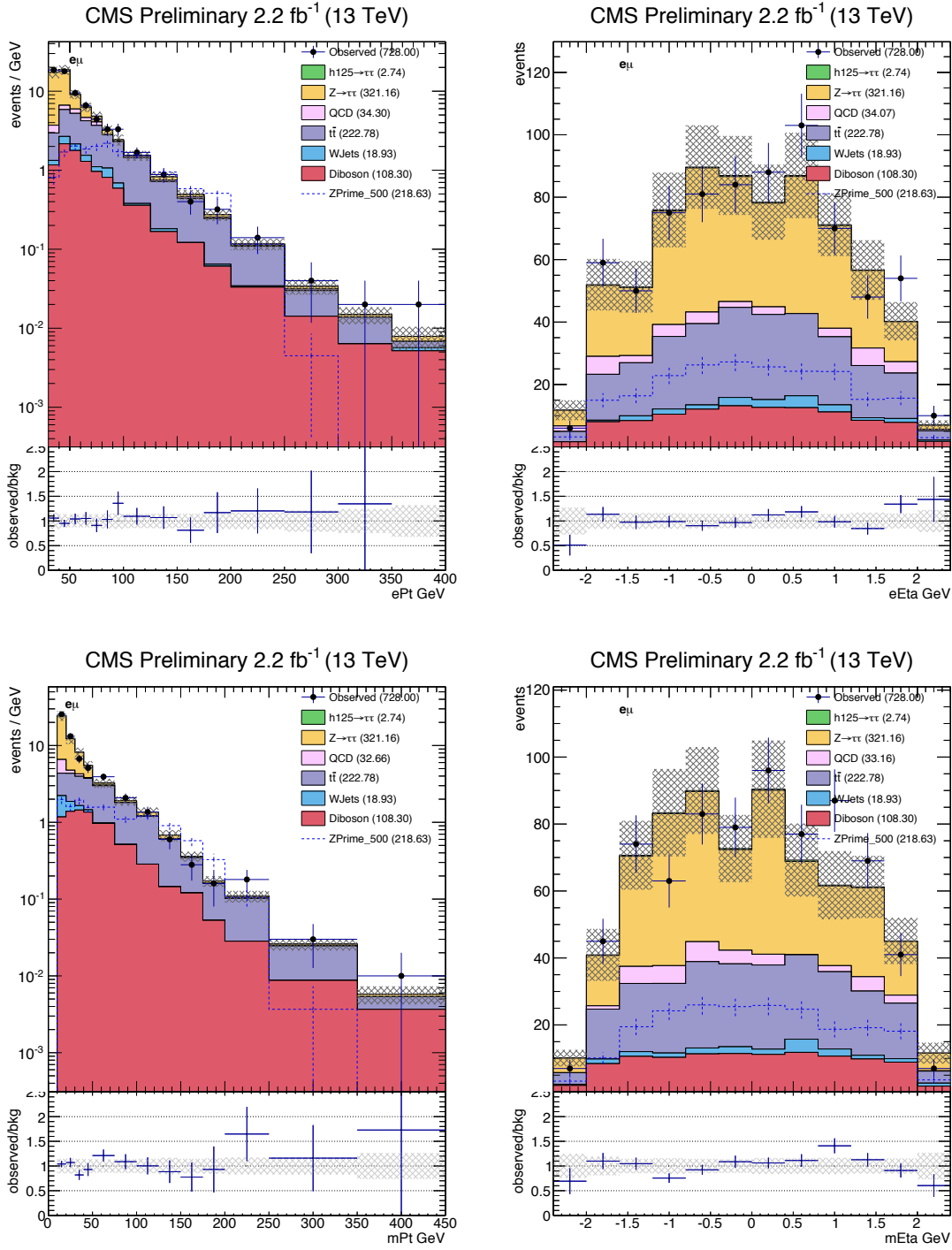
Figure 3.59: Distributions, after final selection, of electron (top left), electron pseudo-rapidity (top right), muon (bottom left), muon pseudo-rapidity (bottom right).
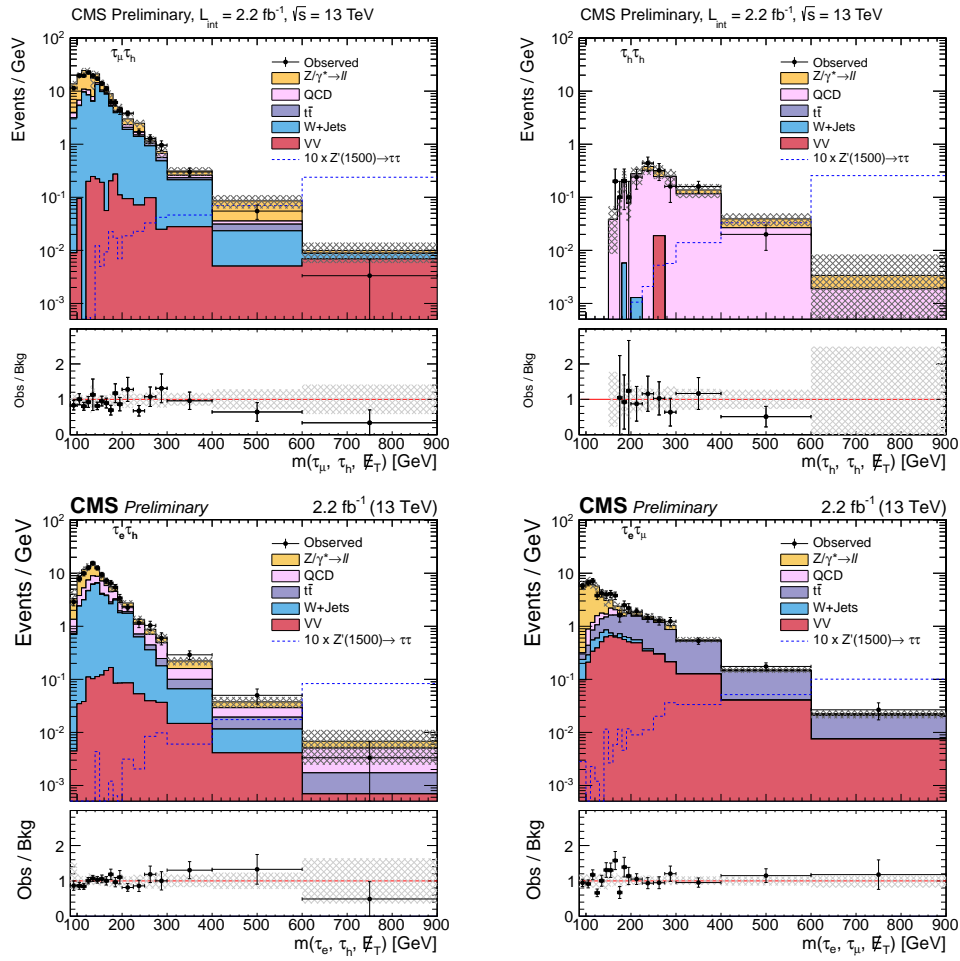
Figure 3.60: Top Left: $m(\mu, \tau_h,)$ distribution in the signal region, in log scale. Top Right: $m(\tau_h, \tau_h,)$ distribution in the signal region, in log scale. Bottom Left: $m(e, \tau_h,)$ distribution in the signal region, in log scale. Bottom Right: $m(e, \mu,)$ distribution in the signal region, in log scale.
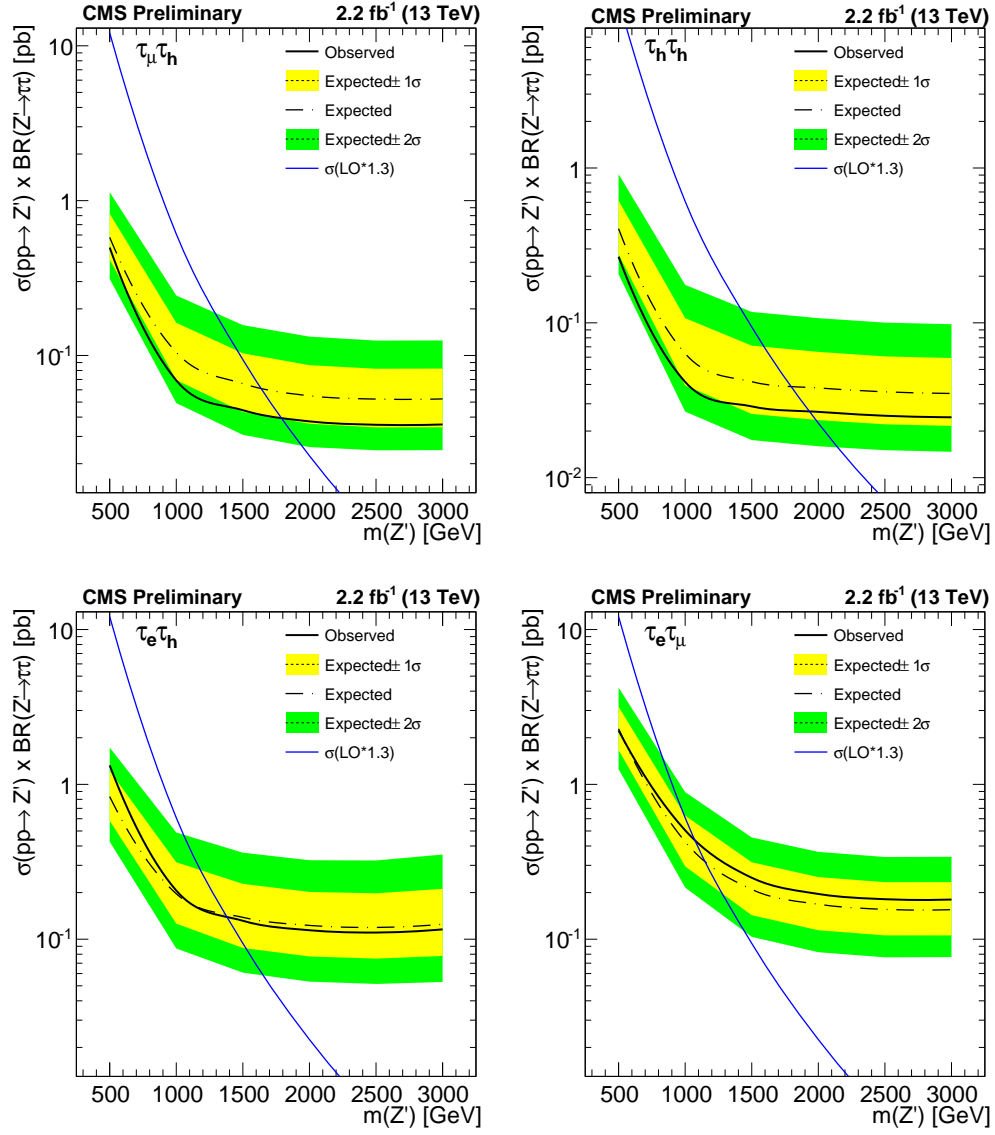
Figure 3.61: Expected and observed limits for the $\tau_\mu\tau_h$, $\tau_h\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$ channels. A k factor of 1.3 has been used to scale the leading order (LO) signal cross-section.
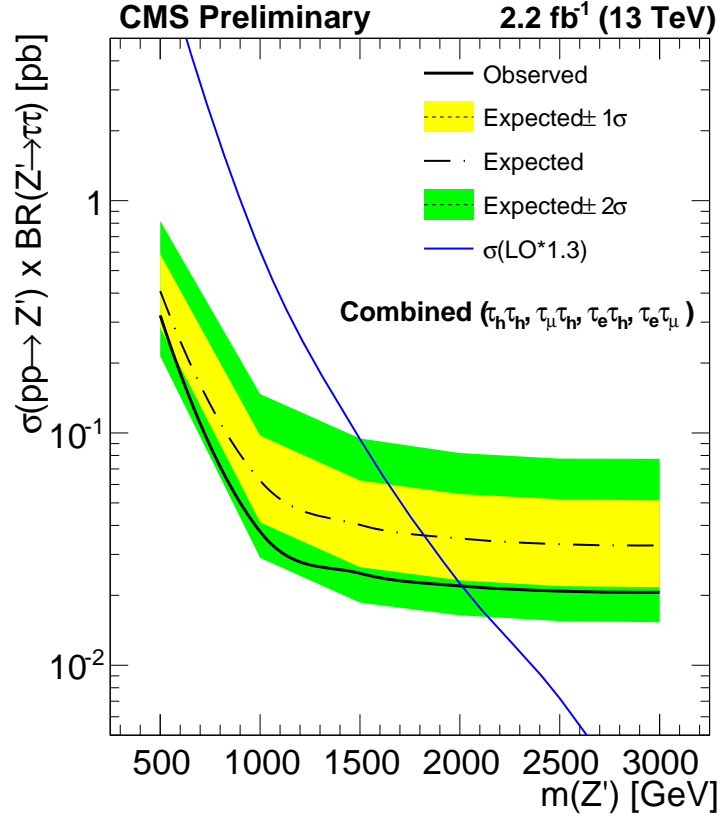
Figure 3.62: Combined expected limit for the $\tau_\mu\tau_h$, $\tau_h\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$ channels.

Figure 3.63: Combined expected limit for the $\tau_\mu\tau_h$, $\tau_h\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$ channels. A k factor of 1.3 has been used to scale the leading order (LO) signal cross-section.